

Development of the Jefferson Scale of Empathy for Teachers (JSE-T) to Measure Empathy in Educationally Relevant Situations

Christoph M. Paulus, PhD

The University of Saarbrücken
Faculty of Empirical Human Sciences and
Economics, Department of Educational Sciences, Germany

Eric Klopp, PhD

The University of Saarbrücken
Faculty of Empirical Human Sciences and Economics
Department of Educational Sciences, Germany

Doi: 10.19044/ejes.v10no2a169

<https://doi.org/10.19044/ejes.v10no2a169>

Submitted: 16 January 2023

Accepted: 28 September 2023

Published: 31 October 2023

Copyright 2023 Author(s)

Under Creative Commons CC-BY 4.0

OPEN ACCESS

Abstract

The Jefferson Scale of Empathy is one of the most commonly used scales in medical education to measure empathy. It is specific to the field of medical education and geared toward orienting medical students to physician empathy in patient care situations. The scale was transferred to the educational context in teacher education. In doing so, the questionnaire was reduced from the original 20 items to 9 because of content and methodological issues. A CFA showed good model-fit parameters for a three-factor model, and correlations with the German version of the Interpersonal reactivity Index were in line with the magnitudes reported in previous literature. In total, the JST-E scales show evidence for their factorial and convergent validity and their reliability. The JSE-T proves to be a good instrument for measuring empathy in educational contexts and thus closes the gap between trait measurement procedures such as the IRI and concrete-situational judgment tests, so that an economical, multidimensional testing of empathy becomes possible.

Keywords: Empathy, Jefferson Scale of Empathy, teacher education

Introduction

Empathy is the ability to understand, comprehend and sympathize with the internal state of another. This includes both cognitive and affective abilities such as perspective taking, empathy, and compassion. Saxena et al. (2017, p.

765) describe empathy as "the ability to understand and share in the internal states of others (...) as well as the abilities to react to the internal states of others, and to distinguish between one's own and others' internal states." Thus, it is a complex construct consisting of several components and, depending on the research direction, the focus is placed more on affective or on cognitive processes.

Empathic individuals are more successful in social interactions with others (Baron-Cohen & Wheelwright, 2004; Davis, 1983a; Mehrabian & Epstein, 1972). A teacher's empathy skills are therefore also among the strongest predictors of positive academic as well as affective and behavioral student outcomes (Cornelius-White, 2007). It is helpful in many phases of teaching, such as classroom management (Emmer & Stough, 2001), in recognizing learners' emotions such as anxiety, anger, or even joy in various learning activities (Weisz et al., 2021), but also in dealing more effectively with disruptions (Stojiljković et al., 2012) or bullying (Bilz et al., 2017). Teachers who rated their own emotion recognition skills higher rated their subjective teaching success more positively (Wu et al., 2019) and were also viewed as "more skilled" by their students (Ghanizadeh & Moafian, 2009; Khodadady, 2012). Therefore, acting empathically is considered an important aspect of educational professionalism (Auernheimer, 2016).

Affective empathy describes empathizing with the internal state of a counterpart and has a high genetic disposition (Farrell & Vaillancourt, 2020; Melchers et al., 2016), which is also supported by the fact that this trait changes little over the course of age (Ziaei et al., 2021). Perspective-taking ability, on the other hand, is a cognitive trait that, like all cognitions, first emerges and changes over development (Farrell & Vaillancourt, 2020; Zahn-Waxler et al., 1992) and also declines somewhat with older age (Ziaei et al., 2021). Therefore, most empathy training also aims to improve perspective-taking (Bas-Sarmiento et al., 2017; Fragkos & Crampton, 2020; Paulus & Meinken, 2022a, 2022b).

The Jefferson Scale of Empathy (JSE) is one of the most commonly used scales in medical education to measure empathy (Costa et al., 2017; Hojat et al., 2002; Hojat et al., 2001; Mehta et al., 2021; Nasr Esfahani et al., 2014; Preusche & Wagner-Menghin, 2013). It was originally constructed to measure medical students' orientation toward physician empathy in patient care situations and, more importantly, to assess prospective physicians' cognitive-empathic skills in dealing with patients. It measures three factors, Perspective Taking, Compassionate Care, and Walking in Patient's Shoes (Hojat et al., 2018)..

Perspective Taking contains 10 items, the content of which is aimed more at cognitive efforts to understand patients (e.g., "I try to think like my patients in order to render better care"), but also explicitly excludes affective

empathy components ("I believe that emotion has no place in the treatment of medical illness").

Compassionate care (8 items) aims at warm-hearted interaction with patients and "is defined as a combination of empathy and sufficient degree of sympathy, (it) is considered an essential dimension of the patient-physician relationship" (Hojat & LaNoe, 2014, p. 75). However, the items of this factor are partly formulated negatively ("I do not enjoy reading non-medical literature and the arts.") or mix cognitive ("I try to imagine myself in my patients' shoes...") and affective aspects ("I have a good sense of humor that I think contributes to a better clinical outcome") within the factor, so that it does not seem to be clearly interpretable to us..

The third factor, "*walking in patient's shoes*", describes the importance of empathy as a therapeutic factor ("I believe that empathy is an important therapeutic factor in medical treatment.").

