Empathy as a selection criterion for medical students: is a valid assessment possible during personal interviews? A mixed-methods study

Thomas Kötter1,2 · Johanna Christine Schulz2 · Nadine Janis Pohontsch3

Received: 15 September 2020 / Accepted: 7 May 2022 / Published online: 20 May 2022 © The Author(s) 2022

Abstract
Places to study at medical schools are scarce, which makes well-designed selection procedures employing criteria with predictive validity for good students and doctors necessary. In Germany, the pre-university grade point average (pu-GPA) is the main selection criterion for medical school application. However, this is criticised. According to a decision by the Federal Constitutional Court, selection must be supplemented with a criterion other than the pu-GPA. Empathy is a core competency in medical care. Therefore, it seems to be an appropriate criterion. This study evaluates the feasibility of an empathy questionnaire and empathy appraisal by a panel for applicant selection. We employed a sequential explanatory mixed-methods design. Results of self- and external assessments of empathy were compared in a quantitative analysis. Thereafter, the concept of empathy and the approach to empathy appraisal by the selection panel members were explored qualitatively in six focus groups with 19 selection panel members using a semi-structured guideline. Transcripts were content analysed using both deductive and inductive coding. We found no significant correlation of self- and external empathy assessment ($\rho(212) = -.031$, $p > .05$). The results of the focus groups showed that, while panel members judged the external empathy assessment to be useful, they had neither a homogenous concept of empathy nor an implicit basis for this assessment. This diversity in panel members’ concepts of empathy and differences in the concepts underlying the Davis Interpersonal Reactivity Index seem to be the main reasons for the lack of correlation between self- and external empathy assessments. While empathy is a possible amendment to established selection criteria for medical education in Germany, its external assessment should not be employed without training panel members based on an established theoretical concept of empathy and an objective self-assessment measure.

Keywords Empathy · Selection · Student · Professional competence · Education · Medical · Schools · Medical · Research · Qualitative

Extended author information available on the last page of the article
1 Introduction

In Germany, there are about five applicants per place at medical schools each year (Foundation for Higher Education Admission, 2018). This high popularity of medical education is seen in many other countries as well. In Germany, medical education at public medical schools is free for the learners but very cost-intensive for the government. Thus, responsible government agencies, as well as medical schools, must develop suitable selection procedures. These procedures should be based on criteria with predictive validity of which graduates are most likely to become the best students and, even more importantly, the best doctors. There is, however, not a single commonly agreed upon definition of a good medical student/a good doctor.

To date, the main selection criterion for medical education in Germany is the pre-university (high school) grade point average (pu-GPA; ranging from 0.7 [best] to 4.0 [worst]) (Federal Ministry of Justice & Consumer Protection, 2018). According to a decision by the Federal Constitutional Court (special court reviewing judicial and administrative decisions/legislation for compliance with the German constitution), the selection must be supplemented with at least one criterion that is independent of the pu-GPA (Federal Constitutional Court, 2017). The revision of the selection process thus provided an opportunity to introduce new criteria that cover, in addition to academic capability, (inter)personal skills as well. This is in line with international initiatives and regulations: e.g. Ireland, Australia and New Zealand use non-academic attributes to complement secondary grades as a selection criterion for medical students (O’Sullivan et al., 2017).

One core quality of both a good medical student and doctor appearing in several respective frameworks is empathy. Empathy is a core component of emotional intelligence, and a general consensus among medical education leaders and professional organisations exists on the high importance of empathy for success in medical education and patient care (Hojat et al., 2011). There is evidence for a positive influence of empathy on Objective Structured Clinical Examinations (OSCE) communication scores (Casas et al., 2017). Several studies show a positive influence of doctors’ empathy on health outcomes in patients (Coulahan et al., 2001; Beckman et al., 1994). Although there is no single definition of empathy (Jeffrey, 1994), different stakeholder groups, such as doctors, patients and students themselves, rate ‘interpersonal qualities’ containing empathy consistently as an important feature of good doctors (Steiner-Hofbauer et al., 2018). Mercer and Reynolds define empathy in the context of healthcare as a complex, multidimensional construct that includes understanding the patient, reflecting your understanding, checking whether you understood the patient correctly and acting upon that understanding in a therapeutic way (Mercer & Reynolds, 2002).

The question as to whether empathy is a stable personality trait or an evolving and changing ability with cognitive and emotional aspects is discussed controversially. We cannot be sure whether empathy scores are stable, increasing or declining during medical education, and there is some criticism on self-report measures to assess empathy in medical students and candidates (Colliver et al., 2010; Costa...
et al., 2013; Ferreira-Valente et al., 2017). Notwithstanding the final outcome of this discussion, it could be worthy of consideration to use empathy as a selection criterion for medical school. If empathy emerges as a trait, it would be preferable to select applicants with high levels of empathy, but also if empathy emerges as evolving and trainable, it would be more worthwhile to have students with high empathy levels to start from to reach even higher levels of empathy.

As long as personal interviews are part of the selection process, the empathy of applicants can be assessed externally. Another widely used possibility would be an empathy self-assessment during the selection procedure. For that, different standardised questionnaires are available, such as the Jefferson Scale of Empathy (JSPE) (Hojat et al., 2002) and the Davis Interpersonal Reactivity Index (IRI) (Davis, 1983). For both instruments, validated German-language versions are available (Neumann et al., 2012) and both have been used widely to assess medical students’ empathy (Quince et al., 2016). None of the two, however, has been validated for selection purposes (Hemmerdinger et al., 2007). Furthermore, self-assessments bear the risk of socially desired answering, especially in selection situations (Edwards, 1957). However, the same problem may occur during selection interviews. Hence, it is until now unclear whether self- or external assessment of empathy is a or the more feasible method in the context of medical student selection.

