The role of the emotive, moral, and cognitive components for the prediction of medical students' empathic behavior in an Objective Structured Clinical Examination (OSCE)

Graupe, Tanja; Giemsa, Patrick; Schaefer, Katharina; Fischer, Martin R.; Strijbos, Jan Willem; Kiessling, Claudia

Published in:
Patient Education and Counseling

DOI:
10.1016/j.pec.2022.06.017

Publication date:
2022

Patient Education and Counseling 105 (2022) 3103–3109

Contents lists available at ScienceDirect

Patient Education and Counseling

journal homepage: www.journals.elsevier.com/patient-education-and-counseling

The role of the emotive, moral, and cognitive components for the prediction of medical students' empathic behavior in an Objective Structured Clinical Examination (OSCE)

Tanja Graupe [a,*], Patrick Giemsa [b], Katharina Schaefer [a], Martin R. Fischer [a], Jan-Willem Strijbos [c], Claudia Kiessling [b]

[a] Institute of Medical Education, University Hospital, LMU Munich, Munich, Germany
[b] Faculty of Health, Chair for the Education of Personal and Interpersonal Competences in Health Care, Witten/Herdecke University, Witten, Germany
[c] Faculty of Behavioural and Social Sciences, Department of Educational Sciences, University of Groningen, the Netherlands

ARTICLE INFO

Keywords:
Medical education
Assessment
Communication skills
Empathy
Verona CodingDefinitions of Emotional Sequences (VR-CoDES)
Situational Judgment Test (SJT)
Objective Structured Clinical Examination (OSCE)

ABSTRACT

Objectives: Investigate whether medical students' emotive abilities, attitudes, and cognitive empathic professional abilities predict empathic behavior in an Objective Structured Clinical Examination (OSCE).
Methods: Linear and multiple regressions were used to test concurrent validity between Interpersonal Reactivity Index (IRI), Jefferson Scale of Physician Empathy (JSPE-S), Situational Judgement Test (SJT-expert-based score (SJT-ES), SJT-theory-based score (SJT-TS)) and empathic behavior in an OSCE measured by Berlin Global Rating (BGR) and Verona Coding Definitions for Emotion Sequences (VR-CoDES).
Results: Highest amounts of explained variance of empathic behavior measured by VR-CoDES were found for the SJT-ES ($R2 = 0.125$) and SJT-TS ($R2 = 0.131$). JSPE-S ($R2 = 0.11$) and SJT-ES ($R2 = 0.10$) explained the highest amount of variance in empathic behavior as measured by BGR. Stepwise multiple regression improved the model for BGR by including SJT-ES and JSPE-S, explaining 16.2% of variance.
Conclusions: The instrument measuring the emotive component (IRI) did not significantly predict empathic behavior, whereas instruments measuring moral (JSPE-S) and cognitive components (SJT) significantly predicted empathic behavior. However, the explained variance was small.
Practice implications: The instrument measuring the emotive component (IRI) did not significantly predict empathic behavior, whereas instruments measuring moral (JSPE-S) and cognitive components (SJT) significantly predicted empathic behavior. However, the explained variance was small.
In a longitudinal assessment program, triangulation of different instruments assessing empathy offers a rich perspective of learner's empathic abilities. Empathy training should include the acquisition of knowledge, attitudes, and behavior to support learner's empathic behaviors.

1. Introduction

Communication is one of the core competencies of health professionals. Effective clinical communication improves the quality of care and facilitates the collection of patients' data, understanding patients' perspective, providing information, making decisions about treatments, and handling patients' emotions [1]. An empathic response to patients' emotions is a key element of patient-centered care [2,3] . Positive effects of providers' empathic communication on patient health outcomes have been shown [4–10] and many catalogues of educational objectives have integrated emotion-handling skills, like empathy, into their guidelines [11–14].

Although the role of empathy in clinical care has been studied for many years, there is no consensual definition. A frequently referenced definition by Mercer and Reynolds [15] defines clinical empathy as a complex multidimensional concept and as the "ability to: (a) understand

Abbreviations: BGR, Berlin Global Rating; IRI, Interpersonal Reactivity Index; JSPE, Jefferson Scale of Physician Empathy; OSCE, Objective Structured Clinical Examination; SJT, Situational Judgment Test; SJT-ES, Situational Judgment Test - Expert-based score; SJT-TS, Situational Judgment Test - Theory-based score; VR-CoDES, Verona Coding Definitions for Emotion Sequences.

* Correspondence to: Pettenkoferstr. 8a, 80336 Munich, Germany.
E-mail address: Tanja.Graupe@med.uni-muenchen.de (T. Graupe).

https://doi.org/10.1016/j.pec.2022.06.017
Received 3 March 2022; Received in revised form 22 June 2022; Accepted 27 June 2022
Available online 28 June 2022

the patients' situation, perspective, and feelings (and their attached meanings); (b) communicate that understanding and check its accuracy; and (c) act on that understanding with the patient in a helpful (therapeutic) way" [15 p. 9]. Empathy can be conceived as a set of professional skills or competencies [15]. Morse et al. [16] distinguish four components: emotive, moral, cognitive, and behavioral. Fig. 1 shows the definitions of the four components.According to Hemmerdinger [17], empathy can be measured from three different perspectives: self-rating, patient-rating, and observer-rating. Fig. 2 gives an overview of the measurement perspectives and examples of relevant assessment instruments. Many instruments have been developed to measure communication skills in an Objective Structured Clinical Examination (OSCE) [22], either in the form of station-specific checklists or global ratings [23–28], like the Berlin Global Rating (BGR) [29]. Empathic behavior in real clinical situations can be evaluated by direct observation or by analyzing video or audio, using coding instruments for sequence analysis, like the Roter Interaction Analysis System (RIAS), which is a more general instrument and less dedicated for the observation of emphatic interaction [30], or the Verona Coding Definitions for Emotional Sequences (VR-CoDES) [31]. VR-CoDES analyze how providers react on patients' emotional cues and concerns by providing or reducing space to further elaborate on emotional issues. Providing space responses, like empathy, demonstrate an observable understanding of patients' feelings and perspective [16]. VR-CoDES can be used to evaluate recorded students' (empathic) responses to simulated patients' emotional cues and concerns in an OSCE [32].