One thing that is often discussed is the validity relation to other empathy questionnaires, especially the Interpersonal Reactivity Index (IRI) (Davis, 1980, 1983; Paulus, 2009, 2021). This approach to empathy measurement, which originates from social psychology, includes two affective (emotional concern, personal distress) and two cognitive factors (perspective taking, fantasy scale) respectively.

Perspective Taking measures the attempt to spontaneously see something from another person's perspective ("Before I criticize someone, I try to imagine how I would feel in their place"). The *Fantasy Scale* captures the tendency of the respondent to put himself in the place of characters in novels or movies ("I can imagine the feelings of a person in a novel very well"). The remaining two subscales represent operationalizations of an observer's typical emotional reactions: The *Empathic Concern* scale is designed to measure other-oriented feelings such as compassion or concern for persons in distress ("I feel warm feelings for people who are less well off than I am"), whereas the *Personal Distress* scale is designed to measure intrinsically focused feelings such as unease or discomfort in close interpersonal relationships ("Being in a tense emotional situation makes me anxious").

At best, moderate correlations (Costa et al., 2017; Hojat & Gonnella, 2017) between the questionnaires are reported. However, it is often forgotten here that the JSE has only situation-typical (i.e., rather state-) item contents, whereas the IRI in its item formulation focuses on (trait-) basic behaviors in different situations (Song et al., 2019). Moreover, the measured factors of the two questionnaires are not identical, although there are overlaps. "The IRI relies on the definition of empathy as a combination of both cognitive and emotional attributes, whereas the JSE was developed based on a definition of empathy in clinical context as a predominantly cognitive (as opposed to

emotional) attribute that involves understanding (rather than feeling) the patient's pain, suffering, experiences, and concerns" (Hojat & Gonnella, 2017, p. 743).

However, this situation specificity can also be utilized when transferring the contents of the JSE to other professional groups.

For our studies, the target group of the questionnaire was transferred from medical professionals to teachers and educational contexts, referring to the adapted version as the JSE-T. As a starting point, the items from Preusche and Wagner-Menghin (2013) were used. The transfer of the wording is shown in the following example item 2:

JSE: "Patients feel better when their physicians understand their feelings."

JSE-T: "Students feel better when their teachers understand their feelings."

However, this was not possible for some items because they were too far removed from the pedagogical context. This concerned items 4 ("Understanding body language is as important as verbal communication in physician-patient relationships."), 5 ("A physician's sense of humor contributes to a better clinical outcome."), and 19 ("I do not enjoy reading non-medical literature or the arts."). In addition, item 18 was not transferred due to negative wording and the associated difficulties of semantically correct answers in German. („Physicians should not allow themselves to be influenced by strong personal bonds between their patients and their family members.") In transferring the item content, it was nevertheless attempted to replicate the three-factor structure with the caveat that our study's sample has greater variance than Hojat's (2018) because student teachers have a greater breadth of subjects than medical students. For the JSE-T, it was expected that the 3-factor structure corresponds to the structure of the JSE (Hojat & LaNoue, 2014), with factor 1 corresponding to the "Perspective Taking" factor, factor 2 corresponding to the items on "Walking in patient's shoes", and factor 3 corresponding to "Compassionate Care".

As described at the beginning, the two theoretical approaches of the IRI and the JSE differ to some extent (see above), so that high correlations are not to be expected. The correlations between the JSE scales and the relevant EC and PT scales of the IRI found in other studies are small to medium (Hojat et al., 2001, p. 361: $r_{EC} = .41$, $r_{PT} = .29$); Costa et al., 2017, p. 865: $r_{EC} = .23$, $r_{PT} = .27$). For our adapted version, medium correlations of the three JSE-factors with the EC and PT scale from a German version of the IRI, i.e., the SPF (Paulus, 2009) were expected.

Method

Sample

The sample consisted of 167 student teachers between the ages of 18 and 40 ($M = 22.81$, $SD = 4.168$), of which 116 were female and 51 were male. The SPF responses were also available from 77 subjects.

Data collection tools

The 16 items of the adapted version of the JSE (Preusche & Wagner-Menghin, 2013) was used as starting point, cf. appendix 1. The items were presented in combination with a 5-point rating scale, ranging from 1 = Strongly Disagree to 5 = Strongly Agree. In addition to that, the German version of the IRI, the SPF (Paulus, 2009), was used to measure trait empathy. The data were collected online to ensure that no missing values occurred.

Statistical analysis

The statistical analysis was performed with R with a nominal α level of $\alpha \leq .05$. The package lavaan (Rosseel, 2012) was used to conduct the exploratory and confirmatory factor analysis with maximum likelihood estimation.

Before conducting factor analyses, an item screening based on the descriptive statistics was considered. In this screening procedure, for each of the three potential dimensions of the JSE-T, those items were excluded which had a mean close to upper end of the rating scale and a small variance.