We therefore aimed to compare the results of two methods of empathy appraisal during the selection process for medical school: self-assessment and external assessment by selection panel members. Furthermore, we aimed to shed light on how the discrepancies between self-assessment and external assessment of empathy can be understood and explained. Therefore, we explored the empathy concepts of selection panel members and compared the results with the empathy concept the self-assessment tool (IRI) is based on.

2 Methods

We conducted a cross-sectional mixed-methods study (sequential explanatory design) (Creswell & Clark, 2017; O’Cathain et al., 2010) with a quantitative study phase conducted in 2016, which resulted in a qualitative study phase to further explore and explain the quantitative findings in 2017.

2.1 Setting

The study was conducted at Lübeck Medical School (LMS), a section of the public University of Lübeck, Germany. About 1,500 students are enrolled in the medical study programme and about 185 freshmen (high school students, in part with a completed vocational training or first university degree) are admitted to LMS each year. Of these, until 2019, 65 (35%) were allocated by a public agency based on pu-GPA or waiting time; 120 (65%) were selected from 240 direct applicants who did not meet the competitive criteria via an internal selection procedure (Auswahlverfahren der Hochschulen, AdH).
The core of the AdH at LMS was a 30-min interview led by two faculty members and one student (selection panel). Selection panel members participated in a mandatory briefing and new selection panel members were encouraged to participate in interview training (Mommert et al., 2020). For the interviews, the selection panel members were provided a standardised interview guide that gave examples of situational and behavioural questions. Five primary dimensions had to be rated by selection panel members: motivation, knowledge about the course of study, social engagement, (self-)reflection and communication. Each dimension contained five items that were rated on a 5-point Likert scale from 0 to 4. The interviews took part each year in the middle of August on the campus of LMS. We collected all data for this study in August 2016 (quantitative data) and 2017 (qualitative data) on site.

2.2 Quantitative study

2.2.1 Participant selection and recruitment

For the survey, we investigated applicants for LMS who participated in the AdH at LMS in August 2016. Overall, 240 applicants were invited to the interviews in 2016, with 228 who accepted the invitation and were asked to take part in the survey (no exclusion criteria). We offered a voucher (5 €) as an incentive for participation.

2.2.2 Data collection

The self-assessment part of the survey was web-based (using SurveyMonkey; Survey Monkey Europe UC, Dublin, Ireland) and was conducted in a computer pool close to the interview location directly after the interview.

For this study, we collected data on age, gender, pu-GPA and a self-assessment of empathy using the German version of the IRI (Neumann et al., 2012). Empathy is measured through four facets covered by four items each on a 5-point Likert scale from 0 = ‘does not describe me well’ to 4 = ‘describes me very well’:

1. Perspective taking (ability to adopt another’s point of view)
2. Fantasy (ability to empathise with fictional characters)
3. Empathic concern (ability for other-oriented emotions, such as pity)
4. Personal distress (self-oriented emotions such as uneasiness, which may occur in close or problematic interpersonal interactions)

Being a self-oriented emotion, personal distress is not considered in the empathy score.

The basic psychometric quality of the German version of the IRI is comparable to the original version (acceptable Cronbach’s alpha for the four subscales; Neumann et al., 2012).

In addition to surveying all applicants, we asked the selection panel members to rate the empathy of each applicant at the end of the interview in addition to the other dimensions mentioned above. 2016 was the first year in which a global empathy
rating was used. Empathy was rated on a 5-point Likert scale (0 = not empathic; 4 = utmost empathic) without further introductory text except the request to rate the applicants’ empathy. For this study, we also extracted the score for overall impression (rated by the selection panel on a 5-point Likert scale from 0 to 4).

2.2.3 Data handling

Data from the web survey was imported into and analysed using IBM SPSS Statistics for Windows Version 22.0 (IBM Corp., Armonk, NY, USA). After a plausibility check (e.g. looking for implausible age and pu-GPA information), data from the interview score sheets was matched to the web survey data using consecutive numbers assigned to the participants.

2.2.4 Data analysis

We analysed the data using descriptive statistics. For gender, we calculated percentages, and for continuous variables, we calculated means (M) and standard deviations (SD). We used t tests to compare the means of continuous variables. In order to express bivariate correlation between the empathy self-assessment (IRI sum score) and the empathy assessment by the selection panel members (external assessment), we used Spearman’s ρ. Effect sizes are reported using Cohen’s d. We considered values of <0.30 small, ≥0.30 medium and ≥0.50 large effect sizes. All statistical tests were performed two-tailed with an alpha of 0.05.

2.3 Qualitative study

Quantitative data showed a non-correlation of externally and self-assessed empathy, and the high correlation of externally assessed empathy and the score for overall impression (see Sect. 2.3.2). To follow this thread in the quantitative data and find an explanation for this somehow rather surprising finding, a qualitative study to further explore and explain quantitative results was deemed necessary. As the qualitative study was not planned in advance, a 1-year time lag lies between the quantitative and qualitative part of our mixed-methods study.