To date, Hemmerdinger's categories for the assessment of empathy have not included assessment with written or video-based tests. Written and video-based instruments typically assess different cognitive abilities [33,34], which play an important role for demonstrating communicative competencies [35,36]. Many of these instruments use scenarios with patients, including communication challenges, and ask learners to analyze them and provide strategies to handle the situation or to judge predefined options for behavior [36–39]. A specific instrument to assess 'non-academic' attributes like empathy is the Situational Judgement Test (SJT) [40]. SJTs were initially developed for selection purposes but appear a feasible approach for assessing communication skills [41–43]. SJTs are based on the behavioral consistency theory, which implies that anticipated behavior can predict future behavior [44]. SJTs confront students with written or video-based work-related scenarios and ask them to evaluate different reactions within the scenarios [40]. Response formats vary from single-best responses to rating and ranking formats [45–47]. Students' scores are typically based on expert panels [40]. SJTs appear effective predictors of performance in practice [42,48–50]. SJTs measure procedural knowledge about empathy, which can be put on a level with the cognitive component of empathy, according to Morse et al. [16].

To our knowledge, evidence is scarce to which extent the emotive, moral, and cognitive components of empathy, as shown in Fig. 1, predict empathic behavior in real-life situations. However, only few studies examined concurrent or predictive validity of knowledge or behavioral attitudes in clinical communication [42,51,52] or empathy [53–56]. Hence, we investigated concurrent validity and, more specifically, the question whether medical students' (a) self-assessed emotive abilities, (b) attitudes towards empathy (manifestation of the moral component), and (c) cognitive empathic abilities predict empathic behavior in an OSCE.

2. Methods

2.1. Setting

We conducted an experimental study at LMU Munich to develop and validate a SJT to assess medical students' empathy [43] and an OSCE to assess communication skills in the field of general medical practice [57]. Medical students from all study years were invited. Eighty-seven medical students completed the SJT, the OSCE, and a questionnaire. The questionnaire contained 77 items covering demographic data, the Interpersonal Reactivity Index (IRI) [18,19], the Jefferson Scale of Physician Empathy (JSPE-S) [20,21], self-rated communicative competencies during the OSCE, and acceptance of the SJT. Participation was voluntary and reimbursed with 20 Euros. The study was approved by the Ethics Committee at LMU Medical Faculty.

2.2. Measurement instruments

Fig. 3 gives an overview of the instruments we used to measure Morse's four components of empathy [16], according to Hemmerdingers measurement perspective [17].

Emotive component: Interpersonal Reactivity Index (IRI).

The IRI measures self-assessed cognitive capabilities and the emotional reactivity of individuals [18,19]. The German version [6] comprises four subscales and 28 items which were answered on a five-point Likert-like scale from 1 (*does not apply at all*) to 5 (*does completely apply*). The subscales 'Perspective Taking' and 'Fantasy' cover cognitive aspects of empathy. The subscales 'Empathic Concern' and

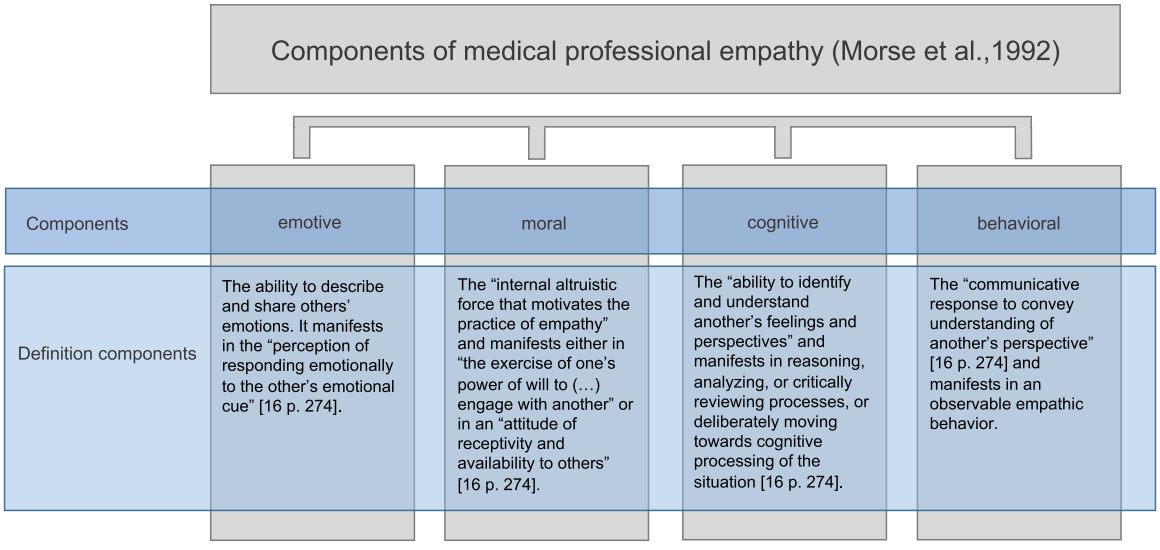

Fig. 1. Components of medical professional empathy according to Morse et al. (1992).