To examine the factorial validity of the JSE-T questionnaire, a two-step procedure was used: In the first step, it was investigated whether the assumed three-factor structure holds by means of exploratory factor analysis (EFA). The information criteria were used to determine the number of factors (Klopp, 2022). Models with 1 up to 5 factors were estimated and considered the AIC, BIC and SBIC. As the items were designed to measure three factors, the information criteria should have their minimum for the 3-factor model. Additionally, the usual fit indices to judge model fit were considered. Fit indices for the EFA model and the following models are made using the guidelines provided in Schermelleh-Engel, Moosbrugger, and Müller (2003). In the second step, a confirmatory factor analysis (CFA) model was set up. A CFA model imposes a strict test of the simple structure in the sense of Thurstone (1947) as all cross loadings are assumed to be zero. Thus, a well-fitting CFA model provides strong evidence for the factorial validity of the JSE-T. In the first CFA model (model 1), the latent variables are measured by the pertinent items. In particular, the factor F1 should capture the scale *Show understanding*, the factor F2 the scale *Empathy as a pedagogical skill*, and the factor F3 the scale *Pedagogy instead of empathy*. To scale the latent variables, the latent variances were restricted to 1. With this scaling, the covariances

between the latent variables can be interpreted as correlations and the loadings reflect the amount of change when the latent variable changes one unit (Klopp & Klößner, 2021). From a measurement perspective, it would be desirable if all items had more or less equal loadings such that each item reflects the same amount of change of the latent variable. From a psychometric perspective, this corresponds to a τ -equivalent measurement model (cf., Brown, 2015). To test if τ -equivalence holds, a second CFA model (model 2) was set up with the restriction that the loadings had to be equal for each factor.

To investigate the convergent validity in the sense of the correlations of the factors with the SPF scales, a second CFA model (model 3) was investigated in which the SPF's scales EC and PT were considered. Again, to scale the latent variables, the latent variances were restricted to 1. Because of the small sample size ($n = 77$) for model 3, bounded estimation (De Jonckere & Rosseel, 2022) were applied. In bounded estimation, bounds are set on certain model parameters to increase the convergence of the model. The bounds were set such that both the manifest and latent variances should be positive. In model 3, the EC and PT scales are entered as manifest variables represented by the standardized sum scores according to Paulus (2009). To judge the convergent validity, the correlation between the three latent variables and the EC and PT scales are considered.

Additionally, internal consistencies in the sense of Cronbach's alpha for the JSE-T scales were calculated. As in typical applications these scales will not be used in a CFA, the sum scores for each of the JSE-T scales were calculated to investigate the convergent validity for these scales. The sum scores for the JSE factors were calculated with the pertinent items for each factor. Pearson correlations were used to examine the correlations between the JSE-T sum scores and the SPF scales.

Results

Table 1. Descriptive data for adapted JSE-T Items cf. appendix 1 (Omitted items are shown in italics)

Item	Intended			M	SD
	factor	Minimum	Maximum		
<i>JSET01</i>	<i>1</i>	<i>3</i>	<i>5</i>	<i>4.63</i>	<i>0.509</i>
JSET02	1	3	5	4.47	0.599
<i>JSET03</i>	<i>1</i>	<i>2</i>	<i>5</i>	<i>3.02</i>	<i>0.620</i>
<i>JSET06</i>	<i>1</i>	<i>1</i>	<i>5</i>	<i>2.94</i>	<i>0.726</i>
<i>JSET07</i>	<i>2</i>	<i>1</i>	<i>5</i>	<i>4.26</i>	<i>0.952</i>
<i>JSET08</i>	<i>2</i>	<i>2</i>	<i>5</i>	<i>4.32</i>	<i>0.641</i>
<i>JSET09</i>	<i>1</i>	<i>2</i>	<i>5</i>	<i>4.50</i>	<i>0.702</i>
<i>JSET10</i>	<i>2</i>	<i>1</i>	<i>5</i>	<i>4.16</i>	<i>0.760</i>
JSET11	3	1	5	1.81	0.819
JSET12	1	2	5	3.96	0.787
JSET13	1	3	5	4.26	0.713

JSET14	3	1	5	1.41	0.704
JSET15	2	2	5	4.53	0.657
JSET16	1	3	5	4.36	0.613
JSET17	1	1	5	3.13	0.933
JSET20	2	2	5	4.66	0.578

The final German items and English translation of the JSE-T questionnaire are shown in the appendix 2.

Because of the item exclusion, the factors of the JSE-T scale assess the following dimensions:

- **Factor 1 (F1, “Show understanding“):** The factor describes perspective taking in conversations with students. This makes the relationship between teacher and student more trusting and more understanding on the part of the teacher.
- **Factor 2 (F2, “Empathy as a pedagogical skill“):** Empathy is generally valued as a pedagogical skill that can increase teacher success.
- **Factor 3 (F3, “Pedagogy instead of empathy“):** The view here is that pedagogical measures and rules, especially in the case of school problems, are more important than empathy and that emotions should be left out of the solution of school problems.

Table 2 shows the descriptive statistics for the items of the JSE-T and the SPF scales. With regard to the first step in investigating the factorial validity of the JSE-T, all three information criteria indicated a three-factor solution, see table 3. As indicated by the fit indices, this model fit the data well.