2.3.1 Participant selection and recruitment

All (n = 48) selection panel members who participated in the AdH in August 2017 (n = 24 faculty members and n = 24 students; n = 23 female and n = 25 male) were eligible for the focus groups. We invited all participants of the mandatory briefing for the selection procedure in July 2017 during a short oral presentation of the study to attend the focus groups. We did not invite the non-attending selection panel members, because we could not give the short oral presentation on our project to them. The briefing session was attended by 30 selection panel members. We offered a voucher (10 €) as an incentive. Participants received (oral and written) information,
could ask further questions and gave their informed consent for the focus groups to be recorded and transcribed verbatim and the results to be published anonymously.

2.3.2 Data collection

For our qualitative study, we (TK, JCS and NJP) developed a focus group topic guide (Table 1) including the following topics:

- Short introduction of interviewer and study
- Subjective definition and meaning of empathy in the clinical setting

| Table 1 | Focus group topic guide |
|---------|-------------------------|
| Introduction: |
| Thank you for taking the time for this focus group |
| All of you are here because you were members of the selection panel for the future members of medical school last year |
| You may recall that last year, in addition to the usual criteria, you made an assessment of empathy |
| In this focus group, we now want to talk about how you assessed empathy at that time |
| I would be pleased if you could write down what empathy is for you and then discuss it in the focus group |
| Concept of empathy: |
| What does empathy mean for you? |
| Does anyone count other aspects of empathy? |
| What do you think empathy is about beyond that? |
| Does anyone of you know a person who seems to be especially empathic? Can you describe this person? |
| What makes this person special? Would any of you describe another person as especially empathic? |
| When you hear the term empathy, what is the first thing you think of? |
| What concepts do you think resemble empathy? |
| Differentiation of empathy and sympathy/overall impression: |
| How do you distinguish between sympathy and empathy? |
| How do you distinguish between the overall impression and empathy? |
| Do you think that someone who has broad knowledge is also particularly empathetic? |
| Do you believe that someone who has excellent communication skills is also highly empathetic? |
| Learnability of empathy during medical education: |
| Do you think that empathy is learnable during medical education? |
| Assessment process: |
| Individual basis for the empathy assessment: |
| What criteria do you use to make your assessment? |
| When an applicant stood in front of you, by which characteristics did you recognise their ability to empathise? |
| When you placed the checkmark, what did you have in mind? |
| What did a candidate have to do to get a good rating? |
| How do you distinguish an applicant who is given 5 points in the assessment from one who is given fewer points? |
| Are the other assessments, such as knowledge of the university or the overall impression, also decisive for how you allocate points? |
| Have you asked certain questions to test empathy or would you do so in the future? |
| What do you think would be the best way to test empathy? |
| Evaluation of the usefulness of empathy appraisal during the selection interviews: |
| Do you find the assessment of empathy in the selection process useful at all? |
Differentiation of empathy and sympathy/overall impression
• Learnability of empathy during medical education
• Individual basis for the empathy assessment
• Evaluation of the usefulness of empathy appraisal during the selection interviews

The focus topic guide was pilot tested during two preliminary interviews and two preliminary focus groups (n = 7 participants) with physicians not acting as selection panel members and selection panel members not available during the actual data collection period. During this development phase, it was found to be suitable to subdivide the topic guide into three parts.

All focus groups were conducted in German. The focus groups took part in seminar rooms with no one present besides the participants and researchers. We used digital audio recording to collect the data. In addition, both facilitators made field notes during the focus groups. All focus groups were transcribed verbatim by JCS and a trained research assistant following designated transcription rules (Mayring, 2000). Interviews were anonymised during transcription. To facilitate a distinction during data analysis, faculty members were given the letter ‘P’ and students the letter ‘S’, followed by a consecutive number. The interviewers were marked ‘I1’ and ‘I2’. Transcripts were checked for accuracy by TK. We did not return the transcript to the focus group participants as this does not seem to be the usual procedure in studies using focus groups and qualitative content analysis and would have meant an unduly demand from the participants.

All focus groups were moderated by both TK and JCS. TK is a male MD who is board certified in Family Medicine, MSc Public Health, and qualified as a professor, with extensive experience in the field of medical curriculum research and in conducting focus groups (Kötter et al., 2015, 2020; Kötter, Carmienke, et al., 2014; Kötter, Ritter, et al., 2016; Kötter, Tautphäus, et al., 2014, 2016). JCS, a female, was at the time of the focus groups a psychology student enrolled in the bachelor program of Lübeck University. She was new to facilitating focus groups, but received a detailed briefing from TK. NJP is a female trained psychologist with comprehensive experience in conducting focus groups and interviews, as well as qualitative data analysis (Pohontsch et al., 2017; Pohontsch, Hansen, et al., 2018; Pohontsch, Stark, et al., 2018; Pohontsch, Zimmermann, et al., 2018), holds a PhD degree and works as a postdoctoral researcher.

2.3.3 Data handling

The qualitative data was managed using QCAmap (http://www.qcamap.org/), an open access web application for systematic text analysis in scientific projects based on the techniques of qualitative content analysis. All transcripts were uploaded to the QCAmap account and were accessible for both coders (TK and JCS).
2.3.4 Data analysis

TK, JCS and NJP analysed the qualitative data conjointly using structuring qualitative content analysis (Hsieh & Shannon, 2005). This systematic procedure is used to reduce large amounts of data while preserving and extracting the main content. Deductive categories were derived from the research question and the focus group topic guide. All other categories were built inductively using summarising content analysis while processing the material (Schreier, 2014). The material was read several times by both TK and JCS before coding to ensure familiarity. For coding, transcripts were broken down into fragments of analysis, which could adopt different sizes. Fragment size ranged from part of a sentence to one or more paragraphs depending on the amount of data needed to understand the content and context of the respective fragment.