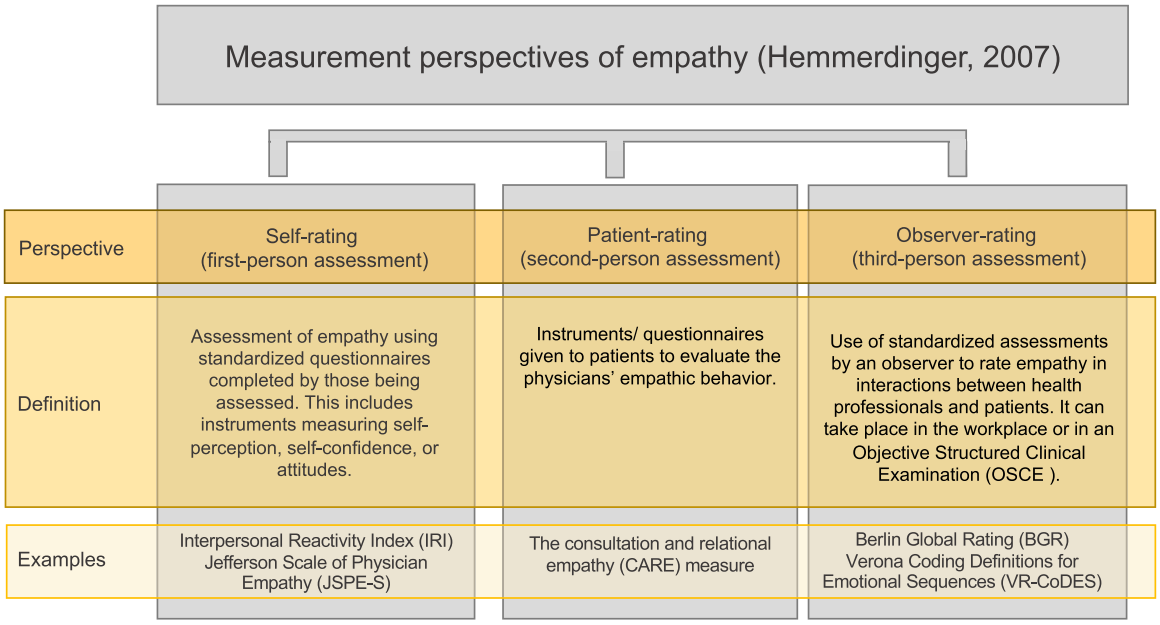

Fig. 2. Measurement perspectives of empathy according to Hemmerdinger et al. (2007).

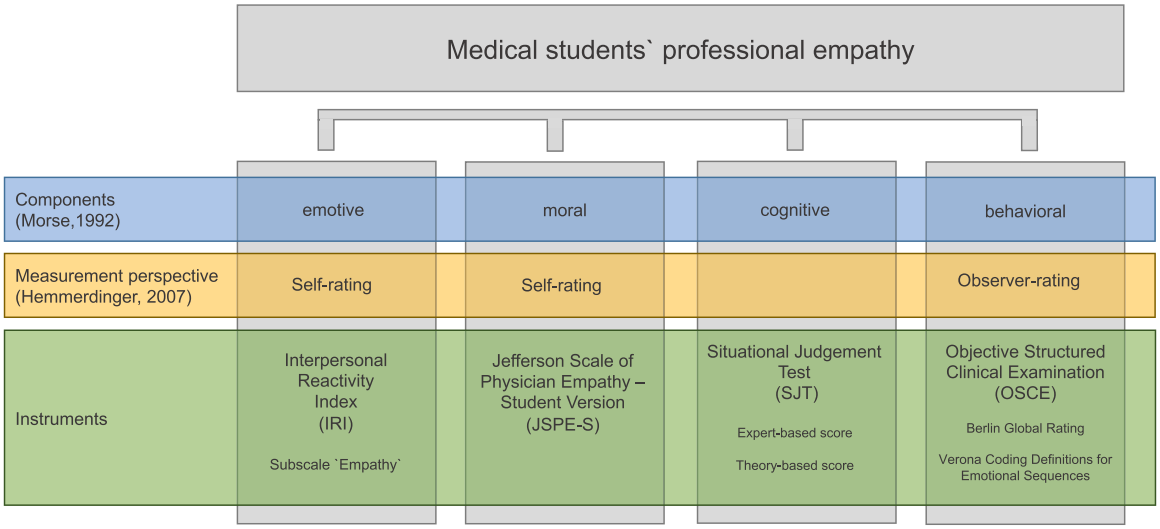

Fig. 3. Instruments used to measure the components of empathy on different perspective.

'Personal Distress' cover emotional aspects of empathy [18,19]. For the purpose of this study, we used the subscale 'Empathic Concern' (7 items, Cronbach's α = .63).

Moral component: Jefferson Scale of Physician Empathy-Student version (JSPE-S).

The JSPE measures health care providers' or students' orientations and attitudes towards empathy [20,21]. In the German student version (JSPE-S) [58] students answered on a seven-point Likert- scale ranging from 1 (*strong disagreement*) to 7 (*strong agreement*) (20 items, Cronbach's α = .79).

Cognitive component: Situational Judgement Test (SJT).

The SJT consists of 12 case vignettes showing real-life situations of physicians and medical students interacting with patients and relatives. Each vignette consists of two consecutive parts and includes (1) a video representing a critical incident containing patients' or relatives' emotional concern(s) and/or cue(s), (2) a standardized lead-in-question, and (3) five response alternatives (115 responses in total). Of these, 28 were 'Non-explicit – Reduce space' (NR), 30 were 'Explicit – Reduce space' (ER), 16 were 'Non-explicit – Provide space' (NP), and 41 were 'Explicit – Provide space' (EP) according to VR-CoDES. Students were asked to rate each of the responses on a slider scale from 1 (*very inappropriate*) to 100 (*very appropriate*). The SJT provides two scores: (1) Expert-based-score (SJT-ES): students' responses were compared with the responses of an expert panel (max. 23 points), (2) Theory-based-score (SJT-TS): students received a point if they identify those response(s), which provided space (according to VR-CoDES) as being appropriate (max. 57 points). The development, validation and psychometric properties of the SJT have been described elsewhere [43].

Behavioral component: Objective Structured Clinical Examination (OSCE).

The OSCE included four stations and was based on typical patient consultations in general practice. To simulate a primary care setting, students did not rotate through a circuit of stations but stayed in one room (like in a consultation room), and one standardized patient after another entered this room and presented their problem. All student interactions with standardized patients were recorded on video. The

development, validation, and psychometric properties of the OSCE have been described elsewhere [57].