Table 2. Correlation matrix of the JSE-T items (cf. appendix 2), EC, and PT scale

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	EC	PT
Item 1	-										
Item 2	.348	-									
Item 3	.400	.409	-								
Item 4	.256	.323	.492	-							
Item 5	.305	.403	.431	.416	-						
Item 6	.266	.323	.230	.345	.380	-					
Item 7	.295	.336	.366	.351	.450	.603	-				
Item 8	.070	-.078	-.115	-.143	-.130	-.063	-.147	-			
Item 9	-.131	-.149	-.176	-.251	-.145	-.180	-.278	.477	-		
EC	.236	.207	.268	.141	.280	.222	.314	-.130	-.143	(.74)	
PT	.066	.276	.316	.182	.221	.052	.244	-.450	-.150	.257	(.78)

Note. Correlation coefficients in bold are significant. The numbers in brackets indicate Cronbach's α of the EC and PT scales

Table 3. Information criteria and fit measures for the EFA model selection

factors	AIC	BIC	SBIC	χ^2	df	p	CFI	SRM R	RMS EA	RMSEA CI		
										lower	upper	p _{close}
1	2848.281	2904.405	2847.414	95.947	27	.000	0.804	0.085	0.124	0.097	0.151	.000
2	2819.710	2900.777	2818.458	51.376	19	.000	0.908	0.051	0.101	0.068	0.135	.007
3	2794.525	2897.418	2792.935	12.191	12	.430	0.999	0.023	0.010	0.000	0.080	.750
4	2799.622	2921.224	2797.744	5.288	6	.507	1.000	0.014	0.000	0.000	0.094	.727
5	2805.609	2942.801	2803.490	1.275	1	.259	0.999	0.007	0.041	0.000	0.215	.352

Note. Bold numbers indicate the factor model with the minimum information criteria.

Table 4 shows that the CFA model 1 fits the data very well indicating evidence for the factorial validity of the JSE-T scale. CFA model 3 provides evidence for the convergent validity of the JSE-T factors. Firstly, the model fits the data, see table 4.

Table 4. Fit measures for the CFA models

	χ^2	df	p	CFI	SRMR	RMSEA	RMSEA CI		
							lower	upper	p _{close}
Model 1	27.735	24	.271	0.989	0.041	0.031	0	0.072	.734
Model 2	41.737	30	.075	0.967	0.077	0.048	0	0.081	.499
Model 3	46.730	36	.109	0.922	0.075	0.062	0	0.108	.328

Table 5 (left panel) shows the factor loadings. As can be seen, all factor loadings load positively on all the factors as intended per item construction. Two notable results are the rather low error variances of item 7 and item 9. Whereas the error variance for item 7 is statistically significant ($p = .008$), the error variance for item 9 is not ($p = .876$). This is in contrast to all other manifest error variances which are statistically significant (all $p \leq .001$) and may indicate potential problems with these items. The factor correlations are as expected, factor 1 and 2 show a strong and significant correlation ($p \leq .001$), where factors 1 and 3 show a medium negative and significant correlation ($p = .006$) and the same applies to the correlation between factors 2 and 3 ($p = .006$).

Concerning the τ -equivalence of the measurement model, model 2 in table 5 (right panel) indicates a rather good model fit providing evidence that each indicator contributes equally in the measurement of the three JSE-T factors. All factor loadings and all manifest residual error variances are significant. The correlation pattern among the factors remains the same, although their numerical magnitudes change slightly.

Table 5. Loadings, manifest error variances θ , and factor correlations in CFA Model 1 and 2

CFA Model 1					CFA Model 2			
Factor loadings								
Variable	F1	F2	F3	θ	F1	F2	F3	θ
Item 1	.304			.264	.411			.257
Item 2	.413			.319	.411			.316
Item 3	.542			.322	.411			.372
Item 4	.447			.305	.411			.322
Item 5	.407			.208	.411			.205
Item 6		.458		.219		.479		.205
Item 7		.497		.086		.479		.099
Item 8			.402	.504			.528	.409
Item 9			.679	.031			.528	.204
Factor correlations								
	F1	F2	F3		F1	F2	F3	
F1	-				-			
F2	.682	-			.708	-		
F3	-.284	-.316	-		-.311	-.348	-	

Note. All loadings are significant, $p \leq .001$, all factor correlations are significant, $p \leq .01$

Table 5 shows that whereas the loading structure mostly remains the same as in Model 1, the correlation between factors 1 and 3 almost vanishes and is no longer significant ($p = .774$). The correlation between factor 1 and 2 is also lightly lower than in model 1, but remains in the medium range and is no longer statistically significant ($p \leq .114$). However, the correlation between factor 1 and 2 remains on a high level and also remains significant ($p \leq .001$). The correlations between the JSE-T factors and the SPF scales are as expected: Factor F1 shows medium correlations with the EC and PT scales ($p = .001$, and $p = .002$) and factor 2 shows a medium, significant correlation with the EC scale ($p = .006$). Additionally, for factor 2, the correlation with EC is also significant ($p = .045$). Lastly, for factor 3, the expected negative medium correlation with the PT scale was found which was significant ($p = .001$).

A noteworthy result in model 3 concerns factor 3. For this factor, the loadings of the items now differ largely in their magnitude. Item 8 now has a manifest error variance of zero which corresponds to the lower bound in the bounded estimation. A further observation for factor F3 is that in model 3, the magnitudes of the loadings now differ largely, which is not the case for the other factors in model 3. Additionally, for factor F2, Item 7's manifest error variance is very small which is an analogous result to model 1. However, in model 3, Item 7's manifest error variance is non-significant ($p = .250$). Again, and in analogy to model 1, this is an indication of potential problems with the items.

With regard to the reliability, the correlations of the JSE-T as operationalized by the scales' sum scores with the SPF scales were considered as shown table 6. The derived scales show medium internal consistencies in sense of Cronbach's α . In particular, the derived scale for *Pedagogy instead of empathy* has a rather low internal consistency.

Table 6. Loadings, manifest error variances θ , and factor correlations in CFA Model 3

Factor loadings					
Variable	F1	F2	F3	θ	
Item 1	.269			.334	
Item 2	.279			.316	
Item 3	.488			.344	
Item 4	.325			.338	
Item 5	.364			.244	
Item 6		.236		.228	
Item 7		.326		.043	
Item 8			.874	0 (lb)	
Item 9			.372	.385	

Factor and SPF scale correlations					
	F1	F2	F3	EC	PT
F1	-				
F2	.673	-			
F3	-.040	-.216	-		
EC	.426	.381	-.129	-	
PT	.412	.271	-.447	.254	-

Note. All loadings are significant, $p \leq .001$, all factor correlations are significant, $p \leq .01$. lb: lower bound, this estimate reached the lower bound of possible values in the bounded estimation.

Regarding the question whether there is also evidence for the convergent validity of the JSE-T's sum scales (table 7), the scale *Show understanding* shows a medium positive and significant correlation with the SPF's EC and PT scales. The scale for *Empathy as a pedagogical skill* also shows a medium positive and significant with the PT scale. However, the scale for *Pedagogy instead of empathy* shows a medium negative, significant correlation with the PT scales. All correlations are in line with theoretical expectations and correspond to the pattern of correlations found in model 2. As described in the introduction, the two theoretical approaches of the SPF and the JSE-T differ to some extent, so that high correlations are not to be expected. For both the correlations of the JSE-T with the SPF in model 2 and in terms of sum scores, the correlations found here are in line with other studies such as Hojat et al. (2001, p. 361) or Costa et al. (2017).

In total, the JST-E scales show evidence for their factory and convergent validity and their reliability.

Table 7. Internal consistencies and correlations between the JSE sum scores and the SPF sum score scales

Sum score scale	Cronbach's α	EC	PT
Show understanding	.75	0.347	0.330
Empathy as a pedagogical skill	.75	0.312	0.159
Pedagogy instead of empathy	.64	-0.156	-0.360

Note. Correlations in bold are significant, $p \leq .001$. EC: Empathic concern, PT: Perspective taking

Discussion

In the context of empathy research, predominantly questionnaires are used that are more trait oriented such as the IRI (Davis, 1980), the E scale (Leibetseder & Laireiter, 2001), MET (Dziobek et al., 2008), or the EQ (Baron-Cohen & Wheelwright, 2004). These capture more general empathic behaviors and attitudes that are exhibited across situations. However, this makes these measurement methods only partially suitable for providing information about the empathic behavior of individuals in very specific contexts such as medicine or even pedagogy. The situation specificity of the JSE was made use of by transferring the item contexts to pedagogical and school situations in order to be able to obtain specific statements about empathy-relevant behavior in school situations as a supplement to the IRI. Thus, the gap between trait-oriented measurement methods and concrete-situational questionnaires such as those represented by the Situational Judgement Tests (Chao et al., 2020; McDaniel et al., 2007; Patterson et al., 2016; Wolf et al., 2020) was closed. The JSE-T, which focuses on the attitudes of prospective teachers, can replicate the factor structure of the original JSE (albeit with different contexts of the factors) and is also more economical to use with only 9 items instead of 20. The correlations to the factors of the SPF are within expectable orders of magnitude and indicate theoretically justifiable directions.

Of course, this study represents only a first approach to the construction of such a scale, and there are also some limitations that should be considered when applying the JSE-T, especially when it is used in the context of latent variable models. This refers to the issues of small and not statistically significant error variances. A possible reason for this finding is the almost identical content of the items. Further investigations should refine the items measuring the second and third factor, in particular their wording. In addition to that, any statements about short- and long-term retest reliability cannot yet be made, since no data on this is available yet. However, the reliability calculations via Cronbach's alpha show good internal consistencies of the three factors. Thus, the scales can be used in typical applications in which sum or mean scores are used, e.g., to investigate correlations with other constructs.

The JSE-T proves to be a good instrument for measuring empathy in educational contexts and thus closes the gap between trait measurement procedures such as the IRI and concrete-situational judgment tests, so that an economical, multidimensional testing of empathy becomes possible. Due to the only medium sample size, further studies on this are needed, but the test-statistical results form a solid basis of the usefulness of the approach chosen here.

Conclusion

The aim of the study was to construct a questionnaire to measure empathy in the specific context of school. In doing so, it was attempted to apply the items of the JSE, a questionnaire frequently used in medical education to measure empathy, to school contexts. This was not possible for all items. The content validity of the remaining items was tested by means of several CFA models and three factors with good model-fit criteria emerged.

With the construction of the JSE-T it succeeded in a first attempt to develop a more state-oriented instrument for empathy in pedagogical contexts. Thus, we add to the repertoire of applicable questionnaires for studies whose research questions aim to explore the relevance and influence of empathy in the education of student teachers, but also in the direct school context. The use of a more state-oriented questionnaire thus complements and extends the previous more trait-specific measurement methods. This is particularly recommended in the context of empathy training to show whether training methods or contents are more likely to prove successful in the short or long term. Short-term changes are more likely to occur in situation-specific (state) contexts, whereas changes in personality structure (traits) are more difficult to achieve.

Human Studies: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Funding Statement: The authors did not obtain any funding for this research.

Data Availability: All the data are included in the content of the paper.

Conflict of Interest: The authors reported no conflict of interest.

References:

1. Auernheimer, G. (2016). *Einführung in die interkulturelle Pädagogik* (Vol. 8). WBG.
2. Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34(2), 163-175. <https://doi.org/10.1023/b:jadd.0000022607.19833.00>
3. Bas-Sarmiento, P., Fernandez-Gutierrez, M., Baena-Banos, M., & Romero-Sanchez, J. M. (2017). Efficacy of empathy training in nursing students: A quasi-experimental study. *Nurse Educ Today*, 59, 59-65. <https://doi.org/10.1016/j.nedt.2017.08.012>
4. Bilz, L., Schubarth, W., Dudziak, I., Fischer, S. M., Niproschke, S., & Ulbricht, J. (2017). *Gewalt und Mobbing an Schulen. Wie sich Gewalt und Mobbing entwickelt haben, wie Lehrer reagieren und welche Kompetenzen sie brauchen*. Klinkhardt.
5. Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
6. Chao, T.-Y., Sung, Y.-T., & Huang, J.-L. (2020). Construction of the situational judgment tests for teachers. *Asia-Pacific Journal of Teacher Education*, 48(4), 355-374. <https://doi.org/10.1080/1359866X.2019.1633621>
7. Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research*, 77(1), 113-143. <https://doi.org/10.3102/003465430298563>
8. Costa, P., de Carvalho-Filho, M. A., Schweller, M., Thiemann, P., Salgueira, A., Benson, J., Costa, M. J., & Quince, T. (2017). Measuring Medical Students' Empathy: Exploring the Underlying Constructs of and Associations Between Two Widely Used Self-Report Instruments in Five Countries. *Acad Med*, 92(6), 860-867. <https://doi.org/10.1097/ACM.0000000000001449>
9. Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. (JSAS Catalogue of Selected Documents in Psychology), 10-85.
10. Davis, M. H. (1983). Measuring Individual-Differences in Empathy - Evidence for a Multidimensional Approach. *Journal of Personality and Social Psychology*, 44(1), 113-126. <https://doi.org/10.1037/0022-3514.44.1.113>
11. De Jonckere, J & Rosseel, Y. (2022) Using bounded estimation to avoid nonconvergence in small sample structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 29, 412-427.

12. Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., & Convit, A. (2008). Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the Multifaceted Empathy Test (MET). *Journal of Autism and Developmental Disorders*, 38(3), 464-473. <https://doi.org/10.1007/s10803-007-0486-x>
13. Emmer, E. T., & Stough, L. M. (2001). Classroom Management: A Critical Part of Educational Psychology, With Implications for Teacher Education. *Educational Psychologist*, 36(2), 103-112. https://doi.org/10.1207/s15326985ep3602_51
14. Farrell, A. H., & Vaillancourt, T. (2020). Adolescent empathic concern and perspective taking: Heterogeneous developmental trajectories and childhood social and psychological factors. *Journal of Personality*, 00, 1-17. <https://doi.org/10.1111/jopy.12607>
15. Fragkos, K. C., & Crampton, P. E. S. (2020). The Effectiveness of Teaching Clinical Empathy to Medical Students: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Academic Medicine*, 95(6), 947-957. <https://doi.org/10.1097/acm.0000000000003058>
16. Ghanizadeh, A., & Moafian, F. (2009). The role of EFL teachers' emotional intelligence in their success. *ELT Journal*, 64(4), 424-435. <https://doi.org/10.1093/elt/ccp084>
17. Hojat, M., & Gonnella, J. S. (2017). What Matters More About the Interpersonal Reactivity Index and the Jefferson Scale of Empathy? Their Underlying Constructs or Their Relationships With Pertinent Measures of Clinical Competence and Patient Outcomes? *Acad Med*, 92(6), 743-745. <https://doi.org/10.1097/ACM.0000000000001424>
18. Hojat, M., & LaNoue, M. (2014). Exploration and confirmation of the latent variable structure of the Jefferson scale of empathy. *Int J Med Educ*, 5, 73-81. <https://doi.org/10.5116/ijme.533f.0c41>
19. Hojat, M., DeSantis, J., Shannon, S. C., Mortensen, L. H., Speicher, M. R., Bragan, L., LaNoue, M., & Calabrese, L. H. (2018). The Jefferson Scale of Empathy: a nationwide study of measurement properties, underlying components, latent variable structure, and national norms in medical students. *Advances in Health Sciences Education*, 23(5), 899-920. <https://doi.org/10.1007/s10459-018-9839-9>
20. Hojat, M., Gonnella, J. S., Nasca, T. J., Mangione, S., Veloksi, J. J., & Magee, M. (2002). The Jefferson Scale of Physician Empathy: Further Psychometric Data and Differences by Gender and Specialty at Item Level. *Academic Medicine*, 77(Supplement), S58-S60. <https://doi.org/10.1097/00001888-200210001-00019>

21. Hojat, M., Mangione, S., Nasca, T. J., Cohen, M. J. M., Gonnella, J. S., Erdmann, J. B., Veloski, J., & Magee, M. (2001). The Jefferson Scale of Physician Empathy: Development and Preliminary Psychometric Data. *Educational and Psychological Measurement*, 61(2), 349-365. <https://doi.org/10.1177/00131640121971158>
22. Klopp, E. & Klößner, S. (2021). The Impact of Scaling Methods on the Properties and Interpretation of Parameter Estimates in Structural Equation Models with Latent Variables. *Structural Equation Modeling: A Multidisciplinary Journal*, 28, 182-206.
23. Klopp, E. (2022). Using information criteria to determine the number of factors in maximum-likelihood exploratory factor analysis. In A. Hernández & I. Tomás (Eds.) *Proceedings from the 9th European Congress of Methodology* (pp. 65-72). Valencia: Universitat de Valencia.
24. Khodadady, E. (2012). Emotional Intelligence and Its Relationship with English Teaching Effectiveness. *Theory and Practice in Language Studies*, 2(10). <https://doi.org/10.4304/tpls.2.10.2061-2072>
25. Leibetseder, M., & Laireiter, A.-R. (2001). E-Skala: Fragebogen zur Erfassung von Empathie - Beschreibung und psychometrische Eigenschaften. *Journal of Individual Differences*, 22(1), 70-85. <https://doi.org/https://doi.org/10.1024//0170-1789.22.1.70>
26. McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60(1), 63-91. <https://doi.org/10.1111/j.1744-6570.2007.00065.x>
27. Mehta, A., Adams, N., Fredrickson, M., Kraszkiewicz, W., Siy, J., Hamel, L., & Hendel-Paterson, B. (2021). Craving Empathy: Studying the Sustained Impact of Empathy Training on Clinicians. *Journal of Patient Experience*, 8, 1-8. <https://doi.org/2374373521104338310.1177/23743735211043383>
28. Melchers, M., Montag, C., Reuter, M., Spinath, F. M., & Hahn, E. (2016). How heritable is empathy? Differential effects of measurement and subcomponents. *Motivation and Emotion*, 40(5), 720-730. <https://doi.org/10.1007/s11031-016-9573-7>
29. Mehrabian, A., & Epstein, N. (1972). Measure of Emotional Empathy. *Journal of Personality*, 40(4), 525-543. <https://doi.org/DOI 10.1111/j.1467-6494.1972.tb00078.x>
30. Nasr Esfahani, M., Behzadipour, M., Jalali Nadoushan, A., & Shariat, S. V. (2014). A pilot randomized controlled trial on the effectiveness of inclusion of a distant learning component into empathy training. *Medical Journal of the Islamic Republic of Iran*, 28, 65. <http://www.ncbi.nlm.nih.gov/pubmed/25405130>

31. Patterson, F., Zibarras, L., & Ashworth, V. (2016). Situational judgement tests in medical education and training: Research, theory and practice: AMEE Guide No. 100. *Medical Teacher*, 38(1), 3-17. <https://doi.org/10.3109/0142159X.2015.1072619>
32. Paulus, C. M. (2009). *Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index*. <http://psydok.psycharchives.de/jspui/handle/20.500.11780/3343>
33. Paulus, C. M. (2021). *Is personal distress a part of empathy? A structural analysis of the multidimensional empathy concept*. <https://doi.org/10.22028/D291-33706>
34. Paulus, C. M., & Meinken, S. (2022a). The effectiveness of empathy training in health care: a meta-analysis of training content and methods. *International Journal of Medical Education*, 13, 1-9. <https://doi.org/10.5116/ijme.61d4.4216>
35. Paulus, C. M., & Meinken, S. (2022b). Empathy training for student teachers - design, structure and initial results. *Journal of Education and Human Development*, 11(2), 1-11. <https://doi.org/10.15640/jehd.v11n2a1>
36. Preusche, I., & Wagner-Menghin, M. (2013). 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). *Advances in Health Sciences Education*, 18(4), 573-587. <https://doi.org/10.1007/s10459-012-9393-9>
37. R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
38. Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36.
39. Saxena, S., Tripathi, S., & Madnavat, A. V. S. (2017). Comparison in the level of empathy in adolescent girls and boys. *Indian Journal of Health and Wellbeing*, 8(7), 765-767.
40. Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research - Online*, 8, 23-74.
41. Song, Y. M., Nie, T. T., Shi, W. D., Zhao, X. D., & Yang, Y. Y. (2019). Empathy Impairment in Individuals with Autism Spectrum Conditions from a Multidimensional Perspective: A Meta-Analysis. *Frontiers in Psychology*, 10. <https://doi.org/ARTN 190210.3389/fpsyg.2019.01902>

42. Stojiljković, S., Djigić, G., & Zlatković, B. (2012). Empathy and Teachers' Roles. *Procedia - Social and Behavioral Sciences*, 69, 960-966. <https://doi.org/10.1016/j.sbspro.2012.12.021>
43. Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
44. Weisz, E., Ong, D. C., Carlson, R. W., & Zaki, J. (2021). Building empathy through motivation-based interventions. *Emotion*, 21(5), 990-999. <https://doi.org/10.1037/emo0000929>
45. Wolf, D., Schneider, S., & Schütz, A. (2020). *SEC-I und SEC-SJT. Inventar und Situational Judgement Test zur Erfassung sozial-emotionaler Kompetenz(en)*. Otto-Friedrich-Universität Bamberg. <https://doi.org/10.20378/irb-48863>
46. Wu, Y., Lian, K., Hong, P., Liu, S., Lin, R.-M., & Lian, R. (2019). Teachers' Emotional Intelligence and Self-efficacy: Mediating Role Of Teaching Performance. *Social Behavior and Personality: an international journal*, 47(3), 1-10. <https://doi.org/10.2224/sbp.7869>
47. Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. (1992). Development of Concern for Others. *Developmental Psychology*, 28(1), 126-136. <https://doi.org/10.1037/0012-1649.28.1.126>
48. Ziaei, M., Oestreich, L., Reutens, D. C., & Ebner, N. C. (2021). Age-related differences in negative cognitive empathy but similarities in positive affective empathy. *Brain Struct Funct*, 226(6), 1823-1840. <https://doi.org/10.1007/s00429-021-02291-y>

Appendix 1

Adapted items of the Jefferson Scale of Empathy for Teachers (The numbers refer to the numbering in Preusche & Wagner-Menghin, 2013).

1. Teachers' understanding of their students' feelings and the feelings of their students' families does not influence pedagogical treatment.
2. Students feel better when their teachers understand their feelings.
3. It is difficult for a teacher to view things from a students' perspectives.
4. Because people are different, it is difficult to see things from a students' perspectives.
5. Attention to students' emotions is not important in history taking.
6. Attentiveness to students' personal experiences does not influence treatment outcomes.
7. Teachers should try to stand in their students' shoes when providing care to them.
8. Students value a teacher's understanding of their feelings, which is helpful in its own right.
9. Only pedagogical measures can solve students' school problems; therefore, teachers' emotional ties to their students do not have a significant impact on solving school problems.
10. Asking students about what is happening in their personal lives is not helpful in understanding their learning complaints.
11. Teachers should try to understand what is going on in their students' minds by paying attention to their non-verbal cues and body language.
12. I believe that emotion has no place in the treatment of school problems.
13. Empathy is a pedagogical skill without which the teacher's success is limited.
14. Teacher's understanding of the emotional status of their students, as well as that of their families is one important component of the teacher-student relationship.
15. Teachers should try to think like their students in order to render better care.
16. I believe that empathy is an important factor in pedagogy.

Appendix 2

Jefferson Scale of Empathy for Teachers (JSE-T)

(5-point Likert-type format, ranging from 1 = Strongly Disagree to 5 = Strongly Agree)

Items (German – Englisch) JSE-T	Item number in Preusche and Wagner-Menghin (2013)	Factor affiliation
1. Schüler fühlen sich besser, wenn ihr Lehrer ihre Gefühle versteht. - Students feel better when their teacher understands their feelings.	02	I
2. Lehrer sollten versuchen, sich in die Lage ihrer Schüler zu versetzen, wenn sie mit ihnen reden. - Teachers should try to put themselves in their students' shoes when they talk to them.	09	I
3. Es ist hilfreich, Schüler danach zu fragen, was gerade in ihrem Leben passiert, um ihre schulischen Probleme zu verstehen. - It is helpful to ask students about what is happening in their lives right now in order to understand their school problems.	12	I
4. Lehrer sollten versuchen zu verstehen, was in den Köpfen ihrer Schüler vorgeht, indem sie auf ihre nonverbalen Hinweise und Körpersprache achten. - Teachers should try to understand what is going on in their students' minds by paying attention to their nonverbal cues and body language.	13	I
5. Eine wichtige Komponente in der Beziehung zwischen dem Lehrer und den Schülern ist es, dass Lehrer für das emotionale Befinden der Schüler Verständnis haben. - An important component in the relationship between the teacher and students is for teachers to be understanding of students' emotional state.	16	I
6. Empathie ist eine pädagogische Fertigkeit, ohne die der Erfolg eines Lehrers eingeschränkt ist. - Empathy is an educational skill without which a teacher's success is limited.	15	II
7. Ich glaube, dass Empathie ein wichtiger Factor in der Pädagogik ist. - I believe that empathy is an important factor in pedagogy.	20	II
8. Nur pädagogische Maßnahmen können die Schulprobleme von Schüler lösen; emotionale Bindungen von Lehrern zu ihren Schülern haben deshalb keinen bedeutsamen Einfluss auf die Lösung schulischer Probleme. - Only pedagogical measures can solve students' school problems; therefore, teachers' emotional ties to their students do not have a significant impact on solving school problems.	11	III
9. Ich glaube, dass Gefühle keinen Platz bei der Lösung schulischer Probleme von Schülern haben. - I believe that feelings have no place in solving students' school problems.	14	III