JCS conducted the inductive coding and theme development process in close consultation with TK. Categories and codes were described in code memos comprising coding rule and typical quotes. The category system was developed in several discussion rounds. In addition to TK and JCS, NJP was involved in the analysis of the qualitative data and interpretation of the results. As findings were summarised over the whole group of participants, participant checking of findings might have been more trouble than worth, especially with respect to the expenditure of the participants’ time and ability to abstract from their own accounts to findings for the whole group. We therefore chose to ensure intersubjective reproducibility and comprehensibility by discussing the results within our interdisciplinary workgroup.

2.4 Reporting

This report was written under consideration of the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) criteria (von Elm et al., 2008), the Consolidated criteria for reporting qualitative research (COREQ; Tong et al., 2007) and the Good Reporting of A Mixed Methods Study (GRAMMS) guideline (O’Cathain et al., 2008). Citations used in this article were translated by TK and double-checked by JCS. Citations are marked with quotation marks. […] marks omissions or amendments in a citation.

3 Results

3.1 Quantitative study

3.1.1 Participants

After the exclusion of incomplete data sets, a total of 214 empathy self-assessments could be matched to external assessments. Fourteen applicants either did not give written informed consent (participation was voluntary, reasons for non-participation
were not collected) or did not fill out the questionnaire completely (response rate: 94%). Of the included individuals, 73% were female. The mean age was 20.7 years ($SD=2.2$). The mean pu-GPA was 1.5 ($SD=0.24$). There was no missing data for these variables.

### 3.1.2 Outcomes

The mean empathy score (self-assessment) was 48.12 ($SD=5.33$). We observed no statistically significant differences between male and female applicants ($t(212)=1.71, p=0.09$). The mean empathy score (external assessment) was 3.80 ($SD=0.86$). Female applicants ($M=3.88$) scored significantly higher when compared to male applicants ($M=3.55$; $t(212)=2.55, p<0.01; r=0.19$).

### 3.1.3 Correlations

We observed no significant associations between the self- and external assessments of empathy ($\rho(212)=-0.031, p>0.05$). The external assessment of empathy did not correlate with any of the subscales of the self-assessment instrument (IRI). The external assessment of empathy showed, however, a strong significant correlation with the rating of overall impression ($\rho(212)=0.697, p<0.01$).

### 3.2 Qualitative study

#### 3.2.1 Participants

We conducted six focus groups with two to five selection panel members (overall $n=19$, $n=10$ students and $n=9$ faculty members; see Table 2 for participant characteristics). The main reason for non-participation was time schedule restrictions. The mean duration of the focus groups was 58 min ($R=54–62$ min).

#### 3.2.2 Main categories

Considering the subjective views of the questioned selection panel members, three main categories, ‘concept of empathy’, ‘distinction from sympathy’ and ‘learnability’, could be identified. Considering the assessment of empathy, two main categories, ‘basis of assessment’ and ‘usefulness’, could be identified (see Table 3 for an overview of categories).

| Table 2 | Characteristics of focus group participants |
|---------|---------------------------------------------|
| Sex     | Students | Faculty members |
| Female  | 6        | 4               |
| Male    | 4        | 5               |
| Median age (range) | 24 (22–27) | 44 (34–63) |
3.2.3 Concept of empathy

Overall, the concept of empathy was found to be very heterogeneous between the interviewed selection panel members. Doctors provided more examples from their day-to-day life when compared to students, who more often referred to communicative aspects of empathy.

Five subcategories (see Table 3) were identified when looking at the concept of empathy of the interviewed selection panel members: sensitivity, being able to put oneself in somebody’s position, helpfulness, individualised communication and active action.

Sensitivity was mentioned as one of the basic mechanisms of empathy. Resonating with your counterpart and being able to feel someone’s feelings were mentioned as key aspects of empathy. Either way, there was disagreement as to whether and how to differentiate empathy from compassion.

‘I would describe empathy basically as sensitivity and have the opinion, that it is essentially the ability to perceive the emotions of the vis-à-vis’. (Faculty member)

‘Empathy doesn’t mean to live through the emotions’. (Student)

‘Sympathise, but not actively commiserate, that is empathy for me’. (Faculty member)

‘And then empathise, commiserate, resonate with the counterpart. I can’t differentiate compassion so clearly’. (Faculty member)
The most important aspect of empathy from the selection panel members was the ability to put oneself in somebody’s position, to be aware of someone else’s feelings, mood and need for help. Adequate assessment of someone’s position and neediness was also relevant.

‘Empathy is the ability to put oneself into the emotional state of the other person’. (Faculty member)
‘For me, compassion means being able to put oneself in the position of one’s fellow human being, as if one were also affected, whereby this ‘as if’ is very important’. (Faculty member)
‘It is about understanding the other who has problems or open questions or conflicts. That one correctly takes in what has been said’. (Faculty member)
‘Not only have compassion but somehow understand it properly’. (Student)
‘Empathy also means being able to recognise the mood or the need for help of others’. (Student)

The concept of individualised communication includes the ability to fit someone’s behaviour and communication style to the counterpart’s needs, showing understanding and keeping communication on equal terms and appreciative. Being authentic, showing social competencies and appropriate physical closeness (touching) were also mentioned.

‘That you can recognise something and also respond to it without being told directly’. (Student)
‘Essentially, an empathetic person is characterised by the feeling that he truly understands what I mean’. (Faculty member)
‘Being empathetic means showing the other person that you understand and accept him or her’. (Student)
‘Empathetic can also have a tactile form. It can also mean to give someone a hug’. (Faculty member)

Helpfulness or the motivation to help others was seen as another basic component of empathy. The motivation to invest time and energy is a prerequisite to be able to be empathetic.