We used two instruments to analyze students' communication skills in the OSCE:

(1) Berlin Global Rating (BGR) [29] is the German version of the Analytic Global Rating [26] and contains four items: response to patient's feelings and needs (empathy), degree of coherence in the interview, verbal expression, and nonverbal expression. Raters used a five-point scale with verbal anchors for each item ranging from 1 (*not competent*) to 5 (*competent*). A training was conducted with four raters, each rating ten stations in parallel. Interrater-reliability was good for all raters (Spearman-Rho >. 800). Subsequently, one rater rated all videos from the OSCE. Inter-station reliability was good for the BGR ($\alpha = 0.87$) and specifically for the item 'empathy' ($\alpha = 0.79$).

(2) Video-recorded student interactions were also analyzed using Verona Coding Definitions for Emotional Sequences (VR-CoDES) [31]. First, units of analyses were defined, and patients' and relatives' cues and concerns analyzed by independent raters. Afterwards, students' responses were coded into 'explicit' versus 'non-explicit' and into 'provide space' versus 'reduce space' answers and subsequently classified into the individual codes in VR-CoDES. Trainings with three raters were conducted for each station. After parallel and independent rating of ten videos, consensus discussions were conducted to ensure inter-rater agreement. For training videos 5–10, Cohen's kappa was good (explicit vs. non-explicit $\varkappa = 0.89$; provide space vs. reduce space $\varkappa = 1.0$; $\varkappa = 0.77$ for the 17 different responses according to the VR-CoDES). Afterwards, one rater coded the remaining videos. Any ambiguous responses were discussed and coded by the researcher team.

For each student, the number and type of responses were counted for each station and overall, across all station. A score for responses, which provide space was calculated based on all responses. As empathic responses according to the VR-CoDES (NPIm, EPAEm) were rare (only 3.4% of all responses), we used the number of providing space responses for further analysis.

2.3. Data analysis

We used R version 3.5.1 for OS X 10.14 with statistical significance levels set to $p < 0.05$ and Shapiro-Wilk's to test assumptions of normal distribution. We checked for multicollinearity using the variance influence factor (VIF < 2). After correlating our predictor and outcome variables (Table 1), we calculated simple linear regressions for all predictor variable (IRI subscale 'Empathic Concern', JSPE-S, SJT-ES, SJT-TS) to show the amout of change of each variable in our criterion 'variables for empathic behavior' (VR-CoDES: provide space; BGR: empathy). We then calculated two multiple regression models to check additional explained variance in 'VR-CoDES: provide space' and 'BGR: empathy', compared to the initial models. Predictor variables were added stepwise in order of correlation with the criterion variable, highest correlation first.

3. Results

Table 2 shows the descriptive statistics for all scales. The simple linear regressions for each of the four predictors are reported per criterion variable in Table 3 (VR-CoDES: provide space) and Table 4 (BGR: Empathy). The highest amounts of explained variance in empathic behavior as measured by 'VR-CoDES: provide space' was found for the SJT-ES (R^2 =.125) and SJT-TS (R^2 =.131), whereas self-rating scales (IRI subscale 'Empathic Concern' and JSPE-S) only explained minor amounts of variance. The JSPE-S (R^2 =.141) explained the highest amount of variance in empathic behavior as measured by 'BGR: empathy'. All other simple linear regressions were not significant as expected by the previous correlation.

The addition of predictors in a stepwise multiple regression did not significantly improve our prediction for empathic behavior as measured by 'VR-CoDES: provide space', leaving the SJT-TS as the single most fitting predictor. A stepwise multiple regression improved our simple model for 'BGR: empathy'. The stepwise multiple regression model included the SJT-ES and JSPE-S, explaining 16.2% of variance. Addition of further predictors did not improve the model.

4. Discussion and conclusion

4.1. Discussion

This study investigated whether instruments capturing Morse's emotive, moral and cognitive components of empathy [16] could predict medical students' empathic behavior in an OSCE (see Fig. 3). Our findings showed that the instrument measuring the emotive component did not significantly predict empathic behavior, whereas the instrument measuring the moral component significantly predicted empathic behavior.

Both scores of the Situational Judgemet Test (SJT-ES and SJT-TS) predicted empathic behavior; however, the theory-based score (SJT-TS) only significantly predicted behavior measured by VR-CoDES and not behavior measured by the BGR. All correlation coefficients were low to medium. Stepwise multiple regression showed that the theory-based score (SJT-TS) predicted empathic behavior measured by VR-CoDES and that the expert-based score (SJT-ES) as well as JSPE-S predicted empathic behavior measured by the BGR.

Overall, it might be questioned whether the selected instruments sufficiently captured the four components of empathy as defined by Morse [16]. Davis, who developed the IRI and its four subscales [18,19], defined empathy in two facets, a cognitive and an affective facet. We used only the subscale 'Empathic Concern', which did not predict empathic behavior measured in an OSCE setting. To our knowledge, validity studies of the IRI only include personality traits and self-reported behavior [59–61] and not empathic behavior assessed by observer-rating.

Furthermore, studies revealed inconsistent evidence about correlations between self-assessment or self-reflection and clinical performance [62–64]. However, in our study, JSPE-S, which is based on self-rating, correlated significantly with empathic behavior measured by BGR. Hojat and colleagues, who developed the JSPE, defined empathy as a cognitive activity and distinguished it from sympathy [20,21,65]. Notwithstanding the definition of empathy, the JSPE was designed to measure orientation and attitudes towards empathy [21]. Therefore, it appears appropriate for measuring the moral component of empathy, because the moral component according to Morse manifests in attitudes

Table 1
Pearson's correlations.

Variable		VR-CoDES Provide Space		BGR-Empathy		SJT-ES		SJT-TS	JSPE-S	
BGR-Empathy	Pearson's r	0.601	**	—						
SJT-ES	Pearson's r	0.379	**	0.341	*	—				
SJT-TS	Pearson's r	0.362	**	0.199		0.597	**	—		
JSPE-S	Pearson's r	0.231	*	0.334	*	0.347	**	0.135	—	
IRI-Empathy	Pearson's r	-0.034		0.108		0.044		-0.006	0.538	**

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2
Mean and Standard Deviation for all scales for the entire sample and split by gender.

Score	VR-CoDES Provide Space		BGR Empathy		SJT-ES		SJT-TS		JSPE-S		IRI Empathy	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Gender												
Female (n = 64)	9.10	3.26	6.89	2.73	11.42	3.72	30.00	11.04	118.22	8.58	21.26	3.28
Male (n = 20)	6.10	2.65	4.85	2.06	9.25	3.67	25.95	10.98	113.15	12.12	18.95	3.20
total sample (n = 87)	8.26	3.41	6.40	2.70	10.83	3.77	28.83	11.09	116.81	9.74	20.69	3.39

Table 3
Simple Regression Results for SJT-ES, SJT-TS, IRI Empathy and JSPE-S on VR Codes – Provide Space.

	B	*SE B*	β	R^2	*t*	*p*
					3.463	
SJT-ES	.012	.004	0.353	0.125	3.556	< .001
SJT-TS	.004	.001	0.362	0.131	-0.315	< .001
IRI Empathy	-.001	.004	-0.034	0.001	2.189	0.753
JSPE-S	.003	.001	0.231	0.053		0.031

Table 4
Simple Regression Results for SJT-ES, SJT-TS, IRI Empathy and JSPE-S on BGR Empathy.

	B	*SE B*	β	R^2	*t*	*p*
					1.920	
SJT-ES	.148	.077	0.205	0.042	1.311	0.058
SJT-TS	.035	.026	0.142	0.020	1.000	0.194
IRI Empathy	.086	.086	0.108	0.001	3.738	0.320
JSPE-S	.105	.028	0.376	0.141		< .001

[16]. In our study, the JSPE-S predicted empathic behavior to a low to moderate level, and Hojat et al. reported significant correlations between the JSPE and an OSCE [66].

SJTs are relatively new in the field of medical education although extended research has been published in the field of job admission and selection procedures [41,44–47,50]. Our results about the low but significant correlations between SJT results and performance-based scores are comparable to other publications in the field of medical education [42,67,68]. However, given that VR-CoDES informed the scoring of the SJT and part of the OSCE, it might not be surprising that the theory-based score of the SJT (SJT-TS) predicted that equivalent part of the OSCE score. However, the SJT-TS was not able to predict the OSCE result measured by the BGR. The expert-based score of the SJT (SJT-ES) was able to predict both OSCE results, the one based on VR-CoDES and the one based on BGR. So, what is the difference between the score based on theory and the score based on an expert panel? Experts apply their knowledge in practice by adapting their behavior to the context. Accordingly, experts do not simply use a theory or theoretical knowledge and apply it. They also adapt their behavior to the specific patient and to the specific situation in accordance with their knowledge and experience. A score based on an expert panel might be closer to real clinical practice and perhaps also student behavior in an OSCE than a score based solely on a theoretical model.

Nevertheless, there might be some bias in our sample. Medical students participated voluntarily and the cohort might be a selection of highly motivated students or students with special interest in communication skills or even empathy. We tried to reduced the bias precautionary by not mentioning the content of our study in the invitation email, i.e. students were invited to help us develop new assessment instruments for medical education. However, women were overrepresented in our sample [43] and female students achieved higher empathy scores than male students, which is concordant with other findings in the literature. Possible reasons could be sociocultural and social learning factors. However, there might also be evolutionary reasons, like women seem to better be equipped with a larger capacity for social relationship, show more sensitivity to social stimuli and emotional cues and concerns and demonstrate more care-oriented abilities from an early age on [69,70].

Also, we used a study design with only one measurement point. Humphris [71] found that Objective Structured Video Examinations (OSVE) results predicted short-term but not long-term OSCE results. Therefore, it would be interesting to replicate our results in a prospective study using the JSPE-S, SJT and other scores in an OSCE and workplace-based assessments with a group of medical students over their course of study and the start of their residency to see if there are any long-term predictions and get a real impression of learners' behavior.

4.2. Conclusion

We conclude that we can approximate to an understanding of empathy and how to measure it by combining different instruments covering its individual components on different measurement perspectives (Fig. 3). We based our study on a definition and on instruments, which are well-established and validated in the field of medical education, and we also tried to find a range of instruments for every component of this conceptualization of empathy. Our study provides evidence that the components in Morses' multidimensional model of empthy [16] are intercorrelated and not separate entities. Although the components of empathy are intercorrelated, assessment instruments measure different facets of empathy. Therefore, within a longitudinal assessment program, triangulation of assessment instruments including self-assessments, tests, OSCEs and workplace-assessments can provide a more complete picture of students' abilities from different perspectives than single stand-alone course assessments [72,73]. Within a longitudinal assessment program these instruments might be used in adaption to the learners' level of competence and to the level of training. Based on our findings, we would assume that for the acquisition of empathic abilities, all components need to be trained and learnt. Procedural and conceptional knowledge, the ability to share others' emotions and the willingness to approach others in an empathic way lead to better empathic behavior towards patients in real-life situations.

4.3. Practice implications

We believe that, like in an assessment program of clinical competence in general, no single instrument or assessment method can cover all aspects. We suggest a longitudinal assessment program with a multisource collection of information about learners using different instruments to get a rich impression of their empathic attitudes, knowledge, and behavior to predict empathic behavior as valid as possible. Training and assessment should address all components of empathy to support the acquisition of empathic abilities in learners.

CRediT authorship contribution statement

Tanja Graupe: Conceptualization, Methodology, Development of SJT (dissertation), Data collection, Writing – original draft. **Patrick**

Giemsa: Methodology, Data analysis, Writing – original draft. **Katharina Schaefer**: Conceptualization, Methodology, Data collection, Coding and data analysis of VR-CoDES (dissertation), Writing – review & editing. **Martin R. Fischer**: Conceptualization, Methodology, Writing – review & editing. **Jan-Willem Strijbos**: Conceptualization, Methodology, Writing – review & editing. **Claudia Kiessling**: Conceptualization, Methodology, Data collection, Writing – original draft.

Declaration of Competing Interest

The role of the emotive, moral, and cognitive components for the prediction of medical students' empathic behavior in an OSCE. All authors declare there are no conflicts of interest. No funding was expended for this study.

Acknowledgment

We thank all students for their willingness to participate in the study, all experts for their time and helpful feedback as well as all colleagues and research assistants, especially Clara Wübbolding and Claire Vogel, who helped us to conduct our study. We also thank Thomas Brendel and Thomas Bischoff for the production of the videos and Maximilian Sailer for statistical advise.

References

[1] Zandbelt L. Patient-centred Communication in the Medical Specialist Consultation. Amsterdam: University of Amsterdam,; 2006. https://pure.uva.nl/ws/files/3779195/45880_Zandbelt.pdf.
[2] Feldman-Stewart D, Brundage M, Tishelman C. A conceptual framework for patient–professional communication: an application to the cancer context. Psychooncology 2005;14(10):801–9. https://doi.org/10.1002/pon.950.
[3] Norfolk T, Birdi K, Walsh D. The role of empathy in establishing rapport in the consultation: a new model. Med Educ 2007;41:690–7. https://doi.org/10.1111/j.1365-2923.2007.02789.x.
[4] Lelorain S, Brédart A, Dolbeault S, Sultan S. A systematic review of the associations between empathy measures and patient outcomes in cancer care. Psychooncology 2012;21:1255–64. https://doi.org/10.1002/pon.2115.
[5] Neumann M, Edelhäuser F, Tauschel D, Fischer MR, Wirtz M, Woopen C, et al. Empathy decline and its reasons: a systematic review of studies with medical students and residents. Acad Med 2011;86:996–1009. https://doi.org/10.1097/ACM.0b013e318221e615.
[6] Neumann M, Scheffer C, Tauschel D, Lutz G, Wirtz M, Edelhäuser F. Physician empathy: definition, outcome-relevance and its measurement in patient care and medical education. GMS J Med Educ 2012;29. https://doi.org/10.3205/zma000781.
[7] Elliott R, Bohart AC, Watson JC, Murphy D. Therapist empathy and client outcome: an updated meta-analysis. Psychotherapy 2018;55:399–410. https://doi.org/10.1037/pst0000175.
[8] Ogle J, Bushnell J, Caputi P. Empathy is related to clinical competence in medical care. Med Educ 2013;47:824–31. https://doi.org/10.1111/medu.12232.
[9] Derksen F, Bensing J, Lagro-Janssen A. Effectiveness of empathy in general practice: a systematic review. Br J Gen Pr 2013;63:e76–84. https://doi.org/10.3399/bjgp13X660814.
[10] Shanafelt T, West C, Zhao X. Relationship between increased personal well being and enhanced empathy among internal medicine residents. J Gen Intern Med 2005;20:559–64. https://doi.org/10.1111/j.1525-1497.2005.0108.x.
[11] Frank J. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. Med Teach 2007;29:42–647. https://doi.org/10.1080/01421590701746983.
[12] Duffy FD, Gordon G, Whelan G, Cole-Kelly K, Frankel R. Assessing competence in communication and interpersonal skills: the Kalamazoo II report. Acad Med 2004;79:495–507. https://doi.org/10.1097/00001888-200406000-00002.
[13] Bachmann C, Abramovitch H, Barbu CG, Cavaco AM, Elorza RD, Haak R, et al. A European consensus on learning objectives for a core communication curriculum in health care professions. Pat Educ Couns 2013;93:18–26. https://doi.org/10.1016/j.pec.2012.10.016.
[14] ⟨https://nklm.de/zend/menu⟩ (accessed Jan 24th 2022).
[15] Mercer SW, Reynolds WJ. Empathy and quality of care. Br J Gen Pr 2002;52(Suppl):S9–12.
[16] Morse JM, Anderson G, Bottorff JL, Yonge O, O'Brien B, Solberg SM, et al. Exploring empathy: a conceptual fit for nursing practice. Image J Nurs Sch 1992;24:273–80. https://doi.org/10.1111/j.1547-5069.1992.tb00733.x.
[17] Hemmerdinger JM, Stoddart SD, Lilford RJ. A systematic review of tests of empathy in medicine. BMC Med Educ 2007;7:1–8. https://doi.org/10.1186/1472-6920-7-24.
[18] M.H. Davis, A multidimensional approach to individual differences in empathy. JSAS Catalog of Selected Documents in Psychology 1980;10: 85. https://www.uv.es/~friasnav/Davis_1980.pdf (accessed Jan 24th 2022).
[19] Davis MH. Measuring individual differences in empathy: Evidence for a multidimensional approach. J Pers Soc Psychol 1983;44(1):113–26. https://doi.org/10.1037/0022-3514.44.1.113.
[20] Hojat M, Mangione S, Nasca TJ, Cohen MJ, Gonnella JS, et al. The Jefferson Scale of Physician Empathy: development and preliminary psychometric data. Educ Psychol Meas 2001;61:349–65. https://doi.org/10.1007/s10459-018-9839-9.
[21] Hojat M., Gonnella J.S., Maxwell K. Jefferson Scales of Empathy (JSE). Professional Manual & User's Guide. Philadelphia: Jefferson Medical College. December 2009 edition.
[22] Harden RM. What is an OSCE. Med Teach 1988;10:19–22. https://doi.org/10.3109/01421598809019321.
[23] Setyonugroho W, Kennedy KM, Kropmans TJB. Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review. Pat Educ Couns 2015;98:1482–91. https://doi.org/10.1016/j.pec.2015.06.004.
[24] Cömert M, Zill JM, Christalle E, Dirmaier J, Härter M, Scholl I. Assessing Communication Skills of Medical Students in Objective Structured Clinical Examinations (OSCE) - a systematic review of rating scales. PLoS ONE 2016;11(3):e0152717. https://doi.org/10.1371/journal.pone.0152717.
[25] Mazor KM, Ockene JK, Rogers HJ, Carlin MM, Quirk ME. The relationship between checklist scores on a communication OSCE and analogue patients' perceptions of communication. Adv Health Sci Educ Theory Pr 2005;10:37–51. https://doi.org/10.1007/s10459-004-1790-2.
[26] Hodges B. Creating, Monitoring, and Improving a Psychiatry OSCE. A Guide for Faculty. Acad Psychiatry 2002;26:134–61. https://doi.org/10.1176/appi.ap.26.3.134.
[27] Burt J, Abel G, Elmore N, Campbell J, Roland M, Benson J, Silverman J. Assessing communication quality of consultations in primary care: initial reliability of the Global Consultation Rating Scale, based on the Calgary-Cambridge Guide to the Medical Interview. BMJ Open 2014;4:e004339. https://doi.org/10.1136/bmjopen-2013-004339.
[28] Edgcumbe DB, Silverman J, Benson J. An examination of the validity of EPSCALE using factor analysis. Pat Educ Couns 2012;87:120–4. https://doi.org/10.1016/j.pec.2011.07.01.
[29] Scheffer S, Muehlinghaus I, Froehmel A. OrtweinH. Assessing students' communication skills: validation of a global rating. Adv Health Sci Educ 2008;13:583–92. https://doi.org/10.1007/s10459-007-9074-2.
[30] Roter D, Larson S. The Roter Interaction Analysis System (RIAS): utility and flexibility for analysis of medical interactions. Pat Educ Couns 2002;46:243–51. https://doi.org/10.1016/s0738-3991(02)00012-5.
[31] Zimmermann C, Del Piccolo L, Bensing J, Bergvik S, De Haes H, Eide H, et al. Coding patient emotional cues and concerns in medical consultations: the Verona coding definitions of emotional sequences (VR- CoDES). Pat Educ Couns 2011;82:141–8. https://doi.org/10.1016/j.pec.2010.03.017.
[32] Zhou Y, Collinson A, Laidlaw A, Humphris G. How do medical students respond to emotional cues and concerns expressed by simulated patients during OSCE consultations – a multilevel study. PLOS ONE 2013;8:e79166. https://doi.org/10.1371/journal.pone.0079166.
[33] Bloom BS, Krathwohl DR, Masia BB. Bloom Taxonomy of Educational Objectives. New York: Longman,; 1984.
[34] Krathwohl DR. A revision of Bloom's taxonomy: An overview. Theory into Pract 2002;41:212–8. https://doi.org/10.1207/s15430421tip4104_2.
[35] Hargie O. Skill in practice: An operational model of communicative performance. In: Hargie O, editor. The Handbook of Communication Skills. London: Routledge; 2018. p41–77.
[36] Hulsman RL, Mollema ED, Hoos AM, de Haes H, Donnison-Speijer JD. Assessment of medical communication skills by computer: assessment method and student experiences. Med Educ 2004;38:813–24. https://doi.org/10.1111/j.1365-2929.2004.01900.x.
[37] Karabilgin OS, Vatansever K, Caliskan SA, Durak H. Assessing medical student competency in communication in the pre-clinical phase: objective structured video exam and SP exam. Pat Educ Couns 2012;87:293–9. https://doi.org/10.1016/j.pec.2011.10.008.
[38] Baribeau DA, Mukovozov I, Sabljic T, Eva KW, Delottinville DB. Using an objective structured video exam to identify differential understanding of aspects of communication skills. Med Teach 2012;34:e242–50. https://doi.org/10.3109/0142159X.2012.660213.
[39] Humphris GM, Kaney S. Assessing the development of communication skills in undergraduate medical students. Med Educ 2001;35:225–31. https://doi.org/10.1046/j.1365-2923.2001.00869.x.
[40] Patterson F., Ashworth V., Zibarras L., Coan P., Kerrin M., O'Neil P. Evaluations of situational judgement tests to assess non-academic attributes in selection. Med Educ 2012; 46:850–868. Patterson F, Ashworth V, Zibarras L, Coan P, Kerrin M, O'Neil P. Evaluations of situational judgement tests to assess non-academic attributes in selection. Med Educ 2012; 46:850–868.
[41] Lievens F, Peeters H, Schollaert E. Situational judgment tests: A review of recent research. Pers Rev 2008;37:426–41. https://doi.org/10.1108/00483480810877598.
[42] Cullen MJ, Zhang C, Marcus-Blank B, Braman JP, Tiryaki E, Konia M, et al. Improving Our Ability to Predict Resident Applicant Performance: Validity Evidence for a Situational Judgment Test. Teach Learn Med 2020;32:508–21. https://doi.org/10.1080/10401334.2020.1760104.

[43] Graupe T, Fischer MR, Strijbos JW, Kiessling C. Development and piloting of a Situational Judgement Test for emotion-handling skills using the Verona Coding Definitions of Emotional Sequences (VR-CoDES). Pat Educ Couns 2020;103:1839–45. https://doi.org/10.1016/j.pec.2020.04.001.
[44] Motowidlo SJ, Hooper AC, Jackson HL. Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. J Appl Psych 2006;91:749–61. https://doi.org/10.1037/0021-9010.91.4.749.
[45] McDaniel MA, Nguyen NT. Situational judgment tests: A review of practice and constructs assessed. Int J Sel Assess 2001;9:103–13. https://doi.org/10.1111/1468-2389.00167.
[46] McDaniel MA, Hartman NS, Whetzel DL, Grubb III WL. Situational judgment tests, response instructions, and validity: A meta-analysis. Pers Psychol 2007;60:63–91.
[47] Christian MS, Edwards BD, Bradley JC. Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. Pers Psych 2010;63:83–117. https://doi.org/10.1111/j.1744-6570.2009.01163.x.
[48] Patterson F, Rowett E, Hale R, Grant M, Roberts C, Cousans F, Martin S. The predictive validity of a situational judgement test and multiple-mini interview for entry into postgraduate training in Australia. BMC Med Educ 2016;16:1–8. https://doi.org/10.1186/s12909-016-0606-4.
[49] Lievens F, Patterson F. The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. J Appl Psychol 2011;96:927–40. https://doi.org/10.1037/a0023496.
[50] Chan D, Schmitt N. Situational judgment and job performance. Hum Perform 2002;15:233–54. https://doi.org/10.1207/S15327043HUP1503_01.
[51] Van den Eertwegh V, van Dulmen S, van Dalen J, Scherpbier AJ, van der Vleuten CP. Learning in context: Identifying gaps in research on the transfer of medical communication skills to the clinical workplace. Pat Educ Couns 2013;90:184–92. https://doi.org/10.1016/j.pec.2012.06.008.
[52] Kinderman P, Humphris G. Clinical communication skills teaching: the role of cognitive schemata. Med Educ 1995;29:436–42. https://doi.org/10.1111/j.1365-2923.1995.tb02868.x.
[53] Berg K, Majdan JF, Berg D, Veloski J, Hojat M. A comparison of medical students' selfreported empathy with simulated patients' assessments of the students' empathy. Med Teach 2011;33:388–91. https://doi.org/10.3109/0142159X.2010.530319.
[54] Blanco JM, Caballero F, García FJ, Lorenzo F, Monge D. Validation of the Jefferson Scale of Physician Empathy in Spanish medical students who participated in an Early Clerkship Immersion programme. BMC Med Educ 2018;18:209. https://doi.org/10.1186/s12909-018-1309-9.
[55] Hojat M, Gonnella JS, Mangione S, Nasca TJ, Veloski JJ, Erdmann JB, et al. Empathy in medical students as related to academic performance, clinical competence and gender. Med Educ 2002;36:522–7. https://doi.org/10.1046/j.1365-2923.2002.01234.
[56] Glaser KM, Markham FW, Adler HM, McManus RP, Hojat M. Relationships between scores on the Jefferson Scale of physician empathy, patient perceptions of physician empathy, and humanistic approaches to patient care: A validity study. Med Sci Monit 2007;13. CR291–294.
[57] Giemsa P, Wübbolding C, Fischer MR, Graupe T, Härtl A, Lenz A, et al. What works best in a general practice specific OSCE for medical students: Mini-CEX or content-related checklists? Med Teach 2020;42:578–84. https://doi.org/10.1080/0142159X.2020.1721449.
[58] Preusche I, Wagner-Menghin M. Rising to the challenge: cross-cultural adaptation and psychometric evaluation of the adapted German version of the Jefferson Scale of Physician Empathy for Students (JSPE-S)-. Adv Health Sci Educ Theory Pr 2012;18:573–87. https://doi.org/10.1007/s10459-012-9393-9.
[59] De Corte K, Buysse A, Verhofstadt LL, Roeyers H, Ponnet K, Davis MH. Measuring empathic tendencies: reliability and validity of the Dutsch version of the interpersonal reacitivity index. Psychol Belg 2007;47:235–60. https://doi.org/10.5334/pb-47-4-235.
[60] Calvi JL. The Relationship Between Self-report and Behavioural Measures of Empathy. Denton Texas: University of North Texas,; 2011. https://shareok.org/bitstream/handle/11244/9406/Calvi_okstate_0664M_11846.pdf?sequence=1.
[61] Fernández AM, Dufey M, Kramp U. Testing the Psychometric Properties of the Interpersonal Reactivity Index (IRI) in Chile. Empathy in a Different Cultural Context. Eur J Psychol Assess 2011;27:179–85. https://doi.org/10.1027/1015-5759/a000065.
[62] Tousignant M, DesMarchais JE. Accuracy of student self-assessment ability compared to their own performance in a problem-based learning medical program: a correlation study. Adv Health Sci Educ 2002;7:19–27. https://doi.org/10.1023/a:1014516206120.
[63] Carr SE, Johnson PH. Does self reflection and insight correlate with academic performance in medical students? BMC Med Educ 2013;13:113. https://doi.org/10.1186/1472-6920-13-113.
[64] Blanch-Hartigan D. Medical students' self-assessment of performance: Results from three meta-analyses. Pat Educ Couns 2011;84:3–9. https://doi.org/10.1016/j.pec.2010.06.037.
[65] Hojat M, Gonnella JS, Nasca TJ, Mangione S, Vergare M, Magee M. Physician empathy: definition, components, measurement, and relationship to gender and specialty. Am J Psychiatry 2002;159:1563–9. https://doi.org/10.1176/appi.ajp.159.9.1563.
[66] Hojat M, Gonnella JS, Mangione S, Nasca TJ, Veloski JJ, Erdmann JB, et al. Empathy in medical students as related to academic performance, clinical competence and gender. Med Educ 2002;36:522–7. https://doi.org/10.1046/j.1365-2923.2002.01234.x.
[67] Cullen M., Zhang C., Sackett P., Thakker K., Young J.Q. Can a Situational Judgment Test Identify Trainees at Risk of Professionalism Issues? A Multi-Institutional, Prospective Cohort Study. Acad Med 2022 May 24. Online ahead of print.
[68] Lievens F, Sacket PR. The Validity of Interpersonal Skills Assessment Via Situational Judgment Tests for Predicting Academic Success and Job Performance. J Appl Psychol 2012;97:460–8. https://doi.org/10.1037/a0025741.
[69] M. Hojat, Empathy in health professions education and patient care, NY: Springer, New York, 2016. https://doi.org/10.1007/978–3-319–27625-0.
[70] Berg K, Majdan JF, Berg D, Veloski J, Hojat M. Medical students' self-reported empathy and simulated patients' assessments of student empathy: an analysis by gender and ethnicity. Acad Med 2011;86:984–8. https://doi.org/10.1097/ACM.0b013e3182224f1f.
[71] Humphris GM. Communication skills knowledge, understanding and OSCE performance in medical trainees: a multivariate prospective study using structural equation modelling. Med Educ 2002;36:842–52. https://doi.org/10.1046/j.1365-2923.2002.01295.x.
[72] Schuwirth L, van der Vleuten C, Durning SJ. What programmatic assessment in medical education can learn from healthcare. Perspect Med Educ 2017;6:211–5. https://doi.org/10.1007/s40037-017-0345-1.
[73] Van der Vleuten CPM, Schuwirth LWT, Scheele F, et al. The assessment of professional competence: building blocks for theory development. Best Pract Res Clin Obstet Gynaecol 2010;24:703–19. https://doi.org/10.1016/j.bpobgyn.2010.04.001.