‘I have to…pull myself together, and that really costs energy, to be adequately empathetic’. (Faculty member)

Selection panel members emphasised that empathy does not end with (non-)verbal communication, and it also includes acting upon the information gained and active conveying of safety.

‘Empathy is not only…you feel bad, I feel with you now, but it is also acting’. (Faculty member)

3.2.4 Distinction to other concepts

Two subcategories (see Table 3) were identified when exploring the distinction of empathy to other concepts, in this case sympathy. Most of the interviewed selection
panel members could not define a clear demarcation between empathy and sympathy/overall impression, but instead named many differences. One student, for example, referred to the learnability:

‘It [sympathy], in contrast to empathy, isn’t learnable’. (Student)

Others saw the difference in the fact that sympathy just happens, whereas empathy is an active process. In the assessment process itself, the interviewed selection panel members had difficulties in differentiating between empathy and sympathy/overall impression, as well:

‘I actually think that for the question whether I found someone to be empathic, sympathy had an influence. I would thus not say that I was able to differentiate this’. (Student)

‘I did not assess someone as overall good but not empathic at the same time. That did not occur. Either she or he was good, then they were also empathic, or not’. (Student)

### 3.2.5 Learnability

Three subcategories (see Table 3) were identified when asking the interviewed selection panel members about the learnability of empathy during medical education.

The interviewed selection panel members see empathy as only limitedly learnable during medical education:

‘My observation is, that in the age of 15 to 16 years, certain personality traits [like empathy] are already fixed.’ (Faculty member)

‘Empathy: you have it or you don’t, that is to some extent like character. Whether you can teach it…I would say, per se, that is difficult’. (Faculty member)

On the other hand, the interviewed selection panel members noticed that there are factors, such as time pressure and stress, that could have an influence on empathy:

‘We can reduce empathy superbly [through stress] in the preclinical phase’. (Faculty member).

### 3.2.6 Basis of assessment

Five subcategories (see Table 3) were identified when assessing the basis of assessment of empathy of the interviewed selection panel members.

The assessment seldom was based on specific questions:

‘I don’t think that one can assess empathy based on a question’. (Faculty member)

For the selection panel members, an authentic presentation that contained both good non-verbal and verbal communication was paramount. Hence, they judged gestures, facial expressions and how the applicants reacted to the selection panel:
‘Gestures and facial expression meant a lot to me’. (Student)

The students found it especially difficult to put in words on what their empathy assessment was based. Seven of 10 students referred to a ‘gut feeling’ in this context.

Doctors, on the contrary, based their assessment on stories told by the applicants:

‘And whether someone is able to be empathic, I try to find out by his stories’. (Faculty member)

### 3.2.7 Usefulness

Two subcategories (see Table 3) were identified when letting the interviewed selection panel members evaluate the usefulness of empathy assessment during the selection process.

Most of the interviewees judged empathy assessment as part of the selection process as useful:

‘I find that significantly more useful to include when compared to asking how the applicant differentiates science from being a doctor’. (Faculty member)

They demanded tools for empathy assessment, although:

‘I think that [empathy] is an important quality, but one has to find a tool [for the assessment]’. (Student)

### 3.2.8 Comparison of the empathy concepts

When comparing the IRI with the empathy concepts of the selection panel members, we found that the two IRI subscales ‘Perspective taking’ (ability to adopt another’s point of view) and ‘Empathic concern’ (ability for other-oriented emotions) found their equivalents in the answers of the interviewees. However, both the facets ‘Personal distress’ (self-oriented emotions that may occur in close or problematic interpersonal interactions) and ‘Fantasy’ (ability to empathise with fictional characters), as the two other subscales, were not present in the empathy concepts of the panel members. The empathy concepts of panel members more closely resembled the empathy concept of Mercer and Reynolds (Mercer & Reynolds, 2002) that is used in the medical context. Figure 1 gives an overview how the selection panel members’ definitions of empathy correspond to the concepts proposed by Mercer and Reynolds (Mercer & Reynolds, 2002).

### 4 Discussion

In our mixed-methods, cross-sectional study investigating empathy as a selection criterion for medical school admission, we found no correlation between self- and external assessments of empathy. According to our data, no common concept for empathy among the selection panel members and discrepancies to the concept underlying the IRI seem to be the main reasons for this finding.
The difference between self- and external assessments of empathy is, to our knowledge, not a subject of scientific studies. In more general papers on this topic (Bogner & Landrock, 2016; Krüger, 1980; Mummendey & Grau, 2014) and a paper describing the comparison of an interpersonal skills appraisal and JSPE scores in the context of medical school admission (O’Sullivan et al., 2017), authors concluded that social desirability may be responsible for such differences. However, social desirability is not limited to self-assessment contexts. Self-assessment of empathy might also be biased by the Dunning-Kruger effect leading unexperienced individuals to overestimate their abilities (here their empathy) or the imposter syndrome (leading individual to underestimate their abilities) (Kruger & Dunning, 1999; Langford & Clance, 1993). Another effect that could have contributed to the lack of correlation between self- and external assessments is the overall very good assessment of empathy by the selection panel members (leniency error) (Hui & Triandis, 1985). An even greater influence may be the halo effect (Thorndike, 1920): in the absence of a concept or rationale for the assessment of empathy, other attributes of the applicants, such as knowledge, communication skills, attractiveness or overall impression, may have played a role in it.

The results of the qualitative part of our study indicate that the lack of a (common) concept of empathy and its appraisal among the selection panel members may have been the main reason for the missing correlation between self- and external

Fig. 1 Participants’ concept of empathy in light of the concept proposed by Mercer & Reynolds (2002)
assessments. However, selection panel members judged empathy as an important concept in the context of selection for medical education. Most of them already had a subjective concept of empathy. The concepts of the different panel members differed, though. In addition, our study showed that the panel members had difficulties in translating their empathy concept into an assessment and differentiating empathy from other competencies. In light of the fact that no consistent concept of empathy exists even among experts and even those have difficulties differentiating between empathy, sympathy and other communication skills (Dohrenwend, 2018), this is rather unsurprising.

We found that the empathy concepts of the panel members only covered two of the IRI subscales (i.e. ‘Perspective taking’ [ability to adopt another’s point of view] and ‘Empathic concern’ [ability for other-oriented emotions]). Yamada and colleagues argued that these two subscales might even be capturing sympathy rather than empathy (Yamada et al., 2018), which may explain why these facets were not covered by the empathy concepts of the interviewees. Even with ‘Perspective taking’ and ‘Empathic concerns’ being present in the empathy concepts of most panel members, we found no correlation between self- and overall external assessments.

As mentioned in Sect. 3, the empathy concepts of panel members more closely resembled the empathy concept of Mercer and Reynolds (Mercer & Reynolds, 2002) than that underlying the IRI. In their conclusions, Mercer and Reynolds describe empathy as a complex, multidimensional concept involving the abilities to understand the patient’s situation, perspective and feelings; to communicate that understanding and check its accuracy; and to act on that understanding in a helpful way. This description resembles the results from our focus group interviews on the concept of empathy more closely when compared to Davis’ concept (Davis, 1983). However, there is no validated instrument for the assessment of empathy developed on the basis of the concept of Mercer and Reynolds.

One major limitation of our study is the participation rate among the panel members. We cannot rule out selection bias to some extent. But with time constraints being the most often mentioned reason for non-participation, we have no reason to assume systematic bias in our data. The questioned panel members were a good mix of female and male participants and students and faculty members, meaning that variables known to influence conceptualisation of empathy and experience were adequately represented in our participants. The variation of the accounts of our focus group participants hints to us having reached the criteria of maximum variation in sampling and data. Even though we only investigated panel members from a single institution, we deem our participants not to be extremely different from panel members in other institutions and believe our results to be cautiously transferable to other faculties in Germany. Furthermore, different panel members were interviewed when compared to those who assessed empathy 1 year earlier in the context of the quantitative study. Hence, the panel members were recruited from the same population and had the same prerequisites in both years.

Using the IRI might be seen as a limitation in terms of generalisability, since for the medical context, the JSPE is more common (O’Tuathaigh et al., 2019). However, its wording makes it easy to guess what it measures, which makes it particularly vulnerable to social desirability effects. This is why we decided to use a less obvious
measurement of empathy in the selection situation (in which social desirability effects may play an even bigger role), even though this measure is not customised to the medical context.

Usually, focus groups should include at least 3–5 participants, going up to 12 participants or more depending on the literature (Krueger & Casey, 2014). We aimed at conducting the focus group sessions in closest time proximity to the selection interviews to guarantee that focus group participants adequately remembered their selection and decision processes. Due to the time constraints of many selection panel members, we decided to conduct questioning with even smaller numbers of participants to include as many selection panel members as possible.

Mixing methods was a strength of our study. Our sequential explanatory design started with quantitative data collection and analysis (applicants’ self-judgement and selection panel members’ judgement of applicants’ empathy), which was then followed by focus groups with selection panel members on empathy and the selection process to further explore and explain the quantitative findings. Therefore, the emphasis was on the quantitative phase of the study (QUAN→qual) (Mayring, 2000; O’Cathain et al., 2008). We chose to apply a mixed-methods approach, because quantitative data alone cannot explain why self- and external judgements (do not) correlate with each other and how external judgement takes place. Therefore, a qualitative exploration of these processes was needed to complement, fully illustrate and understand the background of the quantitative results that provided the basis for the development of the qualitative questioning routes. The qualitative data gives explanations for the rather surprising non-correlation of two measures which, as our research revealed, do not seem to measure the same but very different constructs.

In light of our results and the literature, it seems to be premature to introduce empathy as a selection criterion to the selection procedures for medical education in Germany. Neither self-assessment nor external assessment of empathy can be judged invalid methods for student admission based on our results. However, before any of these is implemented broadly, further insights have to be acquired. As a next step, further research, including prospective, longitudinal studies, is needed to determine which empathy instrument is best suitable for use in this context and regarding the question whether self- or external assessment lead to better results (i.e. better students and doctors).

While empathy is a possible amendment to established selection criteria for medical education in Germany, its external assessment should not be employed without training panel members based on an established theoretical concept of empathy and an objective self-assessment measure in order to ensure a common understanding of empathy.

Acknowledgements We would like to thank Katrin U. Obst for her help during the development, implementation and analysis of the quantitative part of our study; Anne Duus for her help with the transcription of the interviews; and Karen Sievers and Josefin Wagner for their help during data collection for the quantitative part.

Author contribution TK and JCS conceived of and designed the study. NJP gave advice on study design. TK and JCS drafted the interview/focus group guideline and conducted all interviews/focus groups. TK
and JCS analysed the data. NJP helped with interpreting the data. TK drafted the manuscript. JCS and
NJP revised the manuscript for important intellectual content. All authors read and approved the final
manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL. The study was funded by the
University of Lübeck. The funding source had no role in the design of this study and its analysis, interpre-
tation of the data or decision to submit results.

**Data availability** All data and materials are available upon request to the corresponding author.

**Code availability** Not applicable.

**Declarations**

**Ethics approval** The study protocol was approved by the ethics committee of the University of Lübeck (file reference 16–083). It was conducted in accordance with the Declaration of Helsinki.

**Consent to participate** Informed consent was obtained from all individual participants included in the
study.

**Consent for publication** The authors affirm that human research participants provided informed consent
for the publication of data collected in this study.

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License,
which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as
you give appropriate credit to the original author(s) and the source, provide a link to the Creative Com-
mons licence, and indicate if changes were made. The images or other third party material in this article
are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the
material. If material is not included in the article’s Creative Commons licence and your intended use is
not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission
directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licen-
ses/by/4.0/.

**References**

Beckman, H. B., Markakis, K. M., Suchman, A. L., & Frankel, R. M. (1994). The doctor-patient rela-
tionship and malpractice. Lessons from plaintiff depositions. *Archives of Internal Medicine, 154*,
1365–1370.

Bogner, K. & Landrock, U. (2016). Response biases in standardised surveys. *GESIS Survey Guidelines*
Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/
gesis-sg_en_016

Casas, R. S., Xuan, Z., Jackson, A. H., Stanfield, L. E., Harvey, N. C., & Chen, D. C. (2017). Associa-
tions of medical student empathy with clinical competence. *Patient Education and Counseling, 100*,
742–774.

Colliver, J. A., Conlee, M. J., Verhalst, S. J., & Dorsey, J. K. (2010). Reports of the decline of empathy
during medical education are greatly exaggerated: A reexamination of the research. *Academic Medi-
cine, 85*, 588–593.

Costa, P., Magalhaes, E., & Costa, M. J. (2013). A latent growth model suggests that empathy of medical
students does not decline over time. *Advances in Health Sciences Education, 18*, 509–522.

Coulehan, J. L., Platt, F. W., Egener, B., Frankel, R., Lin, C. T., & Lown, B. (2001). ‘Let me see if i have
this right...’: Words that help build empathy. *Annals of Internal Medicine, 135*, 221–227.
Creswell, J. W., & Clark, V. L. P. (Eds.). (2017). Designing and conducting mixed methods research. Sage Publications.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113–126.

Dohrenwend, A. M. (2018). Defining empathy to better teach, measure, and understand its impact. *Academic Medicine, 93*, 1754–1756.

Edwards, A. L. (1957). The social desirability variable in personality assessment and research. Dryden Press.

Federal Constitutional Court (2017). Judgment of the First Senate of 19 December 2017 - 1 BvL 3/14 -., paras. (1–253). http://www.bverfg.de/e/ls20171219_1bv1000314en.html. Accessed 31 March 2022.

Federal Ministry of Justice and Consumer Protection (2018). [Regulation on the central allocation of study places through the Foundation for Higher Education Admission]. http://www.gesetze-recht sprechung.sh.juris.de/portal/?quelle=jlink&docid=MWRE190004162&psml=bsshoprod.psml&max=true. Accessed 31 March 2022.

Ferreira-Valente, A., Monteiro, J. S., Barbosa, R. M., Salgueira, A., Costa, P., & Costa, M. J. (2017). Clarifying changes in student empathy throughout medical school: A scoping review. *Advances in Health Science Education, 22*, 1293–1313.

Foundation for Higher Education Admission (2018). [Data on nationwide admission restricted study programmes at universities]. https://hochschulstart.de/fileadmin/user_upload/bew_zv_ws18.pdf. Accessed 31 March 2022.

Hemmerdinger, J. M., Stoddart, S. D., & Lilford, R. J. (2007). A systematic review of tests of empathy in medicine. *BMC Medical Education, 7*, 24.

Hojat, M., Gonnella, J. S., Nasca, T. J., Mangione, S., Vergare, M., & Magee, M. (2002). Physician empathy: Definition, components, measurement, and relationship to gender and specialty. *American Journal of Psychiatry, 159*, 1563–1569.

Hojat, M., Louis, D. Z., Markham, F. W., Wender, R., Rabinowitz, C., & Gonnella, J. S. (2011). Physicians’ empathy and clinical outcomes for diabetic patients. *Academic Medicine, 86*, 359–364.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*, 1277–1288.

Hui, C. H., & Triandis, H. C. (1985). The instability of response sets. *Public Opinion Quarterly, 49*, 253–260.

Jeffrey, D. (1994). A meta-ethnography of interview-based qualitative research studies on medical students’ views and experiences of empathy. *Medical Teacher, 38*, 1214–1220.

Kim, K., Kim, S. H., Yoon, H. S., Shin, H. S., & Lee, Y.-M. (2020). Assessing the effects of an empathy education program using psychometric instruments and brain fMRI. *Advances in Health Sciences Education, 25*, 283–295.

Kötter, T., Carminenke, S., & Herrmann, W. J. (2014). Compatibility of scientific research and specialty training in general practice. A cross-sectional study. *GMS Zeitschrift für Medizinische Ausbildung, 31*, Doc31.

Kötter, T., Tautphäus, Y., Scherer, M., & Voltmer, E. (2014). Health-promoting factors in medical students and students of science, technology, engineering, and mathematics: Design and baseline results of a comparative longitudinal study. *BMC Medical Education, 14*, 134.

Kötter, T., Pohontsch, N. J., & Voltmer, E. (2015). Stressors and starting points for health-promoting interventions in medical school from the students’ perspective: A qualitative study. *Perspectives on Medical Education, 4*, 128–135.

Kötter, T., Ritter, J., Katalinic, A., & Voltmer, E. (2016). Predictors of participation of sophomore medical students in a health-promoting intervention: An observational study. *PLoS ONE, 11*, e0168104.

Kötter, T., Tautphäus, Y., Obst, K. U., Voltmer, E., & Scherer, M. (2016). Health-promoting factors in the freshman year of medical school: A longitudinal study. *BMC Medical Education, 50*, 646–656.

Kötter, T., Rose, S. I., Waldmann, A., & Steinhäuser, J. (2020). Do Medical Students in Their Fifth Year of Undergraduate Training Differ in Their Suitability to Become a “Good Doctor” Depending on Their Admission Criteria? A Pilot Study. *Advances in Medical Education and Practice, 11*, 109–112.

Krueger, R. A., & Casey, M. A. (Eds.). (2014). *Focus groups: A practical guide for applied research*. Sage Publications.
Krüger, H.-P. (1980). Self or external assessment in diagnostics and psychotherapy. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, 2, 339–350.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one’s own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121–1134.

Langford, J., & Clance, P. R. (1993). The imposter phenomenon: Recent research findings regarding dynamics, personality and family patterns and their implications for treatment. *Psychotherapy: Theory, Research, Practice, Training*, 30, 495–501.

Mayring, P. (2000). Qualitative content analysis. *Forum Qualitative Social Research*, 1, 20.

Mercer, S. W., & Reynolds, W. J. (2002). Empathy and quality of care. *British Journal of General Practice*, 52(Suppl), S9-12.

Mommer, A., Wagner, J., Jünger, J., & Westermann, J. (2020). Exam performance of different admission quotas in the first part of the state examination in medicine: A cross-sectional study. *BMC Medical Education*, 20, 169.

Mummendey, H. D., & Grau, I. (Eds.). (2014). *The survey method: Basics and application in personality, attitude and self-concept research*. Hogrefe.

Neumann, M., Scheffer, C., Tauschel, D., Lutz, G., Wirtz, M., & Edelhäuser, F. (2012). Physician empathy: Definition, outcome-relevance and its measurement in patient care and medical education. *GMS Zeitschrift für Medizinische Ausbildung*, 29, Doc11.

O’Cathain, A., Murphy, E., & Nicholl, J. (2008). The quality of mixed methods studies in health services research. *Journal of Health Services Research & Policy*, 13, 92–98.

O’Cathain, A., Murphy, E., & Nicholl, J. (2010). Three techniques for integrating data in mixed methods studies. *British Medical Journal*, 341, 1147–1150.

O’Sullivan, D. M., Moran, J., Corcoran, P., O’Flynn, S., O’Tuathaigh, C., & O’Sullivan, A. M. (2017). Medical school selection criteria as predictors of medical student empathy: A cross-sectional study of medical students. *Ireland. BMJ Open*, 7, e016076.

O’Tuathaigh, C. M. P., Idris, A. N., Duggan, E., Costa, P., & Costa, M. J. (2019). Medical students’ empathy and attitudes towards professionalism: Relationship with personality, specialty preference and medical programme. *PLoS ONE*, 14, e0215675.

Pohontsch, N. J., Heser, K., Löffler, A., Haenisch, B., Parker, D., Luck, T., Riedel-Heller, S., Maier, W., Jessen, F., & Scherer, M. (2017). General practitioners’ views on (long-term) prescription and use of problematic and potentially inappropriate medication for oldest-old patients—a qualitative interview study with GPs (CIM-TRIAD study). *BMC Family Practice*, 18, 22.

Pohontsch, N. J., Stark, A., Ehrhardt, M., Köttter, T., & Scherer, M. (2018). Influences on students’ empathy in medical education: An exploratory interview study with medical students in their third and last year. *BMC Medical Education*, 18, 231.

Preusche, I., & Lamm, C. (2016). Reflections on empathy in medical education: What can we learn from social neurosciences? *Advances in Health Sciences Education*, 21, 235–249.

Quince, T., Thiemann, P., Benson, J., & Hyde, S. (2016). Undergraduate medical students’ empathy: Current perspectives. *Advances in Medical Education and Practice*, 7, 443–455.

Schreier, M. (2014). Ways of doing qualitative content analysis: Disentangling terms and terminologies. *Forum Qualitative Social Research*, 15, 18.

Steiner-Hofbauer, V., Schrank, B., & Holzinger, A. (2018). What is a good doctor? *Wiener Medizinische Wochenschrift*, 168, 398–405.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4, 25–29.

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 19, 349–357.
von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2008). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Journal of Clinical Epidemiology, 61*, 344–349.

Yamada, Y., Fujimori, M., Shirai, Y., Ninomiya, H., Oka, T., & Uchitomi, Y. (2018). Changes in physicians’ intrapersonal empathy after a communication skills training in Japan. *Academic Medicine, 93*, 1821–1826.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Authors and Affiliations**

**Thomas Kötter**¹,² · **Johanna Christine Schulz**² · **Nadine Janis Pohontsch**³

Thomas Kötter  
thomas.koetter@uni-luebeck.de

1. Institute of Family Medicine, University Medical Centre Schleswig-Holstein, Campus Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany
2. Institute of Social Medicine and Epidemiology, University Medical Centre Schleswig-Holstein, Campus Lübeck, Lübeck, Germany
3. Department of General Practice/Primary Care, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany