

Empathie im Lehramtsstudium: Welchen Einfluss kann Service Learning haben?

Dissertation

zur Erlangung des akademischen Grades Dr. phil.
des Fachbereiches 05 – Psychologie und Sportwissenschaften
der Johann Wolfgang-Goethe-Universität Frankfurt am Main

vorgelegt von

Dipl. - Psychologin Kathrin Kuchta

Tag der Disputation: 06.11.2019

Gutachter:

Professor Dr. Holger Horz

Johann Wolfgang-Goethe-Universität Frankfurt am Main

Professor Dr. Andreas Gold

Johann Wolfgang-Goethe-Universität Frankfurt am Main

Zusammenfassung

Empathie ist ein mehrdimensionales psychologisches Konstrukt, das aus verschiedenen Facetten besteht (Decety & Ickes, 2011). Es ist anzunehmen, dass Empathie ein wichtiger Mechanismus ist, um Menschen miteinander zu verbinden und eine Gruppenkohäsion möglich zu machen (Rameson & Lieberman, 2009). Neben der Fähigkeit die Erlebenswelt des Gegenübers mit eigenen mentalen Repräsentationen nachzuvollziehen, werden dadurch Emotionen ausgelöst, die denen des Gegenübers sehr ähnlich sind. Gleichzeitig unterscheidet sich dieses Gefühlserleben aber beispielsweise von reiner Gefühlsansteckung, da eine Selbst-Andere Differenzierung stattfindet und in einer empathischen Episode immer im Vordergrund steht, dass man sich aufgrund der Gefühle des anderen so fühlt (Altmann, 2015). Hier spielt Imitation eine wichtige Rolle, wenn es darum geht, die Erlebenswelt der anderen Person zu erfassen (Meltzoff & Decety, 2003). Besonders auch bei Lehrkräften zeigt sich eine besondere Wichtigkeit von empathischem Handeln und Verstehen (Tausch & Tausch, 2008). In verschiedenen Studien zeigten sich positive Effekte von Empathie auf die Schülerschaft und die Unterrichtsqualität. Die SchülerInnen trauen sich mehr, es herrscht weniger Angst im Klassenzimmer und die Qualität der Unterrichtsbeiträge steigt (vgl. Tausch & Tausch, 1998). Empathie selbst besteht aus State- und Trait-Anteilen, so dass zumindest Teile davon trainierbar sind (Butters, 2010). Eine potentielle Möglichkeit um Empathie zu fördern scheint das Lehr-Lern-Format Service Learning darzustellen. Hierbei handelt es sich um ein Veranstaltungskonzept, bei dem ein meist fachlicher, akademischer Inhalt mit einem ehrenamtlichen Engagement außerhalb der Universität verknüpft wird (Reinders, 2016). Forschung aus dem angloamerikanischen Raum weist darauf hin, dass Empathie durch derartige Formate gefördert werden kann (Lundy, 2007; Wilson, 2011). Da die meisten Messverfahren von Empathie auf Selbstauskunft basieren und damit nur indirekt Anteile wie das affektive Mitschwingen abbilden können, war es Teil dieser Arbeit im ersten Schritt einen

objektiven, videobasierten Test zu entwickeln, der dann mit anderen Verfahren zur Messung eingesetzt werden sollte. In zwei ExpertInnen-Befragungen wurden aus einem Pool von 29 Videosequenzen mit Unterrichtssituationen erst vierzehn und dann final insgesamt zehn Videoclips mit jeweils vier Items und zugehörigen Antwortoptionen extrahiert. In einer darauf folgenden Validierung mit Studierenden der Goethe-Universität ($N = 112$) wurden diese Vignetten mit dem *Toronto-Empathy-Questionnaire* (Spreng, McKinnon, Mar & Levine, 2009), dem *Interpersonal Reactivity Index* (Davis, 1980), dem *Big-Five Inventory* (Rammstedt & John, 2005), dem *Essener Fragebogen zur Erfassung des subjektiven Empathiekonzepts* (Schönefeld & Roth, 2017) und dem Matrizenest des *Culture Fair Tests* (Weiß, 2006) gemeinsam erhoben und die Zusammenhänge analysiert. Die Reliabilitäten der drei Testscores bewegten sich in den beiden gebildeten Testversionen zwischen Cronbachs $\alpha = .53$ (Verhaltens-Score der Testversion 1) und $\alpha = .76$ (Intensitäts-Score der Testversion 2). Es zeigten sich zu allen Fragebögen erwartungskonforme Zusammenhänge von kleinen bis mittleren Effekten. Zudem zeigten sich Itemschwierigkeiten von 20 bis 88, bei den meisten Items allerdings zwischen 50 bis 65. Die Trennschärfen bewegten sich zwischen .18 bis .70. Wenn man sich die psychometrischen Werte nur für die Lehramtsstudierenden anschaut ($n = 10$) liegen die internen Konsistenzen zwischen $\alpha = .27$ (Verhaltens-Score der Testversion 1) und $\alpha = .85$ (Emotionserkennungs- und Intensitäts-Score der Testversion 2). In dieser Unterstichprobe liegen die Schwierigkeiten im Bereich von 23 bis 90 und Trennschärfen zwischen .10 und .90. Ebenfalls fanden sich Trennschärfen im negativen Bereich, so dass ExpertInnen-Nachbefragungen durchgeführt wurden.

Im nächsten Entwicklungsschritt wurden die Vignetten in neu zusammengestellten Testversionen nur Lehramtsstudierenden ($N = 41$) vorgelegt und zusätzlich Videoaufnahmen der Gesichter der ProbandInnen gemacht, um sie mit Face-Reader zu analysieren und die

Facette *Mitschwingen* abzubilden. Die Reliabilitäten der Testversionen lagen mit einem neuen Scoring nun zwischen $\alpha = .24$ (Emotionserkennungs-Score Prä-Testversion) und $\alpha = .57$ (Intensitäts-Score Prä-Testversion) sowie zwischen $\alpha = .10$ (Emotionserkennungs-Score Post-Testversion) und $\alpha = .77$ (Intensitäts-Score Post-Testversion). Auch die Schwierigkeiten und Trennschärfen änderten sich nach Adaptieren des Scorings und bewegten sich in beiden Testversionen nun von 30 bis 89 (Schwierigkeit) und von .0 bis .5 (Trennschärfe). Die Face-Reader Analysen zeigten nur in Teilen kongruente Emotionen mit den Selbstauskunftsdaten bzw. den eingeschätzten Intensitäten in den Videosequenzen, dann allerdings mittlere bis große Effekte, so dass in Teilen von einem affektiven Mitschwingen ausgegangen werden kann. Da sich die internen Konsistenzen im Vergleich zur Validierung verschlechterten, wurden die Zusammensetzungen der Testversionen für den Praxiseinsatz wieder auf die Validierungs-Versionen umgestellt.

Im Praxiseinsatz wurden Lehramtsstudierende in Service Learning (SL) und Non-Service Learning Veranstaltungen rekrutiert und miteinander verglichen. Insgesamt nahmen $N = 68$ Personen an drei Messzeitpunkten teil ($n = 30$ in Service Learning und $n = 38$ in Non-Service Learning Seminaren). Die Analysen (ANOVA über die Zeit und MANOVA im querschnittlichen Vergleich) zeigten, dass es zwischen den Gruppen keine signifikanten Unterschiede in den genutzten Instrumenten gab. Auch über die Zeit gab es nach der Bonferroni-Korrektur nur einen signifikanten Effekt ($F(2,52) = 6.57, p = .003, \eta^2 = .20$) in der SL-Gruppe von Studierenden mit naturwissenschaftlichem Fach auf die Subskala *Wertepluralismus* des *Inventars sozialer Kompetenzen* (Kanning, 2009). Alle anderen Maße wurden nicht signifikant. Es ist anzunehmen, dass diese Ergebnisse vor allem auf methodische Einschränkungen und Verbesserungsmöglichkeiten des entwickelten Testverfahrens zurückzuführen sind. Die SL-Veranstaltungen beschäftigten sich nicht explizit mit Empathie, so dass angenommen werden kann, dass eine bewusste Auseinandersetzung mit dem

Konstrukt ein wichtiges Kriterium zur Entwicklung darstellt und nicht als Nebeneffekt erworben werden kann. Zudem ist es so, dass gerade in der LehrerInnenbildung SL-Veranstaltungen den Pflichtpraktika und der Praxis sehr ähnlich sind, da auch diese engmaschig von Seminaren und Reflexionsmöglichkeiten begleitet werden. Daher ist es denkbar, dass diese Gründe einen maßgeblichen Einfluss auf die hier gefundenen (oder genauer: nicht-gefundenen) Effekte haben könnten.

Inhaltsverzeichnis

Kapitel 1: Einleitung	9
Kapitel 2: Theoretische Grundlagen	16
2.1 Empathie	17
2.1.1 Empathiemodelle	22
2.1.1.1 Kommunikationsmodell	23
2.1.1.2 Neuropsychologische Modelle	26
2.1.1.3 Empathie-Prozessmodell	28
2.1.2 Abgrenzung zu anderen Konstrukten.....	30
2.1.2.1 Perspektivenübernahme und Theory of Mind	30
2.1.2.2 Gefühlsansteckung	32
2.1.2.3 Mitleid und Mitgefühl	33
2.1.2.4 Prosoziales Verhalten	34
2.1.3 Messmethoden von Empathie	37
2.1.3.1 Carkhuff & Truax Empathy scales	37
2.1.3.2 Hogan Empathy Scale	39
2.1.3.3 The Questionnaire Measure of Emotional Empathy	42
2.1.3.4 Interpersonal Reactivity Index – IRI	45
2.1.3.5 Toronto-Empathy-Questionnaire – TEQ	49
2.1.3.6 Weitere Selbstauskunfts-Verfahren	52
2.1.3.7 Situative Tests	54
2.1.4 Geschlechtereffekte	56
2.1.5 Physiologische Korrelate	57
2.1.5.1 Spiegelneuronen	57
2.1.5.2 Imitation	59
2.1.6 Empathie als trainierbare Kompetenz	62
2.2 Service-Learning als Lehrmethode	67
2.2.1 Begriffsdefinition.....	68
2.2.2 Qualitätskriterien	72
2.2.3 Forschungsstand im Bereich Service Learning.....	76
2.2.3.1 Hauptbefunde und Darstellung der Forschung im deutschen Raum	76
2.2.4 Empathie und Service-Learning	81
2.2.4.1 Rahmenbedingungen für Empathieentwicklung.....	85
2.2.5 Service-Learning in der Lehramtsausbildung	86
2.3 Lehramtsausbildung und Lehrberuf	88
2.3.1 Aufbau der Ausbildung in Deutschland.....	88

2.3.2 Standards in der Lehrerbildung.....	88
2.3.3 Lehramtsstudium in Hessen und an der Goethe-Universität	90
2.3.4 Service-Learning im Lehramtsstudium an der Goethe-Universität	92
2.3.5 Professional Vision	94
2.3.6 Empathie und Lehrberuf	97
2.4 Forschungsleitende Annahmen	102
Kapitel 3: Testentwicklungsschritte	108
3.1 Erster Testentwicklungsschritt: Befragung der Empathie-ExpertInnen.....	110
3.1.1 Interne Pilotierung	110
3.1.2 Stichprobe	111
3.1.3 Aufbau	113
3.1.4 Datenbereinigung.....	115
3.1.5 Ergebnisse.....	116
3.1.5.1 Vignettenauswahl	116
3.1.5.2 Intra-Class-Correlation	119
3.2 Zweiter Testentwicklungsschritt: Expertenbefragung mit LehrerInnen	121
3.2.1 Stichprobe	121
3.2.3 Testaufbau.....	123
3.2.3 Datenaufbereitung.....	124
3.2.4 Ergebnisse.....	125
3.2.5 Auswertung und Implikation für die Weiterentwicklung des Vignettentestes	127
3.3 Dritter Testentwicklungsschritt: Validierung mit Studierenden	131
3.3.1 Stichprobe	131
3.3.2 Instrumente	133
3.3.2.1 Essener Fragebogen zur Erfassung des subjektiven Empathiekonzepts – ESEK	133
3.3.2.2 Toronto Empathy Questionnaire – TEQ.....	134
3.3.2.3 Interpersonal Reactivity Index – IRI	135
3.3.2.4 Culture Fair Test (CFT) – Untertest: Matrizen.....	135
3.3.2.5 Big Five Inventory (BFI-K).....	136
3.3.3 Datenaufbereitung.....	138
3.3.4 Ergebnisse.....	141
3.3.4.1 Itemanalysen.....	144
3.3.5 Auswertung und Implikation für die weiteren Testversionen.....	148
3.4 Vierter Testentwicklungsschritt: Pilotierung des finalen Testungsaufbaus mit Lehramtsstudierenden	151
3.4.1 Stichprobe	153
3.4.2 Testbatterie.....	156

3.4.2.1 Emotionale Kompetenz Fragebogen - EKF.....	157
3.4.3 Datenbereinigung.....	159
3.4.4 Ergebnisse.....	161
3.4.4.1 Itemanalyse.....	162
3.4.4.2 Korrelationen mit anderen Testmaßen	166
3.4.4.3 Mitschwingen	168
3.4.5 Auswertung der Pilotierung und Resümee der Testentwicklung.....	170
Kapitel 4: Praxiseinsatz des entwickelten Tests	173
4.1 Testung in SL-Veranstaltungen.....	174
4.1.1 Stichprobe	174
4.1.2 Durchführung.....	176
4.1.2.1 HEXACO-PI-R-24	180
4.1.2.2 Inventar Sozialer Kompetenzen – ISK	182
4.1.2.3 Aufbau Vignettentest.....	183
4.1.3 Datenaufbereitung.....	186
4.1.4 Ergebnisse.....	188
4.1.4.1 Prüfen der Vorannahmen und längsschnittliche Analysen.....	188
4.1.4.2 Querschnittliche Analysen.....	196
4.1.4.3 Mitschwingen	197
4.1.5 Diskussion.....	214
4.1.5.1 Längs- und Querschnittliche Analysen.....	214
4.1.5.2 Mitschwingen	219
Kapitel 5: Fazit der Untersuchungen	222
5.1 Fazit zur Untersuchung von Empathie in SL Veranstaltungen	223
5.2 Fazit zur Entwicklung einer Testbatterie zur Erfassung von Empathie	227
Literaturverzeichnis.....	234
Anhang	253

Tabellenverzeichnis

Tabelle 1 Übersicht der Rating-Stufen der Carkhuff & Truax Empathy Scales	38
Tabelle 2 Übersicht von Korrelationen zum Nachweis der Validität der Hogan Empathy Scales	41
Tabelle 3 Übersicht der Subskalen und Beispielitems des Questionnaire Measure of Emotional Empathy, Mehrabian & Epstein (1972).....	43
Tabelle 4 Interne Konsistenzen IRI (standardisierte Alpha Koeffizienten) getrennt nach Geschlecht der englischen Originalversion.....	46
Tabelle 5 Korrelationen der IRI-Skalen untereinander	47
Tabelle 6 Interne Konsistenzen (Cronbachs Alpha) der deutschen Version des IRI	48
Tabelle 7 Fragebögen-Grundlage des TEQ.....	50
Tabelle 8 Service-Learning Typisierung	69
Tabelle 9 Entwurf von Qualitätsstandards für SL im deutschen Hochschulbereich den K-12 Kriterien gegenübergestellt.....	75
Tabelle 10 Übersicht der Standards für Lehrerbildung	90
Tabelle 11 Aufbau Testversionen der Empathie-ExpertInnen	114
Tabelle 12 Von den ExpertInnen eingeschätzte Empathie-auslösende Vignetten	116
Tabelle 13 Auswahlkriterien für nicht Empathie-auslösende Vignetten zur Verwendung bei der weiteren Testentwicklung, Befragung Empathie-ExpertInnen	117
Tabelle 14 Aufbau Testversionen der LehrerInnen-Befragung.....	123
Tabelle 15 Übersicht Vignettenkennwerte LehrerInnenbefragung	128
Tabelle 16 Zusammensetzung Testversionen nächster Entwicklungsschritt.....	129
Tabelle 17 Interne Konsistenzen des ESEK.....	134
Tabelle 18 Interne Konsistenzen des BFI.....	137
Tabelle 19 Scoring des Vignettentests	140
Tabelle 20 Korrelationstabelle Vignetten und Empathie-relevante Maße (Validierung).....	142
Tabelle 21 Korrelationstabelle Vignetten und BFI-K sowie CFT (Validierung)	143
Tabelle 22 Vignettentestreliabilitäten, interne Konsistenzen (Validierung)	144
Tabelle 23 Vignettentestreliabilitäten, interne Konsistenzen (Validierung, nur Lehramtsstudierende)	147
Tabelle 24 Gemittelte Trennschärfen, nur Lehramtsstudierende (N=10).....	149
Tabelle 25 Testzusammensetzungen für die Testung im Sommersemester 2017	150
Tabelle 26 Interne Konsistenzen des Vignettentests der Pilotierungsstichprobe mit neuem Scoring.....	163
Tabelle 27 Korrelationen des Vignettentests Prä mit allen anderen Testwerten	167
Tabelle 28 Korrelationen des Vignettentests Post mit allen anderen Testwerten	167
Tabelle 29 Interne Konsistenzen des ISK, Subskala Soziale Orientierung	183
Tabelle 30 Übersicht Aufbau Vignettentests im Praxiseinsatz und interne Konsistenzwerte	184
Tabelle 31 Deskriptiv-statistische Auswertung der Testwerte in der Hauptstudie.....	185

Tabelle 32 Deskriptivstatistische Übersicht der Testwerte über die Messzeitpunkte in der EG (Praxiseinsatz)	190
Tabelle 33 Haupt- und relevante Nebeneffekte in der EG (Praxiseinsatz).....	191
Tabelle 34 Deskriptivstatistische Übersicht der Testwerte über die Messzeitpunkte in der KG (Praxiseinsatz)	193
Tabelle 35 Haupt- und relevante Nebeneffekte in der KG (Praxiseinsatz)	194
Tabelle 36 Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 2, Prä-Version	198
Tabelle 37 Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 2, Post-Version	199
Tabelle 38 Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 2, Prä-Version.....	200
Tabelle 39 Korrelationen der Face-Reader-Daten mit der im Videoclip eingeschätzten Intensität der Vignette 2, Post-Version	200
Tabelle 40 Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 6, Prä-Version	201
Tabelle 41 Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 6, Post-Version	202
Tabelle 42 Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 6, Prä-Version.....	203
Tabelle 43 Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 6, Post-Version	204
Tabelle 44 Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 21, Prä-Version	205
Tabelle 45 Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 21, Post-Version	206
Tabelle 46 Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 21, Prä-Version.....	207
Tabelle 47 Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 21, Post-Version	207
Tabelle 48 Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 25, Prä-Version	208
Tabelle 49 Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 25, Post-Version	209
Tabelle 50 Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 25, Prä-Version.....	210
Tabelle 51 Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion der Vignette 25, Post-Version	210
Tabelle 52 Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 30, Prä-Version	211
Tabelle 53 Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 30, Post-Version	212

Tabelle 54 Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 30, Prä-Version.....	213
Tabelle 55 Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion der Vignette 30, Post-Version.....	213

Abbildungsverzeichnis

Abbildung 1. Grafik des Empathieprozesses als Kommunikationsmodell	24
Abbildung 2. Schematische Darstellung des Empathie-Prozessmodells.....	28
Abbildung 3. Darstellung der Berufsgruppen der teilnehmenden EmpathieexpertInnen.....	113
Abbildung 4. Unterrichtete Fächer der LehrerInnen	122
Abbildung 5. Schwierigkeiten von Items 1 (Escore=Emotionsdekodierungsscore) und 2 (Iscore=Intensitätsscore) über alle Vignetten hinweg in der LehrerInnenbefragung	126
Abbildung 6. Trennschärfen von Item 1 und 2 über alle Vignetten hinweg in der LehrerInnenbefragung	127
Abbildung 7. Prozentuale Verteilung der Studienfächer der teilnehmenden Studierenden der Validierung.....	131
Abbildung 8. Fachsemester der Teilnehmenden der Validierung.	132
Abbildung 9. Itemschwierigkeiten (Validierung).....	144
Abbildung 10. Trennschärfen (Validierung)	145
Abbildung 11. Schwierigkeiten, nur Lehramtsstudierende (Validierung)	146
Abbildung 12. Trennschärfen nur Lehramtsstudierende (Validierung)	146
Abbildung 13. Aufbau der Testung in Präsenz ab dem Sommersemester 2017	152
Abbildung 14. Studierte Fächer der teilnehmenden Studierenden der Pilotierung im SoSe17.....	155
Abbildung 15. Ehrenamtliche Betätigungsfelder (geclustert) der Teilnehmenden der Pilotierung im SoSe17.....	156
Abbildung 16. Itemschwierigkeiten mit altem und neuem E-, sowie V-Scoring zum Messzeitpunkt Prä.....	164
Abbildung 17. Itemschwierigkeiten mit altem und neuem E-, sowie VScoring zu Messzeitpunkt Post	164
Abbildung 18. Trennschärfen des neuen und alten E-, sowie V-Scoring im Vergleich (Pilotierung SoSe17: Prä).....	165
Abbildung 19. Trennschärfen des neuen und alten E-, sowie VScoring im Vergleich (Pilotierung SoSe17: Post)	166
Abbildung 20. Verteilung der studierten Lehrformen der ProbandInnen.....	175
Abbildung 21. Studierte Fächer der ProbandInnen	176
Abbildung 22. Erhebungsdesign des Praxiseinsatzes.....	177
Abbildung 23. Operationalisierung von Empathie bei angehenden Lehrkräften beim Praxiseinsatz des Vignettentests	179
Abbildung 24. Operationalisierung empathischen Handelns bei angehenden Lehrkräften beim Praxiseinsatz des Vignettentests.....	179
Abbildung 25. Längsschnittliche Verläufe der Subskala Wertepluralismus des ISK innerhalb der EG und der Subgruppe, die ein naturwissenschaftliches Fach studieren (NW Cluster) versus denen, die kein naturwissenschaftliches Fach studieren (Kein NW-Cluster).....	192

Abbildung 26. Längsschnittliche Verläufe der Subskala Perspektivenübernahme des ISK innerhalb der EG und der Subgruppe, die ein naturwissenschaftliches Fach studieren (NW Cluster) versus denen, die kein naturwissenschaftliches Fach studieren (Kein NW-Cluster).....	192
Abbildung 27. Längsschnittliche Verläufe der Subskala funktional des ESEK innerhalb der KG und der Subgruppe Alter (mit Median-Split zur Einteilung in zwei Untergruppen).....	195
Abbildung 28. Längsschnittliche Verläufe der Subskala Erkennen der Emotionen anderer - EKF in der KG und der Subgruppe Alter (mit Median-Split zur Einteilung in zwei Untergruppen).....	195

Kapitel 1: Einleitung

„Schaun Sie, ich möchte ihnen helfen. Ich werde es auch tun. Ich weiß, wie ihnen zumute ist, wenn sie auch denken, ich hätte keine Ahnung aber glauben sie mir, ich weiß es. Ich versetze mich einfach in ihre Person.“ "Maggie, das ist Empathie." "Meine Güte. Das war das ganze Geheimnis? Warum haben sie das nicht gleich gesagt?"

Zitat aus „Ein süßer Fratz“, 1957.

Der Mensch ist ein soziales Wesen und steht in vielfältiger Interaktion mit anderen. Doch nicht jedes Zusammentreffen ist mit positiven Gefühlen verbunden. Neue Umgebungen und fremde Menschen sind erst einmal ein potentieller Stressor (Bartens, 2017) und es ist bekannt, dass Stress als Inhibitor von Empathie agiert (Martin et al., 2015). Im Umkehrschluss fördert das Blockieren von Stresshormon-Rezeptoren die emotionale Ansteckung (einen Teil von Empathie) gegenüber Fremden ebenso wie es kurze Episoden gemeinsamer Tätigkeiten vermögen, bei denen ähnliche Effekte auftreten. Aber warum ist die Qualität von menschlichen Beziehungen eigentlich so wichtig? Ist es nicht ausreichend einfach *nur* Beziehungen zu haben, egal wie sie gestaltet sind? Generell lässt sich sagen „Wer keine sozialen Beziehungen hat, wird häufiger und schwerer krank und stirbt früher.“ (Bartens, 2017, S. 43). Ein Blick in die Forschung zeigt, dass der positive Zusammenhang zwischen zufriedenstellenden sozialen Kontakten und Gesundheit, sowie Langlebigkeit unstrittig zu sein scheint (Danner, Snowdon & Friesen, 2001; Holt-Lunstad, Smith & Layton, 2010; Rosengren et al., 2004). Ebenfalls wurde festgestellt, dass sozialer Ausschluss mit der Aktivierung von Gehirnbereichen assoziiert werden kann, die auch bei physischen Schmerzen involviert sind (Eisenberger, Lieberman & Williams, 2003; Eisenberger & Lieberman, 2004). Der Ausspruch, dass Verlust und Loslassen von geliebten Personen *wehtut*, lässt sich also auch

neuropsychologisch gut zeigen und einleuchtend erklären. Im Gegenzug sind Menschen, die liebevoll mit anderen umgehen und ihre Wohlgesonnenheit auch körperlich beispielsweise durch Umarmungen zum Ausdruck bringen, entspannter und leiden weniger unter Stress (Floyd, Pauley & Hesse, 2010). Eine Erklärung dafür könnte wieder auf hormoneller Ebene gesehen werden. In Experimenten, bei denen ProbandInnen mittels Nasenspray Oxytocin verabreicht wurde, waren diese weniger stressanfällig und zeigten sich empathischer (Heinrichs, Baumgartner, Kirschbaum & Ehlert, 2003). Auch sozialer Ausschluss wurde in diesen Fällen als weniger aversiv erlebt. Die unangenehmen Folgen von sozialem Ausschluss konnten in einer Reihe von Experimenten gezeigt werden: Sozialer Ausschluss führte zu weniger proaktivem Verhalten. Nach einem sozialen Ausschluss trat eine Art *emotionale Taubheit* auf und die Fähigkeit zur empathischen Anteilnahme nahm ab bzw. war erst einmal nicht mehr vorhanden (Twenge, Baumeister, DeWall, Ciarocco & Bartels, 2007). Die Autoren erklären das Phänomen in Einklang mit den Ergebnissen von Eisenberger und Lieberman (2004), die zeigten, dass sich physischer und psychischer Schmerz gemeinsame Nervenbahnen teilen, dahingehend, dass diese Taubheit eine ähnliche Funktion haben könnte wie ein Schock nach physischer Verletzung, um die Schmerzen nicht spüren zu müssen. Dieses Phänomen zeigte sich nur temporär und wurde wieder schwächer, allerdings konnte weiterhin eine gewisse Vorsicht im Umgang mit anderen Menschen beobachtet werden. Eine hilfreiche Methode, um sich anderen anzunähern und durch Einfühlung eine Verbindung herzustellen, ist Empathie. Wie Rameson und Lieberman (2009) treffend formulierten

(...) empathy, in some sense, cuts right to the heart of what it means to be human: we are capable of not only recognizing the joys and sorrows of others, but also of reflecting on and sharing in the experience of those emotions. (S. 94)

Die Mechanismen, die Empathie zugrunde liegen und das Mitfühlen und Miterleben möglich machen, sind vielfältig. Durch teilweise unbewusste Verhaltensweisen, die das Gegenüber

imitieren (wie z.B. Kopfneigung, Muskeltonus oder Mimik und Gestik) scheinen wir es zu schaffen, uns einzufühlen und nachzuvollziehen wie es dem anderen geht (Dimberg, 1990). Doch was genau ist Empathie, wie wirkt sie und aus welchen Anteilen besteht sie? Und wie kann sie sich verändern und prosoziales Verhalten fördern?

Die Aufgabe, Prosozialität und konstruktives Konfliktverhalten zu fördern, sowie auf vielfältige Weise die Lebenswelt ihrer SchülerInnen zu verstehen, ist Kernsegment des Berufes von LehrerInnen. Diese sehen sich einer Vielzahl an Aufgaben gegenüber (Kultusministerkonferenz, 2014). Empathie kann dabei unterstützend sein und die Qualität von Lehren und Lernen verbessern (Tausch & Tausch, 1998). Mit der Förderung dieser Kompetenz sollte bereits im Studium begonnen werden. Viele Empathieanteile lassen sich durch Trainings fördern (Butters, 2010). Dadurch ergibt sich die Frage, welches Training zur Entwicklung von Empathiefacetten geeignet ist und sich im Lehramtsstudium implementieren lässt. Eine vielversprechende Veranstaltungsform scheint Service-Learning¹ zu sein (Reinders, 2016). Bei dieser Seminarform wird ein ehrenamtliches Engagement außerhalb der Universität mit einer fachlichen, akademischen Begleitung an der Universität verbunden. Somit können Lerninhalte direkt in der Praxis angewandt und die Erfahrungen wiederum vor dem theoretischen Hintergrund reflektiert werden (Eyler & Giles, 1999). Zudem wird ein realer Bedarf in der Gesellschaft adressiert und die Studierenden praxisorientiert auf ihre zukünftigen Tätigkeiten vorbereitet. Vieles spricht dafür, dass nicht nur akademische Kompetenzen durch dieses Lernen gefördert, sondern auch persönliche Entwicklung angestoßen werden kann. Bereits in den 1940er Jahren wurde in den USA von Projekten berichtet, bei denen Studierende für ehrenamtliches Engagement in der Gemeinde

¹ im weiteren SL abgekürzt

Creditpoints² für ihr Studium erhielten und in den 1960er Jahren wurde das Konzept dann als SL beschrieben (vgl. Reinders, 2016). Seit jeher herrscht in den USA ein anderes Verständnis und eine andere Notwendigkeit für ehrenamtliches Engagement (Community Service) als es in Deutschland der Fall ist (Pritchard, 2002). Das Ziel in den USA stellte dar, das Wohlbefinden der Gemeinde, des Anderen und auch das Eigene zu verbessern (Seifert, 2011). Dieses Vorhaben wurde auch in den 1980er Jahren von der amerikanischen Regierung gestützt, indem Ziele für die Bildung aufgesetzt wurden, in denen ausdrücklich die Stärkung von Citizenship, Community Service und persönlicher Verantwortung genannt und somit eine Brücke zur gesellschaftlichen Verantwortung geschlagen wurde (vgl. Pritchard, 2002). In dieser Zeit vereinigten sich auch einige der angesehensten Universitäten zu dem Netzwerk Campus Compact (2018, November 23), dem sich nach und nach immer mehr Colleges anschlossen (Reinders, 2016). Im Fokus lag nun zunehmend die gesellschaftliche Verantwortung der Bildungseinrichtungen und auch große Wirtschaftsfirmen wie Kelloggs unterstützten diese Bewegung (Pritchard, 2002). Dieser Trend setzte sich fort und mittlerweile hat Campus Compact mehr als 1000 Mitgliedscolleges und -universitäten. Im Laufe dieser Entwicklung stieß die Umsetzung dieser Veranstaltungen nicht immer auf Unterstützung. Für die Durchführenden gab es immer Hürden an ihren Hochschulen zu bewältigen, um sich Unterstützung von der administrativen Ebene zu sichern (Eyler & Giles, 1999). Erst mit neuen lerntheoretischen Erkenntnissen und der damit verbundenen Suche nach adäquaten Umsetzungsmöglichkeiten kam SL eine neue Bedeutung zu (Eyler & Giles, 1999; Seifert, 2011). Zudem bot dieses Format natürlich auch die Möglichkeit, soziale Probleme kennenzulernen und sie aktiv anzugehen (Eyler & Giles, 1999). Auch im deutschen bzw.

² im weiteren CP abgekürzt: Ein CP steht für 30 Stunden Beschäftigung mit einem bestimmten Themengebiet (sowohl in Präsenz in Seminaren, Vorlesungen o.ä., als auch im Eigenstudium, sowie Vor- und Nachbereitungen der Sitzungen und Anfertigen von Leistungsnachweisen wie Hausarbeiten, Lernportfolios, usw.)

europäischen Raum gab es einen Umschwung, so dass studierendenzentriertes Lernen mehr in den Fokus der Hochschullehre rückte (Biggs & Tang, 2011; EHEA, 2012). So wurde im Bologna Prozess neben der *Mobility* auch *Citizenship* in den Mittelpunkt der universitären Aufgaben gestellt (EHEA, 1999). *Mobility* zielte darauf ab, überall in Europa studieren und arbeiten zu können und *Citizenship* auf die Förderung der gesellschaftlichen und menschlichen Entwicklung von Studierenden, um ein Bewusstsein für gemeinsame Werte und Zugehörigkeit zu schaffen (Schubarth & Speck, 2013). Ebenfalls erweiterte sich dieser Anspruch auch auf *Employability*, womit Beschäftigungsfähigkeit im Sinne eines lebenslangen Lernens als Aufgabe der Hochschulen deklariert wurde (EHEA, 1999, 2003). Der Begriff *Employability* war ursprünglich eher negativ konnotiert und stammte aus dem Maßnahmenpaket namens *Youth at Risk*, welches ein Programm für junge Menschen mit schwieriger Familiengeschichte darstellte, die dabei unterstützt werden sollten ein selbstbestimmtes Leben führen zu können (Teichler, 2013). Auch der deutsche Begriff *Beschäftigungsfähigkeit* bezeichnete ursprünglich die Wiedereingliederung von Risikogruppen in den Arbeitsmarkt (HRK-Hochschulrektorenkonferenz, 2014; Rump & Eilers, 2006; Wolter & Banscherus, 2012). Die Bedeutung wurde allerdings neu definiert als der Begriff Einzug in den Bologna Diskurs hielt (Schubarth & Speck, 2013) und bezog sich nun vielmehr auf Aspekte lebenslangen Lernen (EHEA, 2007). Hiermit wird das Vorhandensein von Metakognitionen verstanden, die dabei hilfreich sind, Wissen zu strukturieren, sich aber auch neues Wissen anzueignen (Erpenbeck et al., 2006). SL wirkte und wirkt wie ein vielversprechendes Werkzeug dafür. Im deutschen Raum wurde die erste so deklarierte SL-Veranstaltung an Hochschulen 2003 an der Universität Mannheim durchgeführt (Reinders, 2016). Seitdem hat sich diesbezüglich viel getan und Bestrebungen sich zu organisieren sind erwachsen. Seit 2009 gibt es beispielsweise ein Netzwerk aus Hochschulen und Universitäten, die sich zusammengeschlossen haben, um gemeinsam das

Thema voranzutreiben (Hochschulnetzwerk BDV-Bildung durch Verantwortung, 2018b). Es werden regelmäßig Netzwerktreffen und Tagungen angeboten, um sich auszutauschen und Ideen zu entwickeln wie das Feld bereichert werden kann. Dem genannten Netzwerk, das seit 2015 ein eingetragener Verein ist, sind aktuell (Stand: Januar 2019) 39 Hochschulen aus dem deutschsprachigen Raum angeschlossen. Bereits vor dem Hochschulnetzwerk entstand das Netzwerk Bildung durch Engagement, das zum Ziel hatte, Schulen zu vernetzen, um SL in diesem Kontext zu etablieren. Doch ist dieses Format auch geeignet, um Empathiefacetten zu entwickeln? Und welche Empathiefacetten lassen sich dadurch entwickeln? Die folgende Arbeit wird diesen und weiteren Fragen auf den Grund gehen.

Kapitel 2: Theoretische Grundlagen

2.1 Empathie

Hinter dem Begriff *Empathie* verbergen sich viele Bedeutungen und Definitionen und ebenso viele verschiedene Konstrukte (Decety & Ickes, 2011). Der Duden (Dudenredaktion, 2017) beschreibt Empathie als „Bereitschaft und Fähigkeit, sich in die Einstellungen anderer Menschen einzufühlen“ und gibt als Synonyme *Einfühlungsvermögen*, *Feinfühligkeit*, *Fingerspitzengefühl*, *Mitgefühl*, *Teilnahme*, *Verständnis*, *Zartgefühl* an. Das Wort stammt ursprünglich von dem griechischen Wort *empátheia* (Leidenschaft, Zuneigung/Zuwendung) ab und hat im Englischen mit *empathy* einen ähnlichen Wortlaut. Im Deutschen trat es 1897 noch unter dem Begriff *Einfühlung* in Erscheinung (Lipps, 1897, zitiert nach Jahoda, 2005). Im Lateinischen wurde es unter dem Begriff *pathos* geführt, was mit *Empfindungswahrnehmung* übersetzt werden kann (Barrett-Lennard, 1981). Für Lipps (1897) bezog sich die Idee der Einfühlung auf Gegenstände der Natur oder Kunstgegenstände geschaffen von Menschenhand. Titchener (1912) führte in einem Artikel über die Methode der Introspektion das Wort *empathy* ein, als Darstellung einer Befragungstechnik, die mehr auf Einfühlung basierte und weniger auf nur direkt gestellte Fragen. Im Weiteren taucht das Konzept der Empathie noch unter uneinheitlichen Begriffen auf. Mead (1934) zum Beispiel spricht von *sympathy* und davon, dass sie durch die Erregung in einer anderen Person und nur durch die eigene Reaktion darauf entstehe. Dadurch handle es sich um einen gemeinschaftlichen Prozess, der von der anderen Person abhängig sei. Er geht in seinen Ausführungen auch darauf ein, dass Empathie bzw. Sympathie ein interaktiver und kooperativer Prozess sei und indirekt sogar, dass es sowohl affektive als auch kognitive Facetten besitze. Somit ist er der Erste, der den Weg für eine multidimensionale Definition von Empathie ebnet. Das kognitive Hineinversetzen in eine andere Person und gleichzeitige Mitfühlen, was sie fühlt, gehören zum Konstrukt Empathie dazu. In den Anfängen war das Verständnis des Einfühlens noch oberflächlicher, beispielsweise so wie ein Schauspieler sich

einer Figur nähert oder sich ein Autor in das zu beschreibende Objekt hineinversetzt. Diese Idee machte noch den Begriff *Einfühlung* von Lipps oder *empathy* von Titchener aus (Batson, 2011).

Heutzutage geht man davon aus, dass diese Prozesse weitestgehend unbewusst und automatisch ablaufen, allerdings auch bewusst gemacht werden können (Decety & Ickes, 2011). Personen nutzen ihr eigenes Wissen, um darauf aufbauend zu schlussfolgern, wie die Erlebenswelt der anderen Person aussehen könnte. Dabei kann es natürlich auch zu Fehlern kommen, wie Decety und Ickes (2011) anmerken:

The results of several experiments ... invite the conclusion that people Empathy and Knowledge Projection are quite likely to overestimate the extent to which what they know, or think they know, is known by others as well—which is to say that people are likely to consider their own knowledge to be more representative of the knowledge of others than it really is. (S. 52–53)

Generell lässt sich sagen, dass in entsprechenden Studien die Definition des Konstrukts Empathie wenn überhaupt meist sehr schwammig erfolgt und mittlerweile viele verschiedene Empathie-Modelle hervorgebracht wurden (Batson et al., 1987; zitiert nach Altmann, 2015). Seitens der Entwicklungspsychologie weiß man mittlerweile, dass es sich bei Empathie um eine soziale Fertigkeit handelt, die in der Kindheit gelernt wird (Bowlby, 1998; zitiert nach Butters, 2010). Das mimische Nachahmen von Gesichtsausdrücken ist etwas, das sogar schon in den ersten Lebensmonaten stattfindet und ein Teil davon ist, wie sich Menschen in die Erlebenswelt anderer einfühlen können (Hoffmann & Ludwig, 1982; zitiert nach Bonn, Helbrich, Lohan, Peindl & Schabert, 2008). Dieses Verhalten kann bei Kindern, Jugendlichen und Erwachsenen beobachtet werden und läuft weitestgehend automatisch ab (Hatfield, Cacioppo & Rapson, 1994). Obwohl davon ausgegangen wird, dass es für Empathie auch genetische Dispositionen gibt, scheint also vor allem die Sozialisation eine entscheidende

Rolle zu spielen (Zahn-Waxler, Schiro, Robinson, Emde & Schmitz, 2001). Wenn positive Wertschätzung erfahren wird, steigt die Wahrscheinlichkeit, dass diese auch anderen Personen gegenüber gezeigt wird (Eisenberg, Cumberland & Spinrad, 1998). Hierbei lassen sich bereits Geschlechterunterschiede erkennen. Mütter reden typischerweise mit ihren Töchtern öfter über Gefühle und psychisches Befinden, als sie es mit Söhnen tun, so dass diese das weniger gewöhnt sind und dadurch auch weniger Erfahrung mit dieser Art der Reflexion haben (Aznar & Tenenbaum, 2015). Das ist deswegen relevant, da es sich bei Empathie um einen interaktiven Prozess handelt, der im Kontext mit anderen entwickelt wird und somit erlern- und modifizierbar ist (Steibe, Boulet & Lee, 1979). Üblicherweise liegt der Beginn der Ausbildung der Fähigkeit zur Perspektivübernahme und moralischen Entwicklung zwischen ein bis zwei Jahren (Zahn-Waxler, Radke-Yarrow & Wagner, 1992). Ab diesem Alter lassen sich Verhaltensweisen beobachten, die andere beruhigen oder beschwichtigen sollen. Da die Fähigkeit zur Perspektivübernahme und für komplexe Denkprozeduren vor diesem Alter noch nicht entwickelt sind, haben alle emotionalen Reaktionen vor diesem Alter eher mit Imitation, emotionaler Ansteckung und ähnlichen Phänomenen zu tun (Eisenberg & Strayer, 1987). Dies lässt sich in Studien gut zeigen. Bei Experimenten mit Kindern im Alter zwischen 16-24 Monaten wurden Spiel- oder Essenssituationen geschaffen (Bischof-Köhler, 2009). Die Kinder spielten mit einer instruierten und ihnen schon bekannten erwachsenen Person, die Mütter saßen im Hintergrund. Während dieser Spielsituation gab es einen geplanten Unfall und das Spielobjekt – ein Teddy – verlor durch die Spielgefährtin einen Arm, die darüber deutlichen Kummer zeigte (weinen und schluchzen). Alternativ gab es die Situation, dass Kind und Spielgefährtin ein Dessert aßen und der Löffel der erwachsenen Person brach. Auf dem Tisch lag ein weiterer Löffel bereit und in beiden Versuchsanordnungen wurde nun festgehalten wie sich die Kinder verhielten. Die Reaktionsmuster ließen sich in vier Gruppen aufteilen:

1. helfende Kinder,
2. unbeteiligte Kinder, die einfach weiter machten,
3. Kinder, die sich durch die Gefühle anstecken ließen,
4. verwirrte Kinder, die nicht genau wussten, was sie tun sollten oder was das Weinen und Schluchzen bedeutete.

Die erste Gruppe wurde als empathisch eingestuft, Gruppe zwei und drei als selbstzentriert und Kinder aus der Gruppe vier waren in einer Übergangsphase. Die verschiedenen Reaktionsmuster sind laut der Studienleitung auf Entwicklungsunterschiede zurückzuführen und korrelieren hoch mit der Fähigkeit, sich im Spiegel erkennen zu können. Diese haben ein schon ausgebildetes Ich-Bewusstsein, erst ab dann wird eine Selbst-Andere-Differenzierung möglich. Bischof-Köhler (2009) spricht davon, dass diese Entwicklungen vor allem auf Reifungsprozessen beruhen und weniger mit Sozialisation zu tun haben. Allerdings können die Reaktionen auch auf verschiedene Bindungsstile zurückgeführt werden, so dass beispielsweise unsicher gebundene Kinder eher mit Gefühlsansteckung oder ungebunden reagierten und Kinder wenig empathischer Mütter sich auch ähnlich verhielten. Beim Zeigen von prosozialem Verhalten spielen allerdings auch körperliche Hormonprozesse eine Rolle. Es konnte gezeigt werden, dass Kinder und Jugendliche mit erniedrigten oder gestörten Cortisol-Spiegeln zu weniger empathischem, aber auffälligerem Sozialverhalten neigten (Cusi, Macqueen, Spreng & McKinnon, 2011; Shirtcliff et al., 2009). Zu vergleichbaren Ergebnissen kommt man bei Personen mit psychischen Symptomatiken und hochgradigen Depressionen (Cusi et al., 2011), die ebenfalls gestörte Cortisol-Level haben. Dabei scheint vor allem die Fähigkeit zur Perspektivenübernahme und der empathische Affekt betroffen zu sein.

Wie sich in all diesen Ansätzen zeigt, ist Empathie multidimensional und besteht aus kognitiven, sowie affektiven Aspekten, sowie veränderbaren (State-), als auch

unveränderbaren (Trait)-Anteilen und involviert andere Personen. Es ist möglich, Empathie als Fähigkeit zu definieren, da sie Kommunikation unterstützen kann (Altmann, 2015). Es gibt allerdings auch ForscherInnen, die Empathie nicht als Fähigkeit, sondern eher als vorgegebene genetische, kaum-veränderbare Disposition sehen (Bischof-Köhler, 2009). Vor allem PsychologInnen in helfenden Berufen (wie TherapeutInnen oder BeraterInnen) sind sich weitestgehend einig darüber, dass Empathie drei verschiedene Kompetenzen benötigt: Das Teilen der Emotionen des anderen, kognitiv schlussfolgern wie sich die andere Person fühlt (in diesem Zusammenhang wird auch von empathischer Akkuratheit gesprochen) und in Folge eine mitfühlende Reaktion auf den gezeigten Kummer (Decety & Ickes, 2011). Der Ausdruck bezeichnet das korrekte Schlussfolgern von Emotionen und der Erlebenswelt des Gegenübers (Decety & Ickes, 2011; Rogers, 1957). Dieser Prozess ist wie weiter oben angedeutet natürlich fehleranfällig, da nicht außer Acht gelassen werden darf, dass sich das Gegenüber bzw. die Erlebenswelt des Gegenübers durchaus von der eigenen unterscheiden kann.

Da es keinen Konsens über eine einheitliche Beschreibung von Empathie gibt, schlägt Altmann (2015) ausgehend von der Erklärung von Bischof-Köhler (2009) eine engere, sowie weiter zu verstehende Definition vor. In der engeren Definition fasst er die affektive Komponente im Sinne des emotionalen Mitschwingens als Kernstück von Empathie zusammen. In das etwas weitere Verständnis von Empathie fließen die kognitiven Aspekte der Perspektivenübernahme mit ein, um Fremd- von Eigenemotion trennen zu können und nicht in der Gefühlsansteckung verhaftet zu bleiben. Gleichzeitig entkräftet er allerdings auch die Annahme von Bischof-Köhler (2009), dass es sich bei Empathie um keine Fähigkeit handeln würde, dadurch, dass „... das oben beschriebene Einfühlen durchaus als Fähigkeit verstanden werden kann, die über die empathische Genauigkeit (der englische Fachbegriff ist ‚empathic accuracy‘; Ickes 1993) gemessen und damit in der Güte bewertet werden kann.“ (Altmann, 2015, S. 11). Im bisherigen Diskurs war es auch durchaus üblich Empathie als

Persönlichkeitsmerkmal zu betrachten. Es wird als „... eine nicht bewusste, mindestens mittelfristig stabile Verhaltensdisposition angesehen, in der sich Menschen unterscheiden, aber bei der jeder Wert auf der Dimension zunächst als normal im Sinne von nicht pathologisch gilt (Asendorpf und Neyer 2012).“ (Altmann, 2015, S. 11).

Perspektivenübernahme als Teil von empathischer Akkuratheit scheint eine Fähigkeit zu sein, die bereits im Alter von 14 Jahren umfassend ausgebildet zu sein scheint. In einer Studie von Haugen, Welsh und McNulty (2008) wurden Liebespaare zwischen 14 bis 21 Jahren dahingehend untersucht, ob sich Unterschiede in Kommunikation und Konfliktverhalten aufgrund unterschiedlich ausgeprägter empathischer Akkuratheit nachweisen ließen. Alle Personen konnten sehr zutreffend die Innenwelt ihres jeweiligen Partners wiedergeben und entgegen der Vorhersagen der AutorInnen gab es unabhängig vom Geschlecht keinen Anstieg in der empathischen Akkuratheit über das Alter hinweg. Somit konnte auf der interessierenden Variablen kein Unterschied zwischen den Gruppen festgestellt werden. Das spricht dafür, dass bereits in recht jungem Alter die Fähigkeit der Perspektivenübernahme umfassend entwickelt ist. Ebenfalls eng mit Empathie assoziiert scheinen Motivation und Interesse zu sein (Richter, 2009). So stellten Cheng, Meltzoff und Decety (2007) in einer Studie fest, dass eine empathische Reaktion auf essende Personen vom Hunger der Zuschauenden beeinflusst wurde und eine erhöhte Reaktion im Areal der Spiegelneuronen nachgewiesen werden konnte.

2.1.1 Empathiemodelle

Wie im vorherigen Abschnitt bereits angesprochen gibt es viele verschiedene Sichtweisen auf Empathie. Konsensfähig ist mittlerweile, dass es sich um ein multidimensionales Konstrukt handelt, das zu einem gewissen Anteil aus affektiven und kognitiven Facetten besteht. Wie genau und aus welchen Facetten und Anteilen sich Empathie vermutlich zusammensetzt, ist in der Forschung allerdings höchst unterschiedlich definiert. Rameson und Lieberman (2009)

sehen als einen Grund dafür, dass noch nicht ausreichend geklärt werden konnte, was genau sich bei Empathie in unserem Kopf abspielt. Der Versuch diese Wissenslücke zu schließen, spiegelt sich in den verschiedenen theoretischen Modellen wider, die im folgenden Kapitel vorgestellt werden.

2.1.1.1 Kommunikationsmodell

Empathie kann sehr einfach betrachtet als Kommunikationsprozess betrachtet werden, bei dem vor allem die interaktiven, interpersonellen Prozesse im Vordergrund stehen (Barrett-Lennard, 1981; Kunyk & Olson, 2001). „Empathie ist ... auch aus dyadischer Sicht beschreibbar als Ereignis, das erst durch das Zusammentreffen zweier Personen überhaupt existiert.“ (Altmann, 2015, S. 12). Allerdings lässt sich dieser Prozess auch in Abwesenheit einer anderen Person erleben allein, indem man sich geistig mit der Situation einer anderen Person beschäftigt (Preston & Waal, 2002). Empathie als Kommunikationsfähigkeit zu betrachten, hat in der Forschung Tradition. Diese Sichtweise hat vor allem Relevanz in helfenden Berufszweigen wie der Pflege oder der Medizin, bei denen auf dieses Modell zurückgegriffen wird, um Interaktionen während einer empathischen Episode zu beleuchten (Kunyk & Olson, 2001). In den meisten Fällen wird von einem drei-schrittigen Kommunikationsprozess ausgegangen:

1. Wahrnehmen der Emotion und der Situation der anderen Person.
2. Das Wahrgenommene verbal oder non-verbal zum Ausdruck bringen.
3. Das Gegenüber versteht das zum Ausdruck gebrachte in adäquater Weise (Barrett-Lennard, 1981; Kunyk & Olson, 2001).

Der Empathieprozess läuft dabei in Kreisläufen ab. Nach dem Feedback in Phase 3 kann er erneut starten. Zur Veranschaulichung hat Barrett-Lennard (1981) diesen Prozess graphisch dargestellt (siehe Abbildung 1).

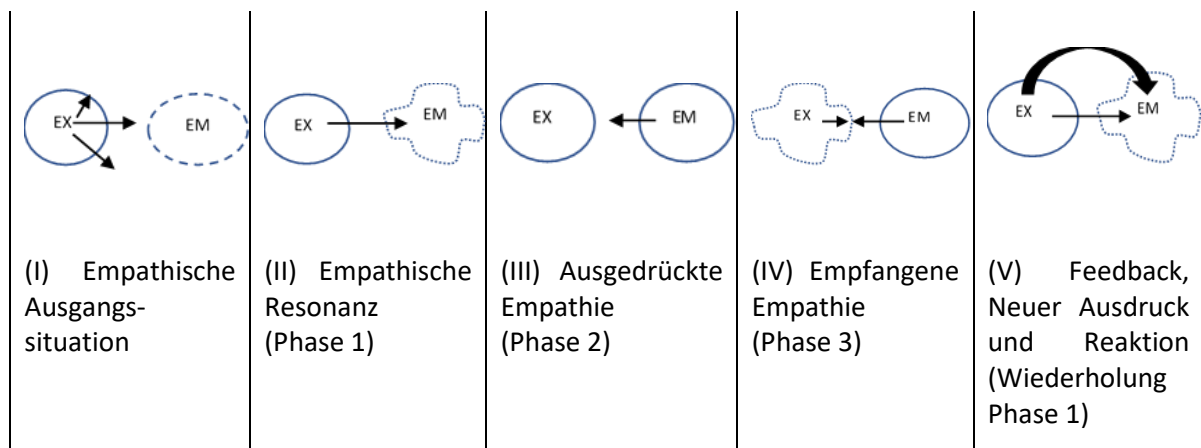


Abbildung 1. Grafik des Empathieprozesses als Kommunikationsmodell, übernommen von Barrett-Lennard, 1981, S. 94; Schematische Darstellung des empathischen Kreislaufes: EX=Person, die eine Emotion hat, die empathisch aufgenommen wird; EM=Person, die die gezeigte Emotion empathisch aufnimmt; gepunktete oder gestrichelte Linien symbolisieren die Bereitschaft Emotionen anzunehmen bzw. bereits erfolgtes Mitschwingen; die Pfeilrichtung zeigt die Richtung an, in die die Emotion gesendet bzw. kommuniziert wird.

Barrett-Lennard (1981) beschreibt den Prozess folgendermaßen:

Described conditions initiate this sequence, which has three main interior phases: (Phase 1) empathic resonance by Person A in response to B (who is personally expressive in some way), (Phase 2) A's attempt to convey his responsive understanding to B, and (Phase 3) B's actual reception/awareness of A's communication. When the process continues, Phase 1 is again the core feature, and 2 and 3 follow in cyclical mode. (S.91)

Empathie-Begriffe, die mit diesem Modell einhergehen und durch Forschung gemessen werden konnten, sind Beziehungs-Empathie, empathisches Verstehen und akkurate Empathie.

Dem Modell zugrunde liegend kann man zusätzlich von einer Metakommunikation ausgehen, bei der die Echtheit der Motive der Empathie-gebenden Person (Person A) für die Empathie-empfangende Person (Person B) eine entscheidende Rolle spielt. Hier entscheidet sich, ob sich Person B in ihren Bedürfnissen verstanden fühlt oder der Prozess erneut durchlaufen wird (Barrett-Lennard, 1981). Mit dem Aufbau aus Sender und Empfänger baut dieses Modell auf dem klassischen Kommunikationsmodell auf (Shannon, 1948). Empathie als interpersoneller Mechanismus kann zu einer Aufwertung der Kommunikation beitragen und Konflikte entschärfen (Haugen et al., 2008). Das zeigt sich auch bei einer Studie von Tamblyn et al. (2007), die in Kanada einen Zusammenhang feststellen konnten zwischen Kommunikationsfähigkeiten der Ärzten und Beschwerden der PatientInnen bei der Ärztekammer. Ärzte, die sich mit ihren PatientInnen beschäftigten und ihnen einfühlsam zuhörten, bekamen deutlich weniger Beschwerden als Ärzte mit geringen Kommunikationsfähigkeiten. In einer weiteren Studie mit Liebespaaren (Perrone-McGovern et al., 2014) konnte zudem festgestellt werden, dass Personen mit höheren Empathiewerten gegenüber ihren PartnerInnen häufiger konstruktive Konfliktlöse-Strategien nutzten. Die AutorInnen der Studie gehen davon aus, dass eine höhere Bereitschaft zur Perspektivenübernahme zu einem besseren Verständnis und letztendlich pro-sozialerem Verhalten führt. In dem Ansatz des Kommunikationsmodells lässt sich auch die Aufteilung von Empathie in State und Trait-Aspekte wiederfinden. Wie Kunyk und Olson (2001) beschreiben, braucht es bereits angeborene Fähigkeiten, allerdings auch eine Sensitivierung für die Gefühle des anderen und zusätzlich erlernte Möglichkeiten auf diese Wahrnehmungen auf eine empathische Art und Weise reagieren zu können, um Empathie adäquat zeigen zu können. Im Gegensatz zu anderen Modellen wird bei dieser Anschauung von Empathie bereits die Verhaltensantwort als Teil des Empathie-Prozesses mit angenommen. Ebenfalls spielt dafür das Alter eine Rolle, da einhergehend mit der Entwicklung des präfrontalen Cortexes

und erlernten sozialen Verhaltensregeln automatische, offen gezeigte empathische Antworten mit dem Alter abnehmen und eher von jüngeren Personen gezeigt werden (Preston & Waal, 2002).

2.1.1.2 Neuropsychologische Modelle

Weitere Empathie-Modelle, die mehr auf die intrapersonellen Vorgänge fokussiert sind, finden sich im neuropsychologischen Bereich. Einige ForscherInnen bieten ein Modell an, bei dem Empathie vorrangig als intrapersoneller Prozess betrachtet wird (Preston & Waal, 2002). Beim *perception-action model of empathy* wird es möglich, Empathie als übergeordneten Faktor zu sehen, dem Teilprozesse untergeordnet werden können (bspw. emotionale Ansteckung, kognitive Empathieanteile usw.) und der sowohl motorisches, als auch emotionales Verhalten beinhaltet. Ganz konkret wird in diesem Modell davon ausgegangen, dass eine Person die Emotionen bzw. den Zustand einer weiteren Person wahrnimmt und versteht. Dadurch würden automatisch eigene Repräsentationen mit dazu passenden Situationen aktiviert, die äquivalente autonome und somatische Reaktionen zum Zustand der anderen Person hervorrufen würden, welche allerdings auch blockiert werden könnten. Als wichtiger Auslösefaktor wird die Verbundenheit zu der anderen Person gesehen. Je näher und vertrauter man sich ist, desto eher werden die Prozesse ausgelöst und kongruente Repräsentationen aktiviert, um geeignete Verhaltensantworten zu erzeugen. Damit stimmt dieses Modell mit der *simulation theory of empathy* überein, bei dem ebenfalls durch das Wahrnehmen der Erlebenswelt einer anderen Person eigene Repräsentationen aktiviert werden, die ähnlich den wahrgenommenen Empfindungen sind (Rameson & Lieberman, 2009). Ebenso wie im *perception action model* postuliert, werden daraufhin autonome, körperliche Reaktionen ausgelöst, die denen der anderen Person sehr ähnlich sind. Im Unterschied dazu werden die ablaufenden Prozesse allerdings als unbewusst betrachtet und

das Teilen der Emotionen aufgrund der aufgebauten mentalen Repräsentationen gesehen. Auf diesen Modellen aufbauend legen Rameson und Lieberman (2009) dem empathischen Prozess ein *dual process Model* zugrunde. Sie gehen aufgrund neurophysiologischer Befunde davon aus, dass sich Empathie aus einem automatisch ablaufenden, affektiven und einem kontrolliert gesteuerten, kognitiven Anteil zusammensetzt. Sie regen an, diese Prozesse zukünftig genauer zu untersuchen, eine Forderung, der sich auch andere ForscherInnen anschließen (z.B. Jolliffe & Farrington, 2004; Spreng et al., 2009).

Ein weiteres umfassendes, neuropsychologisches Modell wird von Decety und Moriguchi (2007) vorgestellt. In diesem Modell wird von vier Hauptfaktoren ausgegangen, die Empathie möglich machen: 1. *Affective sharing*, 2. *Self-awareness*, 3. *Mental flexibility* und 4. *Emotion regulation*. Unter *affective sharing* versteht man (aufbauend auf dem *perception-action model*) einen automatischen Prozess, mit dessen Hilfe affektive Repräsentationen zwischen einem selbst und dem anderen geteilt werden. „Dieses Netzwerk erklärt also die emotionale Ansteckung (emotional contagion), beispielsweise die eigene Amüsiertheit, wenn jemand anderes lacht.“ (Altmann, 2015, S. 14). Der zweite Prozess beschreibt eine stabile Selbstaufmerksamkeit, so dass zwischen Eigen- und Fremdanteil unterschieden werden kann. Dieser intakte Automatismus hilft emotionale Ansteckung zu vermeiden. Die *mental flexibility* dient dazu, die eigenen Repräsentationen an den anderen anzupassen, um eine Perspektivenübernahme zu ermöglichen und die Geschehnisse bestmöglich aus den Augen des anderen nachvollziehen zu können. Und die *emotional regulation* schließlich macht es möglich, die selbst erlebten Gefühle zu regulieren. Damit wird zusätzlich die *Self-awareness* unterstützt, so dass persönliche Belastung geringer ausfällt und damit die emotionale Ansteckung vermindert wird.

An diesem Modell wird bemängelt, dass es zwar bereits umfassend sei, aber voneinander isolierten Prozessen ausgehe und keinerlei Aussagen über Wechselwirkungen oder ähnliches gemacht würden (vgl. Altmann, 2015).

2.1.1.3 Empathie-Prozessmodell

Im Versuch ein Modell zu konzipieren, das dem Kommunikations- und dem Prozessgedanken Rechnung trägt und somit die verschiedenen Instanzen vereint, wurde das Empathie-Prozessmodell entwickelt (Altmann & Roth, 2013). Grundlage bildet hierbei auch das Sender-Empfänger Modell (Shannon, 1948), so dass Empathie erst im Zusammentreffen zweier Menschen entstehen kann und einer der Personen einen emotionalen Ausdruck zeigen muss (Altmann, 2015). Dieses Modell geht ebenfalls von verschiedenen Stufen des empathischen Prozesses aus (siehe Abbildung 2).

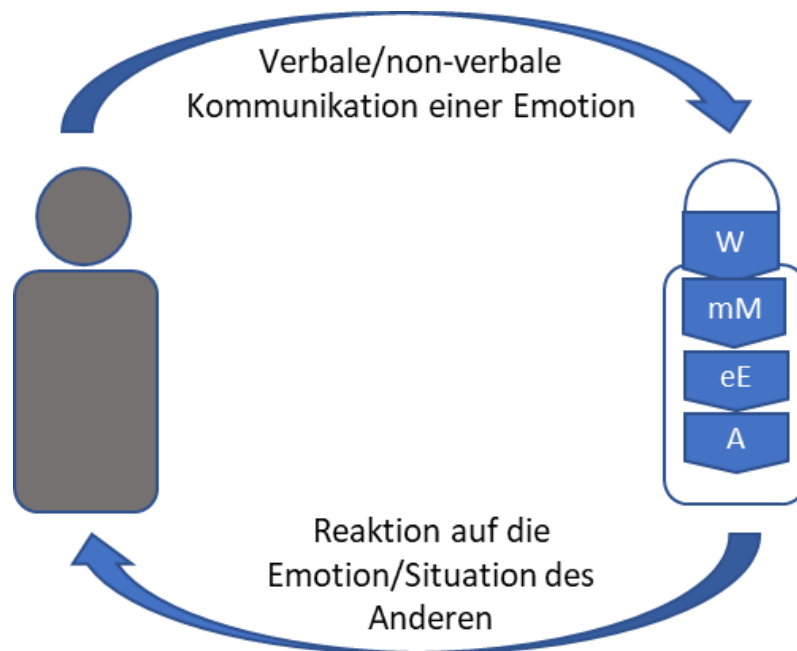


Abbildung 2. Schematische Darstellung des Empathie-Prozessmodells: Eine Empathie-Episode zwischen zwei Personen aus Senden, Empfangen, Verarbeiten und Reagieren auf eine kommunizierte Emotion innerhalb einer mindestens dyadischen Interaktion (W=Wahrnehmung, mM=mentales Modell, eE=empathische Emotion, A=Antwort); übernommen von Altmann, 2015, S.17.

Im ersten Schritt geht es auch wieder darum Informationen über den emotionalen Zustand des Gegenübers zu erlangen. Wie ist die Erlebenswelt des Gegenübers und welche Emotionen sind gerade präsent. Hierbei können emotionale (bspw. Schluchzen) und situative Informationen (z.B. eine Verletzung) unterschieden werden. Ebenfalls wichtig in dieser Sequenz sind Aufmerksamkeitsprozesse (welche Information wird tatsächlich aufgenommen). Im zweiten Schritt werden daran anknüpfend eigene mentale Repräsentationen aktiviert, die die erkannte Situation und Stimmung der anderen Person widerspiegeln. Dieses Prinzip spiegelt das Konzept der *Theory of Mind* und *Perspektivenübernahme* wider (siehe Kap. 2.1.2.1). Hier werden die wahrgenommenen Informationen verarbeitet und in Relation mit eigenen Erfahrungen sowie evtl. vorhandenen zusätzlichen Informationen gesetzt. Dieser Prozess läuft in den meisten Fällen unbewusst ab, ist allerdings bewusstseinsfähig (Altmann & Roth, 2013). Durch diese Verarbeitung werden in einem dritten Schritt Emotionen ausgelöst, die eine ähnliche Qualität und Intensität haben wie die des Gegenübers, um dann im weiteren Verlauf ein Verhalten initiieren zu können. Die Reaktion auf die zuvor wahrgenommene Situation kann sehr unterschiedlich aussehen: „...“, die Antwort, umfasst alle Arten von Reaktionen, sei es internal (z. B. Gedanken, Bewertungen) oder external (z. B. Gesichtsausdruck, verbale Kommunikation), solange sie sich auf die Situation der anderen Person beziehen.“ (Altmann, 2015, S. 22). Die verschiedenen Phasen lassen sich zusätzlich noch einmal in Trait-Aspekte und erlernbare Fertigkeiten einteilen: Die Verarbeitung der Situation des Gegenübers und das Einordnen in die aktivierte mentale Repräsentation kann als Fertigkeit verstanden werden, ebenso wie die an den Prozess angegliederten empathischen Antworten. Die Qualität der durch die Verarbeitung ausgelösten ähnlichen Emotion kann als Persönlichkeitsmerkmal angesehen werden, ebenso wie die Vielfalt und Vielschichtigkeit der mentalen Modelle selbst (Altmann, 2015). Es ist zu vermuten, dass sich diese auch interpersonell unterscheiden.

2.1.2 Abgrenzung zu anderen Konstrukten

Dadurch, dass es bislang keine eindeutige Empathiedefinition gibt, finden sich viele Begriffe und Konstrukte, die mit Empathie gleichgesetzt werden, von denen sie allerdings abzugrenzen ist. Teilweise handelt es sich um Teilprozesse von Empathie, die allerdings ohne die anderen Komponenten zu kurz greifen, oder es sind mit Empathie verwandte Phänomene, die allerdings keine Facetten davon darstellen. Die wichtigsten werden im folgenden Kapitel erläutert.

2.1.2.1 Perspektivenübernahme und Theory of Mind

Der Begriff der Perspektivenübernahme fällt eigentlich immer im Zusammenhang mit Empathie. Das Wissen, dass andere Menschen Gedanken, Gefühle und Einstellungen haben, die sich durchaus von den eigenen unterscheiden können, wird als *theory of Mind* bezeichnet (Premack & Woodruff, 1978; Rameson & Lieberman, 2009). Häufig ist eine Unterscheidung der beiden Begriffe kaum möglich und die Definition sehr nah beieinander (vgl. z.B. Richter, 2009), so dass sowohl Theory of Mind als auch Perspektivenübernahme in dieser Arbeit synonym genutzt werden. Auch bei Perspektivenübernahme wird ein multidimensionales Konstrukt zugrunde gelegt und es wird zwischen räumlich-visueller (Piaget & Inhelder, 1999), konzeptueller bzw. kognitiver sowie affektiver bzw. emotionaler Perspektivenübernahme unterschieden (Oswald, 1996; Richter, 2009; Underwood & Moore, 1982). Letztere beiden werden vor allem in Verbindung mit Empathie häufig genannt und sind im Zusammenhang mit empathischer Akkuratheit von Interesse (Decety & Ickes, 2011). Perspektivenübernahme kann als aktiver Prozess verstanden werden, um die Erlebenswelt einer anderen Person zu verstehen (Kerkenbusch, 2012). Immer wieder wird dabei auch von einer Art *Gedankenlesen* gesprochen (Hickok, 2015). Decety und Ickes (2011) beschreiben es folgendermaßen:

Empathic inference is the everyday mind reading that people do whenever they attempt to infer other people's thoughts and feelings. It is a concept that other writers address under such headings as 'mentalizing' or 'theory of mind' (Stone, 2006; Stone & Gerans, 2006). Empathic accuracy is the extent to which such everyday mind reading attempts are successful (Ickes, 1997, 2003). To put it simply, empathically accurate perceivers are those who are good at 'reading' other people's thoughts and feelings. (S.57)

Bischof-Köhler (2009) beschreibt diesen Prozess als einen rein rationalen Vorgang, bei dem kein emotionales Nachempfinden stattfindet, sondern man sich lediglich mental in die Position der anderen Person versetzt, um die Erlebenswelt des anderen nachvollziehen zu können. Im Prozess werden eigene mentale Repräsentationen aktiviert, die dann auf das Gegenüber projiziert werden, um so die Gedanken, Einstellungen und Gefühle des anderen zu verstehen (Decety & Ickes, 2011; Preston & Waal, 2002; Rameson & Lieberman, 2009). Es wird hierbei als Grundlage für Empathie gesehen (Batson et al., 1997; Batson, 1997) und als ein Teilaspekt dessen verstanden (Davis, 1980, 1983), bei dem affektive Inhalte fehlen (Bird & Viding, 2014). Zur Erfassung werden vor allem Methoden vorgeschlagen, in denen die Überzeugungen von Person A und B ungleich sind (Rameson & Lieberman, 2009). Dadurch ließe sich dann tatsächlich ableiten, wie gut bspw. Person A's Fähigkeiten zur Perspektivenübernahme sind je nachdem, wie nah die Repräsentation an Person B's Gedanken und Gefühlen ist. Der Prozess, die eigene Repräsentation an die vermutliche Erlebenswelt des anderen anzupassen, ist nämlich ein störanfälliger Vorgang. So berichten Decety und Ickes (2011) Befunde, nach denen Personen die Kenntnisse ihres Gegenübers in den Aspekten überschätzen, in denen sie selbst Wissen besitzen und die eigenen Fähigkeiten somit für repräsentativer halten als sie es eigentlich sind. In diesem Prozess können auch beispielsweise bei Personen mit Autismus größere Defizite beobachtet werden, die letztendlich vermutlich zum Fehlen empathischer Reaktionen führen (Bird & Viding, 2014). „They [Baron-Cohen,

1995; Baron-Cohen, Wheelwright, Skinner, Martin & Clubley, 2001] argue that severe autism can be characterized as mindblindness— an inability to accurately infer, or perhaps even to recognize the existence of, other people’s thoughts and feelings.“ (Decety & Ickes, 2011, S. 60). Außerdem scheinen soziale Einflüsse weniger relevant zu sein, als in der Entwicklung von Empathie (vgl. Richter, 2009).

2.1.2.2 Gefühlsansteckung

Gefühlsansteckung oder *emotional contagion* ist ein weiteres Konstrukt, das in engem Zusammenhang mit Empathie steht. Es handelt sich hierbei um einen automatischen, instinktiven Prozess, bei dem das Gefühl einer anderen Person übernommen wird (Steins, 2005). Diese Übernahme kommt vermutlich durch Imitationsprozesse zustande, bei denen unbewusst Mimik, Gestik oder Intonation des Gegenübers kopiert werden (Bird & Viding, 2014; Hatfield, Cacioppo & Rapson, 1993; Lamm, Decety & Singer, 2011; Rempala, 2013). Allerdings scheint es sich um eine abgeschwächte Version des Originalgefühls zu handeln (Hatfield et al., 1994). “However, some forms of primitive emotional contagion are far more subtle, automatic, and ubiquitous than previous theorists have supposed. (Hatfield et al., 1994)” (Hatfield et al., 1993, S. 97). Altmann (2015) beschreibt das Phänomen dahingehend, dass wenn bspw. in einer Sitcom Lachen eingespielt wird, dieses dafür sorgen soll, dass die Serie als lustiger empfunden wird, was tatsächlich der Fall ist. Es kann davon ausgegangen werden, dass Gefühlsansteckung vermutlich dazu dient einen Gruppeneinklang herzustellen und schnelle Handlungsfähigkeit in Gefahrensituationen zu gewährleisten (Bartens, 2017; Zillmann, 2004). Verantwortlich gemacht dafür werden die sogenannten Spiegelneuronen, die im Kapitel 2.1.5.1 ausführlicher beleuchtet werden. In früherer Forschung (ab Ende des 18. Jahrhunderts) wurde noch davon ausgegangen, dass für Gefühlsansteckung bewusste, kognitive Prozesse notwendig sind (Hatfield et al., 1993). Diese Sichtweise wird mittlerweile

nicht mehr geteilt (Altmann, 2013; Bird & Viding, 2014; Manera, Grandi & Colle, 2013). Vielmehr herrscht Einstimmigkeit darüber, dass “durch die Gefühlsansteckung (der englische Fachbegriff ist *emotional contagion*; Hatfield et al., 1994) werden Emotionen von einer Person automatisch (ohne intentionale Steuerung) auf eine andere übertragen.” (Altmann, 2015, S. 8). Gefühlsansteckung lässt sich schon bei Neugeborenen beobachten (vgl. Richter, 2009). Ein wesentlicher Unterschied zu Empathie besteht in der fehlenden sogenannten Selbst-Andere-Differenzierung (vgl. Altmann, 2015). Diese Selbst-Andere-Differenzierung sorgt dafür, dass die empfundene Emotion klar zugeordnet werden kann und die Ursache im anderen erkannt wird (Altmann, 2015; Corcoran, 1982; Bird & Viding, 2014). Gefühlsansteckung wird also als weiteres vermittelndes Element angesehen, um über eine kognitive Reflexion dieses Erlebens die Emotionswelt des Anderen besser verstehen zu können (Richter, 2009).

2.1.2.3 Mitleid und Mitgefühl

Zu weiteren Begriffen, die immer wieder synonym zu Empathie genutzt werden, obwohl sie davon abgrenzbar sind, gehören Mitleid und Mitgefühl (im englischen das Wort *sympathy*; Decety & Ickes, 2011). Die beiden Ausdrücke selbst lassen sich nicht weiter differenzieren und werden immer wieder deckungsgleich gebraucht. Es handelt sich hierbei um eine emotionale Reaktion, bei der die Ähnlichkeit der eigenen Emotionen mit denen des Gegenübers (im Sinne einer Gefühlsansteckung) nicht gegeben sein muss, sondern Sorge zum Ausdruck gebracht wird (Eisenberg, Shea, Carlo & Knight, 1991). Damit ist es vornehmlich von Anteilnahme geprägt und eher auf die negative Gefühlslage einer anderen Person bezogen (Richter, 2009). Mitgefühl scheint damit eine Zentrierung auf das Gegenüber zu beinhalten und mit gezeigtem, prosozialem Verhalten in Verbindung zu stehen (Decety & Ickes, 2011). Damit rückt das Mildern des unangenehmen Erlebens des anderen in den Vordergrund (Eisenberg et al., 1991; Bischof-Köhler, 2009). Somit würden sowohl empathische

Akkuratheit, als auch Selbstaufmerksamkeit bei Sympathie geringer werden (Wispé, 1986). Richter (2009) erläutert, dass Mitgefühl und Empathie verwandt sind, ersteres jedoch über Empathie hinausgehe und ein gezeigtes Verhalten zum Ziel hat - somit könne Mitgefühl im Anschluss an Empathie folgen. Die kategorialen Übergänge scheinen fließend zu sein, sind allerdings trennbar (Jahoda, 2005). Zudem zeigt sich, dass entgegen der empathischen Episode, bei der die Gefühlsqualität des empathischen Mitschwingens vom Gegenüber bestimmt wird, die Gefühlsqualität bei Mitgefühl immer recht ähnlich ist (Ulich & Volland, 1998). „Sympathie ist lediglich ein Mitgefühl, das man für das Schicksal eines anderen empfinden kann, ohne die Gefühle des anderen teilen zu müssen.“ (Cierpka, 2004, S. 38) . Dem schließt sich auch Eisenberg (1986) an, die auch die Sorge um die andere Person und die Unterschiedlichkeit der empfundenen Emotionen als entscheidendes Unterscheidungskriterium erachtet. Letzteres wird auch von anderen Forschern als wichtigstes Abgrenzungskriterium gewertet (Bird & Viding, 2014). Nur wenn die eigenen Emotionen und die des Gegenübers sehr ähnlich sind, ließe sich dann von Empathie sprechen. Wispé (1986) drückt es so aus, dass man sich im empathischen Prozess zum anderen hinwendet und bei Mitleid vom anderen bewegt wird. Damit gehe es bei Empathie um das Verstehen des anderen und bei Mitleid um das Wohlbefinden der anderen Person, beides auf die Emotionen des anderen bezogene Prozesse, aber mit unterschiedlichem Fokus.

2.1.2.4 Prosoziales Verhalten

Im Zusammenhang mit Empathie fällt auch immer wieder der Begriff prosoziales Verhalten, was im weitesten Sinne jedes Verhalten bezeichnet, das anderen Personen zu Gute kommt und freiwillig gezeigt wird (Bierhoff, 2005; Markus Wirtz, 2018). Hierbei wird häufig noch einmal anhand der Motivation unterschieden, ob sie das eigene Wohlbefinden zum Ziel hat (egoistische Motivation) oder das Wohlbefinden anderer (altruistische Motivation) (Batson,

Fultz & Schoenrade, 1987). Grundsätzlich unterstützt prosoziales Verhalten andere Personen und benötigt einen individuellen Einsatz (Twenge et al., 2007). Diese Art Verhalten sorgt für ein angenehmes Miteinander und kann die Zugehörigkeit zu einer Gruppe sichern, was viele positive Effekte mit sich bringt. Im Gegensatz dazu konnte gezeigt werden, dass es sich negativ auf prosoziale Verhaltenstendenzen auswirkt, wenn man aus sozialen Gruppen ausgegrenzt wird. In verschiedenen Experimenten, in denen prosoziales Verhalten unterschiedlich operationalisiert wurde (Geld spenden, Freiwilligendienst, Bereitschaft beim Aufräumen zu helfen, Kooperation in einem Spiel) konnten Twenge et al. (2007) zeigen, dass es in Abhängigkeit der Zugehörigkeit oder des Ausschlusses aus der jeweiligen Gruppe abnahm (mit einer Effektstärke von $d=1.49$, was einen starken Effekt anzeigt).

Neuere, umfangreiche Studien konnten außerdem zeigen, dass sich Prosozialität in drei verschiedene Subgruppen aufteilen lässt: 1. Altruistisch motiviertes prosoziales Verhalten (wirklich altruistisch), 2. Strategisch motiviertes prosoziales Verhalten (Kosten-Nutzen Erwägungen) und 3. Norm-motiviertes prosoziales Verhalten (durch soziale Normen bedingtes Verhalten) (Boeckler, Tusche & Singer, 2016). Es lässt sich beobachten, dass je höher die Kosten werden oder wenn die Unsicherheit ansteigt, welche Hilfe in einer bestimmten Situation angebracht wäre, die Wahrscheinlichkeit prosoziales Verhalten zu zeigen sinkt (vgl. Bierhoff, 2005). Prosoziales Verhalten wird zwar in den allermeisten Definitionen nicht als Teil von Empathie angesehen, sondern als begünstigende Bedingung (Bird & Viding, 2014; Jolliffe & Farrington, 2004), trotzdem zeigen sich immer wieder Zusammenhänge. In Studien konnte gezeigt werden, dass höhere Empathiewerte zudem mit konstruktivem Konfliktlöseverhalten in Zusammenhang stehen, was als ein unterstützendes Element prosozialen Miteinanders angesehen werden kann. Alexander (2001) konnte in einer Studie von Neunt- bis Zwölf-Klässlern zeigen, dass es eine moderate Beziehung zwischen Werten in Perspektivenübernahme bzw. empathischer Reaktion und Kompromiss suchendem

Konfliktlöseverhalten gibt. SchülerInnen mit hohen Werten zeigten häufiger kompromissbereites Verhalten gegenüber ihren Familien und MitschülerInnen. Wied, Branje und Meeus (2007) konnten ebenfalls ähnliche Zusammenhänge bei jungen Erwachsenen zeigen. Empathie kann also prosoziales Verhalten bewirken (Batson et al., 1991; Rameson & Lieberman, 2009). Allerdings variieren diese Zusammenhänge je nachdem wie Empathie erfasst wird (Eisenberg & Miller, 1987). Diese Anschauung wird von weiteren Studien gestützt (vgl. Jolliffe & Farrington, 2004).

Die Forschung konnte außerdem zeigen, dass prosoziales Verhalten in der Kindheit einen reliablen Prädiktor für gezeigtes Verhalten im jungen Erwachsenenalter darstellt (Eisenberg et al., 1999). Eine verbreitete Theorie prosozialen Verhaltens vermittelt die Empathie-Altruismus-Hypothese von Batson (1987), in der die Auffassung vertreten wird, dass Empathie immer altruistisches, prosoziales Verhalten fördert und egoistisch motivierte Prosozialität nur in Situationen erfolgt, in denen keine Empathie stattfand. Diese Ansicht wird allerdings immer wieder in Frage gestellt (Cialdini, Brown, Lewis, Carol & Neuberg, 1997; Neuberg, Cialdini, Brown, Luce et al, 1997). Batson (1987) wird kritisiert, die im verborgenen wirkenden Prozesse wie bspw. soziale Erwünschtheit nicht ausreichend erfasst zu haben (Neuberg et al., 1997). Es wird außerdem eine Verbindung von prosozialem Verhalten zu dem Persönlichkeitsmerkmal *Verträglichkeit* gesehen (Bierhoff, 2005). Unstrittig scheint auch zu sein, dass es in diesem Konstrukt, sowohl genetische Dispositionen, als auch sozial erlernte Anteile gibt (vgl. Bierhoff, 2005; Matthews, Batson, Horn & Rosenman, 1981). Zusammenfassend lässt sich sagen, dass prosoziales Verhalten ein von Empathie unterscheidbares Phänomen ist, jedoch ein hoher Zusammenhang gegeben zu sein scheint.

2.1.3 Messmethoden von Empathie

Den verschiedenen Messmethoden liegen unterschiedliche und diverse Verständnisse von Empathie zugrunde. Einige dieser Messverfahren sind im Kontext psychotherapeutischer Arbeit entstanden, im Versuch TherapeutInnen-PatientInnen Interaktionen und die Therapie-Ergebnisse besser zu verstehen (z.B. Rogers, 1975; Truax, 1972). „Scales have been developed in an attempt to measure transsituational dispositional differences in empathic responsiveness (Davis, 1983; Mehrabian & Epstein, 1972), and the construct measured by such scales has often been called empathy.” (Batson et al., 1987, S. 21). Im Folgenden wird die Entwicklung einiger bekannter Empathie-Messverfahren vorgestellt und das zugrunde liegende Empathieverständnis erläutert.

2.1.3.1 Carkhuff & Truax Empathy scales

Das entwickelte Verfahren zur Messung von empathischer Akkuratheit entstand im TherapeutInnen-PatientInnen-Setting und wurde aufgrund dieses Anwendungskontextes häufig in Trainingsevaluationen eingesetzt (Butters, 2010; Truax, 1972). Es handelt sich bei dieser Skala um ein Ratingverfahren, anhand dessen Interviews bezüglich der zu erkennenden empathischen Akkuratheit eingeschätzt werden können. Neben Empathie galten Wärme und Aufrichtigkeit als die wichtigsten Attribute eines Psychotherapeuten (Truax, 1972). Im Jahr 1961 entwickelte Truax eine 9-stufige, vorläufige Skala zur Messung von Empathie, die von Carkhuff verfeinert und auf 5-Stufen reduziert wurde, um empathisches Verstehen in interpersonellen Interaktionen zu erheben. Die verschiedenen Stufen sehen wie folgt aus (siehe Tabelle 1):

Tabelle 1
Übersicht der Rating-Stufen der Carkhuff & Truax Empathy Scales

Level 1: Low Level of Empathic Responding	<ul style="list-style-type: none"> • Communicating little or no awareness or understanding of the caller's feelings • Responses are irrelevant or abrasive • Changing the subject, giving advice, etc.
Level 2: Moderately Low Level of Empathic Responding	<ul style="list-style-type: none"> • Responding to the surface message of the caller but omitting feelings or factual aspects of the message. • Inappropriately qualifying feelings (e.g., "somewhat," "a little bit," "kind of") • Inaccurately interpreting feelings (e.g., "angry" for "hurt," "tense" for "scared"). • Level 2 responses are only partially accurate, but they show an effort to understand
Level 3: Interchangeable or Reciprocal Level of Empathic Responding	<ul style="list-style-type: none"> • Verbal and nonverbal responses at level 3 show understanding and are essentially interchangeable with the client's obvious expressions, accurately reflecting the client's story and surface feelings or state of being
Level 4: Moderately High Level of Empathic Responding	<ul style="list-style-type: none"> • Somewhat additive, accurately identifying the client's implicit underlying feelings and/or aspects of the problem. • Volunteer's response illuminates subtle or veiled facets of the client's message, enabling the client to get in touch with somewhat deeper feelings and unexplored meanings and purposes of behavior. • Level 4 responses thus are aimed at enhancing self-awareness.
Level 5: High Level of Empathic Responding	<ul style="list-style-type: none"> • Reflecting each emotional nuance, and using voice and intensity of expressions finely attuned to the client's moment-by-moment experiencing, the volunteer accurately responds to the full range and intensity of both surface and underlying feelings and meanings • Volunteer may connect current feelings and experiencing to previously expressed experiences or feelings, or may accurately identify implicit patterns, themes, or purposes. • Responses may also identify implicit goals embodied in the client's message, which point out a promising direction for personal growth and pave the way for action. • Responding empathically at this high level facilitates the client's exploration of feelings and problems in much greater breadth and depth than responding at lower level"

Anmerkung. Entnommen von MacDonald, 2018 (nach Patterson, 1985).

In diesem Kontext umfasst empathische Akkuratheit „both a sensitivity to what the patient is currently feeling or experiencing, and the verbal facility to communicate this understanding in a language attuned to the patient's current feelings.“ (Truax, 1967, S. 1781). Das Ratingverfahren ist so konzipiert, dass es für die Bewertung ausreichend ist, in psychotherapeutischen Settings nur den Therapeuten zu hören, um valide Aussagen machen zu können (Truax, 1972). Da es hier vor allem um die Einschätzung der Qualität der interpersonellen Interaktion geht, lässt sich das zugrunde-liegende Empathieverständnis als Kommunikationsmodell ansehen (siehe Kap. 2.1.1.1). Ein positiver Aspekt dieser Messmethode ist, dass es sich um eine Fremdeinschätzung handelt, die sich mit dem empathischen Ausdruck beschäftigt (Butters, 2010) und kein Selbst-Auskunftsverfahren ist, das aufgrund sozialer Erwünschtheit zu Verzerrungen führen kann (Schmidt-Atzert & Amelang, 2012).

2.1.3.2 Hogan Empathy Scale

Die Entwicklung der Hogan Empathy Scale basiert auf einem moralischen Konstruktverständnis, für das empathisches Verhalten als Ausdruck dient (Hogan, 1969), da moralische Entwicklung entscheidend dafür angesehen wurde, prosoziales Verhalten zeigen zu können. Dieser Definition liegt das Empathieverständnis von Mead (1934) zugrunde, welches Empathie als einen Teil sozialer Intelligenz versteht. Das finale Instrument besteht aus 64 Items, die in verschiedenen Bewertungsschleifen aus zwei verschiedenen Instrumenten extrahiert bzw. selbst entwickelt wurden und zur Einschätzung eine *Wahr-Falsch* Skala hat. 31 Items wurden dem *Minnesota Multiphasic Personality Inventory* (MMPI; Hathaway & McKinley, 1943) entnommen (ein Persönlichkeitsfragebogen, der hauptsächlich für den klinischen Sektor und Störungsdiagnostik gedacht ist), 25 Items aus dem *California*

Psychological Inventory (CPI - einen Persönlichkeitsfragebogen mit rund 20 Skalen, der auf dem MMPI beruhte; Gough, 1956) und 8 Items wurden selbst entwickelt. Aus dem 100-Item *California Q-Sort-Test* (Block, 1961) wurden 50 Items ausgewählt, die inhaltlich Empathie zuzuordnen waren. Die Grundlage für die Auswahl lieferten 14 Nicht-PsychologInnen (Laien), denen eine Empathie-Definition vorgelegt wurde und die darauf aufbauend ihre Definition einer sehr empathischen Person mit Hilfe des Q-Sort Tests abgeben sollten (Hogan, 1969). Als Gegenprobe bekamen neun fortgeschrittene Psychologiestudierende (ExpertInnen) dieselbe Aufgabe. Die Korrelation der jeweils ausgewählten 50 Items lag bei .86, so dass die AutorInnen davon ausgingen, dass es eine große Übereinstimmung bzgl. des gesuchten Konzeptes zwischen Laien und ExpertInnen zu geben schien. In weiteren Befragungsrunden wurden noch die Konzepte einer *guten* und einer *geistig gesunden* Person erhoben, die mit .65 geringer ausfielen, so dass Hogan (1969) es als erwiesen ansah, dass es sich bei Empathie um ein gut abgrenzbares Konstrukt handelte, das sowohl Laien als auch ExpertInnen ähnlich definierten. Darauf aufbauend wurde eine Skala entwickelt, die hoch empathische und wenig empathische Gruppen unterscheiden sollte, um allgemein Empathie messen zu können. Im Entwicklungsprozess wurden die Itemantworten hoch-empathisch-, sowie wenig-empathisch eingeschätzter Gruppen miteinander verglichen (Hogan, 1969). Für den finalen Test wurden insgesamt 64 Items ausgesucht. Diese dadurch entstandene Skala wiederum wurde mit verschiedenen anderen Verfahren korreliert, um Angaben über die Validität machen zu können. Die eingesetzten zusätzlichen Verfahren umfassten das *CPI* (Gough, 1956), den *MMPI* (Hathaway & McKinley, 1943), und den *Chapin Social Insight Test*, der die Fähigkeit einer Person erfassen soll, wie gut er/sie eine zwischenmenschliche Situation analysieren kann. Das Verfahren zeigte die größten Korrelationen mit Skalen des CPI (zwischen .20 bis .24). Positive Korrelationen gab es ebenfalls mit dem CPI, aber auch mit dem Chapin Test. Korrelationen in negativer Richtung gab es mit dem MMPI. „These correlations suggest that

empathy ratings define a dimension which includes social competence, intellectual promise, and feelings of self-worth.“ (Hogan, 1969, S. 309). Damit sieht er die Ähnlichkeit zu dem zugrunde-liegenden Empathieverständnis als Teil sozialer Intelligenz als gegeben an. Zu weiteren Validitätsprüfungen wurden noch folgende Verfahren herangezogen, die in Tabelle 2 aufgelistet sind.

Tabelle 2

Übersicht von Korrelationen zum Nachweis der Validität der Hogan Empathy Scales

Verfahren	<i>r</i> Hogan Empathy Scale
Myers-Briggs Type Indicator (Myers, 1962) – geht auf die psychologische Typenbildung von Jung zurück und ist aufgrund mangelhafter Konstruktion wissenschaftlich umstritten	-.61 (Introversion) bis .63 (Extraversion)
Study of Values (Allport, Vernon & Lindzey, 1951) – einer der frühesten Fragebögen zur Messung persönlicher Werte	-.15 (Theoretical) bis .19 (Social)
Ego Strength scale (Barron, 1954) – ein im Psychopathologischen Bereich eingesetztes Verfahren, um Personen auszumachen, die von verbaler Psychotherapie profitieren könnten	.34 bzw. .38 (Testung in zwei Gruppen)
Manifest Anxiety scale (Taylor, 1953) – Fragebogen zur Identifikation von Personen mit Angststörungen	-.49 bzw. -.20 (Testung in zwei Gruppen)
Edwards' scale for Social Desirability (Edwards, 1957) – Fragebogen zum Aufdecken sozial, erwünschter Antworttendenzen	.50 bzw. .37 (Testung in zwei Gruppen)
Neuroticism and Extraversion (Eysenck, 1959)	-.28 (Neuroticism) und .51 (Extraversion)
California F scale (Adorno, Frenkel-Brunswik, Levinson & Sanford, 1950; Bass, 1955) – Persönlichkeitstest, um die Tendenzen von faschistischen Tendenzen zu messen	-.52 (männliche Probanden) und -.30 (weibliche Probandinnen)
Dogmatism scale (Rokeach, 1960) – Persönlichkeitsfragebogen zur Messung von Autoritarismus	-.31 (männliche Stichprobe)

In einer späteren Untersuchung von Cross und Sharpley (1982) konnte gezeigt werden, dass die interne Konsistenz des Tests mit $\alpha=.61$ unter den empfohlenen und weithin anerkannten Richtlinien von mindestens $\alpha=.70$ für Persönlichkeitstests liegt (Moosbrugger & Kelava, 2007). Es zeigte sich auch, dass insgesamt 13 Items negativ mit dem Testwert korreliert waren und 30 Items überhaupt nicht mit dem Testwert korrelierten, so dass der Test kritisch gesehen werden kann. Davis (1983) schränkt ein, dass es sich bei dieser Skala eher um ein Messinstrument handelt, das kognitive Empathiefacetten misst.

2.1.3.3 The Questionnaire Measure of Emotional Empathy

Ein weiterer Selbstauskunftsfragebogen zur Messung von Empathie wurde von Mehrabian und Epstein (1972) entwickelt. Ihr Empathieverständnis grenzen sie von kognitiven Modellen zur Rollenübernahme ab und fokussieren sich auf den Ansatz der *affektiven Reaktivität*. Sie sehen als Teil davon nicht nur das Teilen, sondern auch das Wahrnehmen unangenehmer Gefühlsqualitäten. Ebenfalls stellen sie die These in den Raum, dass anti-aggressives und helfendes Verhalten mit dem Konzept affektiver Empathie verknüpft sein müsste. „... a person who has a high level of emotional empathy is less likely to engage in aggressive behavior, ... , and ... he is more likely to engage in helping behavior when he notices distress in another.“ (Mehrabian & Epstein, 1972, S. 526). Das Instrument besteht aus 33 Items, die verschiedene Subskalen abbilden und auf einer Likert Skala von +4 (*sehr starke Zustimmung*) bis -4 (*sehr starke Ablehnung*) bewertet werden. Die Subskalen und beispielhafte Items sind in Tabelle 3 abgebildet.

Tabelle 3
 Übersicht der Subskalen und Beispielitems des *Questionnaire Measure of Emotional Empathy*,
 Mehrabian & Epstein (1972)

Subskala	Items
1. Susceptibility to Emotional Contagion	<ul style="list-style-type: none"> • (10) The people around me have a great influence on my moods. (—) • (20) I am able to remain calm even though those around me worry. (—)
2. Appreciation of the Feelings of Unfamiliar and Distant Others	<ul style="list-style-type: none"> • (15) Lonely people are probably unfriendly. (—) • (28) Becoming involved in books or movies is a little silly. (—)
3. Extreme Emotional Responsiveness	<ul style="list-style-type: none"> • (8) Sometimes the words of a love song can move me deeply. • (23) Sometimes at the movies I am amused by the amount of crying and sniffing around me. (—)
4. Tendency to Be Moved by Others' Positive Emotional Experiences	<ul style="list-style-type: none"> • (14) I like to watch people open presents. • (22) Another's laughter is not catching for me. (—)
5. Tendency to be Moved by Others' Negative Emotional Experiences	<ul style="list-style-type: none"> • (16) Seeing people cry upsets me. • (30) I become more irritated than sympathetic when I see someone's tears. (—)
6. Sympathetic Tendency	<ul style="list-style-type: none"> • (26) It is hard for me to see how some things upset people so much. (—) • (33) Little children sometimes cry for no apparent reason. (—)
7. Willingness to be in Contact with Others Who Have Problems	<ul style="list-style-type: none"> • (12) I would rather be a social worker than work in a job training center. • (21) When a friend starts to talk about his problems, I try to steer the conversation to something else. (—)

Anmerkung. (—) = invertierte Items.

Als Testwert wird der Summenwert der Items gebildet. Die finalen Items wurden aus einem größeren Itempool gewählt. Auswahlgründe waren nicht vorhandene Korrelationen mit einer Skala zur Sozialen Erwünschtheit (Crowne & Marlowe, 1960), signifikante Korrelationen mit den Items der jeweiligen Skala und die Ergebnisse einer Faktorenanalyse. In zwei

Experimenten wurden die Thesen überprüft. Im ersten Experiment, das im Aufbau Ähnlichkeiten zu dem bekannten Experiment von Milgram (1965) aufwies, gab es eine Lehrer-/SchülerInnen Situation, in der der *Lehrer* (ProbandIn) den *Schüler* (eingeweihter Akteur) mit Elektroschocks bestrafen konnte, wenn dieser falsch antwortete. Die vermeintlich vergebenen Elektroschocks wurden allerdings genauso wie die adäquaten Schmerzäußerungen nur simuliert bzw. gespielt. Die Lehrkraft konnte selbst entscheiden, in welcher Intensität die Elektroschocks vergeben werden sollten. Es gab in diesem Setting zwei Bedingungen. In Bedingung 1 war der Schüler in Sichtweite und im selben Raum und in Bedingung 2 in einem anderen Raum und lediglich zu hören. Es zeigte sich, dass hoch empathische ProbandInnen in Bedingung 1 weniger aggressiv reagierten (in Form der Auswahl der Stärke des Elektroschocks) als wenig empathische ProbandInnen. Der Empathiegrad wurde über den *Questionnaire Measure of Emotional Empathy* erhoben. An einem zweiten Experiment nahmen nur Frauen teil, die eingangs einen Fragebogen zu ihren Einstellungen bzgl. verschiedener Themen ausfüllen sollten und im Anschluss noch eine Fragebogenbatterie vorgelegt bekamen. Anhand des Einstellungs-Fragebogens wurde ein falscher Fragebogen erstellt, der angeblich einer Person (eine Komplizin der Versuchsleitung) gehörte, die eine sehr ähnliche Einstellung (Bedingung 1) oder eine sehr unterschiedliche Einstellung (Bedingung 2) hatte. Der Fragebogen wurde der Probandin ausgehändigt, so dass sie sehen konnte, dass die unmittelbar am Experiment teilnehmende Frau ähnliche oder unterschiedliche Einstellungen hatte. Beide sollten gemeinsam Musik hören und während die Versuchsleitung den Raum verließ, berichtete die Komplizin von einem persönlichen Problem, für das sie Unterstützung bräuchte. Sie fragte die Probandin, ob sie ihr helfen könne. Unabhängig von den Bedingungen zeigte sich, dass die empathische Tendenz gemessen über den Fragebogen ein Prädiktor für das helfende Verhalten war und die vermeintliche Ähnlichkeit zur Komplizin keinen Einfluss auf das Verhalten hatte. Die Autoren sehen darin ihre Hypothesen bestätigt und ein starkes

Argument dafür, dass ihr Fragebogen reliabel affektive Empathie misst (Mehrabian & Epstein, 1972). Davis (1980) kritisiert, dass obwohl die Autoren nur affektive Empathie messen wollten, er durchaus Items sehe, die auch kognitive Anteile abfragten. Deswegen sieht er es auch als problematisch an, dass lediglich ein einziger Empathiescore gebildet würde. Aus diesem multidimensionalen Verständnis heraus, entwickelte Davis einen Fragebogen, der seit der Entwicklung das mit am meisten genutzte Verfahren zur Messung von Empathie darstellt.

2.1.3.4 Interpersonal Reactivity Index – IRI

Der IRI ist ein Selbstauskunftsfragebogen zur Erfassung von Empathiefacetten (Davis, 1980). Er bedient sich als einer der ersten Verfahren (im Gegensatz zur Hogan Empathy Scale und The Questionnaire Measure of Emotional Empathy) einer multidimensionalen Sichtweise auf Empathie, enthält 28 Items (neun davon in invertierter Form) und gliedert sich in vier Subskalen: *Perspective Taking (PT)*, *Empathic Concern (EC)*, *Fantasy (FS)* und *Personal Distress (PD)*. Jede Skala besteht aus sieben Items und der Testwert wird als Mittelwert jeder Skala gebildet. PT bezieht sich auf die Fähigkeit, kognitiv die Perspektive anderer Personen einzunehmen und psychologisch nachzuvollziehen, wie sich die Erlebenswelt der anderen Person gestaltet. Bei FS können sich Personen mit hohen Werten sehr mit fiktiven Charakteren aus Büchern, Filmen oder Theaterstücken identifizieren. Die Items der Skala EC fragen die Fähigkeit ab, warme Gefühle und Mitgefühl anderen gegenüber zu empfinden. Personen mit hohen Werten in der PD-Skala tendieren dazu, mit Ängstlichkeit und Unwohlsein zu reagieren, wenn sie bei ihrem Gegenüber negative Gefühle wahrnehmen. Die finalen Items des IRI wurden aus einem Pool von 50 Items extrahiert, die eigens entwickelt oder aus anderen Empathie-Messverfahren entlehnt wurden (unter anderem aus Mehrabian & Epsteins, 1972, emotional empathy scale oder Stotland's Fantasy-Empathy scale; Davis, 1980). Bei den selbstentwickelten Items wurde Wert daraufgelegt, dass sie sowohl kognitive als auch affektive

Facetten von Empathie enthielten. Die Items werden auf einer 5-stufigen Likert Skala von *beschreibt mich überhaupt nicht* (1) bis *beschreibt mich sehr gut* (5) bewertet. Diese 50 Items wurden einer Stichprobe von 452 Personen vorgelegt (251 Frauen, 201 Männern) und getrennt nach Geschlecht in einer ersten Faktorenanalyse geprüft. Als Ergebnis zeigten sich bereits hier die vier Dimensionen, die auch in der finalen Testversion vorhanden sind. In einem zweiten Schritt wurde eine Version mit 45 Items erstellt, die aus den Items der ersten Version, überarbeiteten Items der Vorgängerversion und drei neuen Items bestanden. Die umformulierten und neuen Items orientierten sich inhaltlich an den Items, die auf den Subskalen der ersten Untersuchung die höchste Ladung aufwiesen. Bei einer weiteren Faktorenanalyse zeigten sich erneut die vier Faktoren und die Items luden sowohl in der männlichen, als auch in der weiblichen Teilstichprobe auf ähnliche Weise. In das finale Instrument flossen die Items ein, die bei beiden Geschlechtern ähnlich hoch luden bzw. bei einem der Geschlechter besonders hoch waren. Items, die auf mehreren Faktoren hoch luden wurden exkludiert. In einer weiteren Testung mit über 1000 ProbandInnen ($N=1161$) bestätigte sich nochmals die Faktorenstruktur, die gute bis akzeptable Werte der internen Konsistenzen standardisierter Alpha Koeffizienten ergaben (siehe Tabelle 4).

Tabelle 4

Interne Konsistenzen IRI (standardisierte Alpha Koeffizienten) getrennt nach Geschlecht der englischen Originalversion, entnommen aus Davis, 1983

Subskala	Cronbachs α	
	Männlich	Weiblich
Fantasy (FS)	.78	.75
Perspective Taking (PT)	.75	.78
Empathic Concern (EC)	.72	.70
Personal Distress (PD)	.78	.78

Die Retest-Reliabilität liegt je nach Dimension und Geschlecht der ProbandInnen zwischen .61 (PT, männliche Probanden) und .81 (FS, weibliche Probanden). Wie sich durch diese

Trennung andeutet, konnte ein Geschlechtereffekt beobachtet werden, bei dem Frauen signifikant höhere Werte zeigten als Männer ($F(1,12) = 96.28; p < .001$; Davis, 1980). Die verschiedenen Skalen selbst sind nur moderat miteinander korreliert (vgl. Tabelle 5).

Tabelle 5

Korrelationen der IRI-Skalen untereinander, entnommen aus Davis, 1983

Subskala	Pearsons' s <i>r</i>					
	PT		EC		PD	
	Männlich	Weiblich	Männlich	Weiblich	Männlich	Weiblich
Fantasy (FS)	.10	.12**	.30**	.31**	.16**	.04
Perspective Taking (PT)	-	-	.33**	.30**	-.16**	-.29**
Empathic Concern (EC)			-	-	.11**	.01

Anmerkungen. **= signifikant < 0.01 ; männlich $N=582$, weiblich $N=587$.

Die in dieser Arbeit genutzte Übersetzung wurde von Neumann et al. (2012) ins Deutsche übersetzt. Hierbei wurden die Items von drei Übersetzern so wörtlich wie möglich ins Deutsche übertragen und von einem professionellen, muttersprachlichen Übersetzer wieder ins Englische übersetzt. Die Items, die dem englischen Original am ähnlichsten waren, wurden in die deutsche Version aufgenommen. Die finale Version wurde nochmals ins Englische übersetzt und vom Entwickler Davis zur Nutzung freigegeben. Beispielitems sind „Ich habe Tagträume und stelle mir recht regelmäßig Dinge vor, die mir passieren könnten.“ (Subskala Fantasy), „Ich empfinde oft warmherzige, sorgende Gefühle für Leute, denen es weniger gut geht als mir.“ (Subskala Empathic-Concern), „Mir fällt es manchmal schwer, Dinge aus der Sicht einer anderen Person zu sehen.“ (Subskala Perspective-Taking, inverses Item) und „In Notfallsituationen fühle ich mich ängstlich und unbehaglich.“ (Subskala Personal-Distress). Die internen Konsistenzen wurden getrennt nach Professionen bzw. Studienfächern berechnet und sind in Tabelle 6 zu sehen.

Tabelle 6

Interne Konsistenzen (Cronbachs Alpha) der deutschen Version des IRI, entnommen aus Neumann et al., 2012

Subskala	Cronbachs α	
	Medizinstudierende	Studierende anderer Disziplinen
Fantasy (FS)	.74	.78
Perspective Taking (PT)	.76	.76
Empathic Concern (EC)	.69	.62
Personal Distress (PD)	.70	.70

Auch wenn der IRI eines der meistgenutzten Verfahren zum Erfassen von Empathie darstellt (Jolliffe & Farrington, 2004; Paulus, 2009), gibt es immer wieder Kritik an ihm. Zum einen wird das mit dem Fragebogen abgebildete Empathiekonzept in Frage gestellt, da einige Studien die Multidimensionalität als nicht adäquat abgebildet ansehen (Albiero, Matricardi, Speltri & Toso, 2009; vgl. Butters, 2010). Zum Anderen konnte die 4-Faktoren Struktur in einigen Studien nicht repliziert werden, was vermutlich darauf zurückzuführen ist, dass die Studien nicht mit studentischen Stichproben durchgeführt wurden (Cliffordson, 2002; Lauterbach & Hosser, 2007; Siu & Shek, 2005). Es scheint also fraglich wie weit der Fragebogen eingesetzt werden kann. Ebenfalls sind die Reliabilitätswerte häufig unzureichend (Beven, O'Brien-Malone & Hall, 2004; Ireland, 1999; Paulus, 2009). Bei verschiedenen Übersetzungen wurde zusätzlich die Itemanzahl angepasst, um eine bessere Faktorenladung zu erhalten (Paulus, 2009; Siu & Shek, 2005). Außerdem wurde kritisch bemerkt, dass bei der Konstruktion des IRI nicht ausreichend zwischen Empathie und Sympathie unterschieden wurde, obwohl es sich um unterschiedliche Konstrukte handle (Jolliffe & Farrington, 2006). Des Weiteren fasst Butters (2010) zusammen, dass zwar verschiedene Empathiefacetten erhoben werden, aber ein großer Fokus auf Perspektivenübernahme liegt und weniger auf dem Erkennen der Emotionen anderer, was bezogen auf kognitive Empathiefacetten damit zu kurz kommt.

2.1.3.5 Toronto-Empathy-Questionnaire – TEQ

Da der *IRI* weiterhin unzureichende Ergebnisse lieferte, gab es Bemühungen reliablere Empathie-Messverfahren zu entwickeln. Ein Ansatz dazu kam von Spreng et al. (2009), die einen neuen Selbstauskunftsfragebogen mit 16 Items entwickelten - den *TEQ*. Dem Instrument liegt das Verständnis von Empathie als primär emotionalem Prozess zugrunde. Die Items werden auf einer 5-stufigen Likert Skala von *nie* (1) bis *immer* (5) beantwortet. Das Verfahren wurde aus einer Vielzahl bislang bestehender Instrumente zur Messung von Empathie entwickelt. Die Items wurden statistisch mit Hilfe von Faktorenanalysen ausgewählt. Ziel war es einen Fragebogen zu entwickeln, der nicht nur wie die bisher bestehenden einzelne Teile von Empathie abbildete, sondern das Konstrukt möglichst umfassend beschrieb. Aus verschiedenen Selbstauskunftsfragebögen (Tabelle 7) wurden hierfür Items entnommen.

Tabelle 7
Fragebögen-Grundlage des TEQ; entnommen aus Spreng et al., 2009

Fragebogen	Anzahl entnommene Items
Interpersonal Reactivity Index (Davis, 1983)	28 Items
Hogan´s empathy Scale (Hogan, 1969)	15 Items
The Questionnaire Measure of Emotional Empathy (Mehrabian & Epstein, 1972)	9 Items
Reworded Balanced Emotional Empathy Scale (Mehrabian, 2000)	12 Items
Scale of Ethnocultural Empathy (Wang et al., 2003)	4 Items
Jefferson Scale of Physician Empathy (Hojat et al., 2001)	6 Items
Nursing Empathy Scale (Reynolds, 2000)	8 Items
Japanese adolescent Empathy Scale (Hashimoto & Shiomi, 2002)	10 Items
Measure of Emotional Intelligence (Schutte et al., 1998)	3 Items
Dysexecutive Questionnaire (Burgess, Alderman, Evans, Wilson & Emslie, 1996)	4 Items
Measure of emotion comprehension (Hornak, Rolls & Wade, 1996)	7 Items

Zusätzlich wurden noch weitere 36 Items entwickelt, die auf Basis der Literatur zu Personen mit Läsionen oder Störungen Probleme mit empathischem Verhalten zeigten, erstellt wurden. Insgesamt wurden 142 Items untersucht, die zuvor von einer studentischen Stichprobe ($N=200$) ausgefüllt wurden. Mit Hilfe einer exploratorischen Faktorenanalyse - (Hauptachsenanalyse) - und einer gewollten Ein-Faktorenlösung/Einfachstruktur wurden 55 Items aus jeder Skala identifiziert, die auf einem Faktor mit einer Ladung von mindestens .40 luden. Da das Ziel ein kurzes Verfahren sein sollte, wurden die Items nochmals auf insgesamt 16 Items reduziert, die aufgrund einer Ladung von $>.40$ ausgewählt wurden (Mehrabian & Epstein, 1972). Acht Items liegen invertiert vor. Alle inkludierten Items stammten zu gleichen

Anteilen aus verschiedenen Skalen, sowie den neu entwickelten Items. Die internen Konsistenzen zeigten mit $\alpha = .85$ gute Werte und die Retest-Reliabilität zeigte mit $r = .81$ eine hohe Stabilität. Es wurden zur Überprüfung von diskrimanter und konvergenter Validität noch weitere Instrumente vorgelegt. Positive und auch erwartete Korrelationen gab es mit der Subskala *Empathic Concern* des IRI (Davis, 1983) mit $r = .74$, die auf einem ähnlichen Niveau blieb auch wenn die vier umformulierten Items, die im TEQ enthalten sind, herausgenommen wurden ($r = .71$). Auch mit der IRI Skala *Perspective Taking* konnte ein mittlerer Zusammenhang festgestellt werden ($r = .35$). Negative Korrelationen ($r = -.30$) gab es erwartungskonform mit dem *Autism Quotient* (Baron-Cohen, Wheelwright, Skinner et al., 2001), bei dem vor allem die Einschätzung des eigenen Verhaltens im sozialen Kontext und bei Kommunikationsprozessen abgefragt wird und entsprechend Personen mit hohen TEQ-Werten eher niedrige Werte in diesem Verfahren erhielten. In einer weiteren Untersuchung mit 79 Studierenden fanden sich mittlere Korrelationen von $r = .35$ mit dem *Reading the Mind in the Eyes Test-Revised /MIE* (Baron-Cohen, Wheelwright, Hill, Raste & Plumb, 2001) und $r = .23$ beim *Interpersonal Perception Task-15/IPT-15* (Costanzo & Archer, 1994). Der MIE ist ein Test zur Mentalisierung, d.h. dem Verstehen der mentalen Zustände des eigenen und des Verhaltens anderer. Er besteht aus 36 Bildern mit Augenpartien von Schauspielern, denen mentale Haltungen zugeordnet werden sollen. Richtige Antworten in diesem Test legen nahe, dass die Probanden dazu in der Lage sind die mentale Haltung anderer zu verstehen und non-verbale Indikatoren zuordnen zu können. Beim IPT-15 geht es auch um die Bewertung nonverbaler Signale diesmal allerdings anhand von Videosequenzen, zu denen Multiple Choice Fragen mit *richtigen* Antworten gestellt werden. Hierbei geht es z.B. um das Einordnen von Familienzugehörigkeit („Zu welchem Erwachsenen gehört das Kind.“). Hohe Werte werden als Indikator für interpersonelle Sensitivität und soziale Kompetenz gesehen (Spreng et al., 2009). Die hier genutzte deutsche Version des TEQ wurde von einer Forschungsgruppe

an der Universität Duisburg-Essen rund um Professor Dr. Andreas Roth ins Deutsche übertragen (Schönefeld, Deckers, Altmann & Roth, 2016). Bei dem Prozess wurden die Items von einem Muttersprachler, sowie professionellen Übersetzern ins Deutsche und dann wieder ins Englische zurückübersetzt und auf Ähnlichkeiten überprüft. Beispiele für Items sind „Wenn jemand anderes aufgeregt ist, neige ich dazu, auch aufgeregt zu sein.“ (Item 1), „Es verärgert mich zu sehen, wenn jemand respektlos behandelt wird.“ (Item 3) und „Ich bleibe unberührt, wenn eine mir nahestehende Person glücklich ist.“ (Item 4). Faktorenanalytisch konnte die Struktur des englischen Originals bestätigt werden (Schönefeld & Roth, 2018). Aufgrund der Ergebnisse bei der Faktorenanalyse gibt es allerdings die Empfehlung Item 2 „Vom Pech anderer Menschen lasse ich mich nicht aus der Fassung bringen.“ aus der Skalenbildung herauszunehmen. Bei der psychometrischen Überprüfung der deutschen Version konnte eine interne Konsistenz von $\alpha=.70$ ($N=833$) ermittelt werden (Schönefeld & Roth, 2018). Kritisch wurde das Instrument betreffend angemerkt, dass Erkenntnisse aus den kognitiven Neurowissenschaften nicht ausreichend Beachtung fanden (Gerdes, Lietz & Segal, 2011).

2.1.3.6 Weitere Selbstauskunfts-Verfahren

Neben den bereits beschriebenen Verfahren gibt es noch weitere, die es zumindest zu erwähnen gilt. Ein weiteres, schon älteres Verfahren, der Empathy Rating Test, kommt von Dymond (1949). Er legt ein eher kognitives Verständnis von Empathie zugrunde. „Empathy will be used in this paper to denote the imaginative transposing of oneself into the thinking, feeling and acting of another and so structuring the world as he does.“ (Dymond, 1949, S. 127). Sein Verfahren besteht aus sechs Items, die in vier verschiedenen Runden auf einer 5-stufigen Skala eingeschätzt werden sollen: „1. self-confidence, 2. superior-inferior, 3. selfish-unselfish, 4. friendly-unfriendly, 5. leader-follower, 6. sense of humour“ (Dymond, 1949, S.

128). Im ersten Schritt soll die Person A sich selbst einschätzen, im zweiten Person B, im dritten wie sich Person B vermutlich selbst einschätzt und im letzten Schritt wie Person B vermutlich Person A einschätzt (Dymond, 1949). Als Testwert wird die Nähe der Einschätzungen und Übereinstimmung genommen. Da die Ergebnisse von Person B auch vorliegen, kann die Differenz der Einschätzungen genommen und als Testwert summiert werden.

Gerade bei und mit Kindern wurde vielfach die *Picture-Story-Method* eingesetzt (Feshbach & Roe, 1968). Bei diesem Verfahren werden Kindern eine Reihe von Bildern begleitet von der Erklärung einer Situation gezeigt, die den Emotionen Freude, Traurigkeit, Angst oder Ärger zugeordnet werden können. Das Alter der Personen in den Bildern liegt bei 7 Jahren. Ein Beispiel für Traurigkeit wäre im englischen Original: "Slide I: Here is a boy and his dog. This boy goes everywhere with his dog, but sometimes the dog tries to run away. Slide II: Here the dog is running away again. Slide III: This time the boy cannot find him, and he may be gone and lost forever." Direkt nach der Präsentation der Bilder wurden die Kinder dann gefragt wie sie sich fühlen. Die Antwort darauf wurde wortwörtlich als Indikator für Empathie festgehalten. In manchen Settings wurden die Kinder auch darüber befragt, wie sich die Person in der Bildergeschichte fühlen würde. Auf diesem Wege war es möglich herauszufinden, wie gut die Kinder die Emotionen in der Geschichte erkennen konnten und ob sie fähig waren die Emotionen adäquat in Worte zu fassen (Feshbach & Roe, 1968). Kritisch gesehen wurde bei diesem Verfahren, ob wirklich Empathie ausgelöst wurde, da es sich in den erfundenen Bildergeschichten um fiktive Personen handelte, die vermutlich andere Wirkungen erzeugten als reale Personen mit echten Emotionen (Eisenberg & Miller, 1987).

Ein weiteres recht bekanntes Verfahren ist die *Basic Empathy Scale* (Jolliffe & Farrington, 2006). Das Selbstauskunftsverfahren erhebt empathische Empfänglichkeit bei Erwachsenen und inkludiert damit sowohl kognitive als auch affektive Facetten. Faktorenanalysen

bestätigen diese zwei-faktorielle Annahme. Das finale Verfahren besteht aus 20-Items, die auf einer Likert-Skala von *Starke Ablehnung* (1) bis *Starke Zustimmung* (5) eingeschätzt werden. Besonderheit des Verfahrens ist, dass neben negativen Emotionen auch positive gemessen werden, was eine bisherige Lücke schließt (Butters, 2010). Bei der Konstruktion des Fragebogens wurde auch darauf geachtet, die bekannten Zusammenhänge zwischen Selbstauskunftsfragebögen und sozialer Erwünschtheit mit zu berücksichtigen (Jolliffe & Farrington, 2006; vgl. Richter, 2009).

2.1.3.7 Situative Tests

Etwas objektivere Messverfahren wurden schon mit dem *Reading-the-mind-in-the-eyes*-Test und dem *Picture-Story-Verfahren* (Feshbach & Roe, 1968) vorgestellt. Gerade bei Messungen einer Fähigkeit können objektivere Verfahren im Gegensatz zu Selbst-Auskunftsverfahren Vorteile mit sich bringen. Bei Selbsteinschätzungen bleibt immer die Frage offen, ob tatsächlich die in Frage kommende Kompetenz oder das Ganze eher „als Maße für ‚wahrgenommene Kompetenz‘ (perceived competence)“ (Hartig & Jude, S. 25) betrachtet werden müssen. Deswegen sollten auch im Bereich der Messung von Empathie objektivere Messweisen in Betracht gezogen werden. Üblicherweise werden für derartige Instrumente Videoaufnahmen von Personen genutzt (vgl. Richter, 2009). Diese können entweder tatsächliche Filmaufnahmen (Davis, Hull, Young & Warren, 1987) sein oder Aufnahmen realer Personen mit authentischen Emotionen (Ickes, Gesn & Graham, 2000; Klein & Hodges, 2001; Levenson & Ruef, 1992). Meist werden Ausschnitte von fünf bis 15 Minuten genutzt, „die Personen in emotionalen, dramatischen und problematischen Lebenssituationen zeigen und bei den Teilnehmern teilweise sehr starke Emotionen auslösen können.“ (Richter, 2009, S. 11). Der Vorteil der Aufnahmen realer Personen besteht in der Authentizität des emotionalen Erlebens und der Möglichkeit, die Personen zu ihren Gedanken und Emotionen

in der jeweiligen Situation zu befragen, um die Antworten der ProbandInnen daran messen zu können. Es wird vermutet, dass reale Personen eine andere Wirkung auslösen als SchauspielerInnen, die die Gefühle lediglich spielen (Richter, 2009). Richter (2009) hat seine eigenen Studien dementsprechend mit Filmaufnahmen von *echten* Personen durchgeführt. Die Personen in den Videoaufnahmen sprachen über ein für ihr Leben bedeutungsvolles Thema und die ProbandInnen sollten eine Einschätzung darüber abgeben, welche Emotionen die erzählenden Personen in dieser Situation durchlebten. Da die gefilmten Personen selbst bereits Informationen darüber abgegeben hatten, konnten die Antworten der ProbandInnen damit abgeglichen werden. Zusätzlich wurden die Aufnahmen noch durch drei Beobachter anhand positiver/negativer Emotionen, sowie Komplexität der verbalen Stellungnahme beurteilt. Die Interrater-Reliabilitäten lagen im Schnitt bei $ICC=.89$.

Als weitere Möglichkeit für situative Messverfahren können Interaktionssituationen genutzt werden (Decety & Ickes, 2011). Diese lassen sich unterteilen in unstrukturierte und strukturierte Interaktionen. Richter (2009) beschreibt es folgendermaßen:

Bei Untersuchungen zu unstrukturierten dyadischen Interaktionen interagieren Personen miteinander, werden währenddessen (teilweise ohne ihr Wissen) auf Video aufgenommen, geben im Nachhinein eigene Gedanken und Gefühle an und schätzen die Gedanken und Gefühle des Gegenübers ein (z.B. Ickes, Bissonette, Garcia & Stinson, 1990). (S.11)

Bei strukturierten Interaktionen, die überwiegend mit Kindern durchgeführt wurden, beschäftigte sich eine Versuchsleitung auf unterschiedliche Weise mit den Kindern, die währenddessen auf Video aufgenommen wurden (Richter, 2009). Die Videoaufnahmen der Kinder wurden dann von geschulten Ratern bzgl. verschiedener Konstrukte eingeschätzt. Solche Settings werden vor allem für Messungen zu empathischer Akkuratheit genutzt (Decety & Ickes, 2011). Richter (2009) weist darauf hin, dass vor allem bei den

Interaktionsmethodiken die unzureichende Standardisierung zu bemängeln ist, so dass eine Vergleichbarkeit schwierig wird. Selbst bei der strukturierteren Version bleibt fraglich, ob die Versuchsleitung tatsächlich bei allen Kindern das gleiche Verhalten zeigt oder durch unterschiedliche Verhaltensweisen die Interaktionen beeinflusst werden.

Ebenfalls mit Videoaufnahmen arbeitet die *Affective Sensitivity Scale* (Kagan & Kratwohl, 1967). Die Videosequenzen zeigen echte Interviewsituationen von TherapeutInnen und KlientInnen, die danach ausgewählt wurden, ob eine emotionale Situation gezeigt wurde und eine Stimmungsveränderung beobachtet werden konnte. Zur Einschätzung sollte aus einer Multiple-Choice-Liste das Gefühl gewählt werden, das das letztgezeigte Gefühl der KlientIn widerspiegelte. In ersten Versionen konnten Reliabilitäten von $\alpha=.70$ gezeigt werden.

2.1.4 Geschlechtereffekte

In vielen Studien wurden immer wieder Geschlechtereffekte gefunden. So wurde beispielsweise herausgefunden, dass sich männliche Probanden, die als weniger empathisch eingestuft werden konnten, aggressiver verhielten als Frauen oder generell empathischere ProbandInnen (Mehrabian & Epstein, 1972). Ebenfalls Unterschiede in den Ergebnissen fanden Jolliffe und Farrington (2006) bei Entwicklungsstudien zu ihrer *Basic Empathy Scale*, so dass Frauen sowohl in kognitiven, als auch affektiven Empathiefacetten höhere Werte erhielten. Die Autoren führen die Unterschiede im Abschneiden nicht auf soziale Erwünschtheit zurück, da sie kaum Korrelationen mit Skalen zur Sozialen Erwünschtheit finden konnten. Baron-Cohen, Wheelwright und Hill et al. (2001) gehen sogar so weit zu sagen, dass im Schnitt bei Männern häufiger autistische Züge beobachtet werden können, so dass das schlechtere Abschneiden in Fragen zur empathischen Akkuratheit auch darauf zurückgeführt werden könnte. Entgegen diesen Befunden konnten Ickes et al. (2000) diesen Effekt bezüglich empathischer Akkuratheit in einer gemischten Studierendenstichprobe nicht

feststellen. Vermutlich lassen sich diese Befunde damit erklären, dass sich diese Effekte in normal entwickelten Stichproben nicht finden lassen (Decety & Ickes, 2011). Geschlechtsspezifische Effekte lassen sich allerdings durch entsprechende Instruktionen herstellen, wenn Frauen bspw. darauf hingewiesen werden, dass sie das empathischere Geschlecht sind (Decety & Ickes, 2011), verschwinden aber wieder sobald die ProbandInnen als Belohnung für richtig gezogene Schlüsse Geld bekommen (Klein & Hodges, 2001). Motivation kann also als ein wichtiger Faktor für empathische Akkuratheit angesehen werden. Auch im IRI werden immer wieder Geschlechtseffekte berichtet (Alexander, 2001; Davis, 1983). Ebenfalls spielt die Art des Messinstrumentes eine Rolle. Eisenberg und Lennon (1983) konnten zeigen, dass diese Effekte auftreten, sobald sehr offensichtlich aus dem Instrument hervorgeht, was gemessen werden soll. Diese vermutlich sozialisierten Effekte zeigen sich schon im Kindesalter. Dabei wurde festgestellt, dass Mütter und Väter in Gesprächen mit ihren 4-jährigen Töchtern häufiger Emotionsworte nutzten als in Gesprächen mit ihren 4-jährigen Söhnen (Aznar & Tenenbaum, 2015).

2.1.5 Physiologische Korrelate

2.1.5.1 Spiegelneuronen

Ende der 1990er Jahre gab es eine vermeintlich bahnbrechende Entdeckung, die gerade im Zusammenhang mit Empathie eine große Rolle zu haben schien und scheint: Die Spiegelneuronen. 1988 veröffentlichten Rizzolatti et al. eine Studie, in der sie die Entdeckung einer neuen Neuronenart im prämotorischen Kortex des Gehirns von Makaken beschrieben. Dabei handelte es sich um eine Zufallsentdeckung, da diese Beobachtung nicht Teil der Studie war, sondern zufällig bei Umbauten während der Pause festgehalten wurde (Hickok, 2015). Das besondere an den entdeckten Neuronen war, dass sie nicht nur feuerten, wenn die Affen selbst eine Aktion ausführten, sondern schon aktiviert wurden, wenn die Affen den Forschern

bei Bewegungsabfolgen lediglich zuschauen. Die ForscherInnen schlossen daraus, dass diese Neuronen Handlungen antizipieren und spiegeln konnten und auf neuronaler Ebene eine Handlung imitierten. Aus diesen Ergebnissen entstanden neue Versuchsreihen, welche die Zufallsbefunde bestätigen konnten und postulierten, dass es auch beim Menschen dieses System geben müsste (Cochin, Barthelemy, Lejeune, Roux & Martineau, 1998; Hickock, 2008; Rizzolatti & Sinigaglia, 2008). Es wurde äquivalent zu den Funden im Affengehirn bei den Menschen im Broca-Areal verortet (Hickock, 2008), obwohl klar war, dass sich aufgrund der methodischen Einschränkungen beim Menschen (keine Einzelneuronenmessungen) die Gesamtstruktur des Spiegelneuronensystems nicht abbilden lässt (Rizzolatti & Sinigaglia, 2008). Entgegen der Vermutung, sei diese Lokalisation jedoch kein Zeichen dafür, dass verbale, internale Repräsentationen des Gesehenen erfolgten, sondern eine neuronale Imitation der gesehenen Bewegungen stattfände (Buccino et al., 2001; Rizzolatti & Sinigaglia, 2008). Zudem wurde bemerkt, dass das System beim Menschen ausgeprägter zu sein scheint und neben dem Nachvollziehen von Handlungen auch Rückschlüsse auf die Intention und darauf folgende Bewegungsmuster gemacht würden (Iacoboni et al., 2005; Rizzolatti & Sinigaglia, 2008). Zusätzlich wurde angenommen, dass Spiegelneuronen auch dazu in der Lage sind, Emotionen des Gegenübers nachzuahmen, so dass sie eine Basis für Gefühlsansteckung sein könnten (Frith & Frith, 2006). Es lässt sich also annehmen, dass auch bei Empathie Spiegelneuronen involviert sind. Ein weiterer Bereich, bei dem eine Spiegelung der Empfindungen des Gegenübers eine Rolle spielt und vermutlich das Spiegelneuronensystem aktiviert wird, ist Schmerz (Lamm et al., 2011). „Beobachten wir, dass andere Menschen physische Schmerzen erleiden oder seelische Qualen, werden auch bei uns die Schmerzzentren aktiviert.“ (Bartens, 2017, S.125). Wie bereits Eisenberger und Lieberman (2004) feststellten, teilen sich physischer und psychischer Schmerz teilweise die gleichen Nervenbahnen und auch die kortikalen Erregungsmuster sind ähnlich.

Nach der Entdeckung des Spiegelneuronensystems wurden diese für die verschiedensten Prozesse verantwortlich gemacht: vom Spracherwerb bis zum Miteifern bei Sportereignissen (Hickok, 2015). Diese damalige Neuentdeckung war allerdings auch nicht ganz unumstritten. Kritik wurde geäußert, dass die Beobachtungen beim Affen eins zu eins auf den Menschen übertragen wurden. Dies erschien schwierig, da im Broca-Areal die Sprachproduktion des Menschen angesiedelt ist und Affen bekanntermaßen nicht sprechen können – das Areal bei ihnen hat also andere Funktionen (Hickock, 2008; Hickok, 2015). Es wurde zeitweise gemutmaßt, dass auch bei den Affen eine rudimentäre Sprachverarbeitungsform vorläge, was allerdings einer empirischen Basis entbehrte. Und tatsächlich lässt sich diese Neuronenart im menschlichen Gehirn nicht genauso gut nachweisen wie beim Affen (Hickock, 2008; Rizzolatti & Sinigaglia, 2008). Das lässt sich hauptsächlich darauf zurückführen, dass bei Makaken aufgrund der geringeren Anzahl von Neuronen das Messen einzelner Nervenzellen möglich ist, beim Menschen allerdings nicht. Hier können nur Neuronen-Cluster über bildgebende Verfahren beobachtet werden, so dass der eindeutige Nachweis beim Menschen schwierig ist. Kurzum: Es scheint unstrittig, dass es in irgendeiner Form ein Spiegelneuronensystem zu geben scheint. Was und wieviel dadurch allerdings erklärt werden kann, ist nach wie vor fraglich (Hickock, 2008). Was das Auftreten von Empathie anbelangt kann sicherlich davon ausgegangen werden, dass Spiegelneuronen beteiligt sind, allerdings wird angemerkt, dass vermutlich noch andere Prozesse involviert seien (Rameson & Lieberman, 2009). Ansonsten ließe sich schwer erklären, warum die Emotion der empathischen Person nur ähnlich und nicht deckungsgleich mit der der beobachteten Person ist.

2.1.5.2 Imitation

Die Spiegelneuronen scheinen damit einen Prozess möglich zu machen, der als solcher als gesichert gilt und der immer wieder in Verbindung zu Empathie beschrieben wird: die

Imitation. „Die Entdeckung der Spiegelneurone scheint eine mögliche Neueinschätzung des Prinzips der ideomotorischen Kompatibilität nahelegen: Das gemeinsame Repräsentationsschema wäre dann nicht als abstraktes, amodales Schema zu betrachten, sondern als Mechanismus der direkten Transformation der visuellen Transformation in potentielle motorische Akte.“ (Rizzolatti & Sinigaglia, 2008, S.146). Die Imitation oder auch der *Chamäleon-Effekt* (Chartrand & Bargh, 1999) sorgt dafür, dass sich Menschen sympathischer sind und es ihnen möglich ist, andere zu verstehen (Decety & Ickes, 2011). In unterschiedlichsten Experimenten konnten diese Effekte gezeigt werden. Man fand beispielsweise heraus, dass sich Sprachmuster von Interviewpartnern, die sich verstehen, anpassen, wenn sie sich sympathisch sind (Cappella & Planalp, 1981; Hatfield, Rapson & Le Yen-Chi, 2009). Das Phänomen der Imitation gibt es nicht nur bei Menschen, sondern bspw. auch bei Affen (Palagi, Norscia & Demuru, 2014). Hier fanden die ForscherInnen heraus, dass sich die Anzahl und die zeitliche Kontingenz des Gähnens (als Operationalisierung von Gefühlsansteckung) erhöhte je enger die soziale Bindung der beobachteten Affen oder Menschen war. Auch wenn die Reaktionen bei Affen und Mensch weitestgehend vergleichbar waren, zeigte sich bei den engeren Verbindungen, dass Menschen schneller und häufiger gähnten. Die AutorInnen sahen das als Indiz, dass empathische Verhaltensweisen beim Menschen doch anders und möglicherweise höher ausgeprägt seien als bei Affen. Auch Emotionsausdrücke werden imitiert und führen dadurch zu einem eigenen Nachempfinden der Emotion (Adolphs, Tranel & Damasio, 2003; Hatfield et al., 2009; Phillips et al., 1997). Hierbei scheint es ein Zusammenspiel von Gestik, Mimik, Körperhaltung und sogar Sprachgestus zu geben. Dass besonders auch verbale Imitation Emotionen induzieren konnte, zeigte eine Studie von Hatfield, Hsee und Costello (1995). Sie ließen ihre ProbandInnen eines von sechs auf Kasette aufgenommenen Geräuschen nachahmen und erfragten im Anschluss die Gefühle ihrer Versuchspersonen. Diese berichteten ähnliche Emotionen wie die, deren

Geräuschmuster sie imitiert hatten. Diese Resultate bekräftigen Funde von Ekman, Levenson und Friesen (1983), die in ihren Studien zeigen konnten, dass Erregungsmuster im autonomen Nervensystem bestimmten Emotionen zugeordnet werden können und sich diese Erregungsmuster bereits zeigen, wenn Gesichtsmuskelkontraktionen dieser Emotionen gezeigt werden (ohne eine vorhergegangene erlebte Emotion). Da sich dieses Phänomen auch bei anderen Wirbeltieren zeigte, kann diese emotionale Ansteckung als Schutzmechanismus betrachtet werden, um die eigene Gruppe schnell auf Gefahren hinweisen zu können (Hatfield et al., 2009). "Imitation is, in essence, the bridge leading to empathy. Indeed, developmental research indicates that we are hardwired for imitation with our conspecifics, and that such a mechanism is the stepping-stone to intersubjectivity (Meltzoff & Decety, 2003)." (van Baaren, Decety, Dijksterhuis, van der Leij & van Leeuwen, 2009, S.33). Dass Imitation eine Brücke zu Empathie darstellt, mag auch sicherlich der Grund sein, dass Personen mit höheren empathischen Fähigkeiten eine höhere Tendenz für Imitation zeigen (Chartrand & Bargh, 1999). Auch in der Beratungspraxis konnte gezeigt werden, dass BeraterInnen, die Arm- und Beinposition ihrer jungen KlientInnen imitierten, als empathischer eingeschätzt wurden (Maurer & Tindall, 1983). Es scheint also ein hoher Zusammenhang zwischen Imitation und der Tragfähigkeit einer Beziehung zu bestehen. Es wird davon ausgegangen, dass Menschen sich nach derartigen Spiegelungen ähnlicher fühlen, was die Bewertung unseres Interaktionspartners wohlwollender ausfallen lässt (van Baaren et al., 2009). Im Gegenzug zeigten Studien, dass es für Personen häufig unangenehm und ungewohnt ist, wenn sie nicht imitiert werden (vgl. van Baaren et al., 2009). Die neurophysiologische Forschung zu Imitationsprozessen gestaltet sich etwas schwierig, da bildgebende Verfahren häufig nicht auf Bewegungen ausgerichtet sind, die bei dieser Forschung aber quasi unerlässlich sind. Behelfsmäßig wird deswegen hauptsächlich mit Aufgabenstellungen zu mimischem Ausdrucksverhalten gearbeitet (van Baaren et al., 2009). In einer Studie von Dimberg,

Thunberg und Elmehed (2000) wurden ProbandInnen für 30 ms Bilder mit freudigen, ärgerlichen oder neutralen Gesichtsausdrücken präsentiert. Auch wenn sie diese nicht bewusst wahrnahmen, konnten in den Ergebnissen typische Muskelkontraktionen, die passend zu der dargebotenen Emotion waren, nachgewiesen werden. Es kann also davon ausgegangen werden, dass Imitation hauptsächlich unbewusst stattfindet, allerdings auch bewusst eingesetzt werden kann, um empathischer (und auch sympathischer) vom Gegenüber wahrgenommen zu werden.

2.1.6 Empathie als trainierbare Kompetenz

Bisher wurde Empathie eher als mittelfristig stabiles Persönlichkeitsmerkmal angesehen (Altmann, 2015), wird allerdings wie bereits beschrieben in manchen Definitionen auch als soziale Fertigkeit betrachtet, so dass zumindest von trainierbaren Anteilen im Sinne einer Fähigkeits- oder Kompetenzentwicklung ausgegangen werden kann (Bowlby, 1969; Zahn-Waxler et al., 2001). Schaut man sich den Begriff *Kompetenz* genauer an, gibt es auch hier keine eindeutige Definition. „Kompetenz bezieht sich sowohl auf Handlungsvollzüge als auch auf die ihnen zugrunde-liegenden mentalen Prozesse und Kapazitäten, zu denen Kognition, Motivation und Volition bzw. Wissen und Können gehören.“ (Klieme & Hartig, 2008, S. 13). In diesem Kompetenzverständnis werden sowohl Dispositions-, als auch veränderbare Anteile angenommen. Zudem sollte immer ein Handlungsfeld bzw. Anforderungsbereich definiert werden, da diese Anteile inhaltspezifisch ausgebildet werden. Da Kompetenz auf die Anwendbarkeit ausgerichtet ist, sollten die Wissensinhalte auch dementsprechend konzipiert und für Anwendung und Transfer vorgesehen sein (Klieme & Hartig, 2008). Kompetentes Handeln bedeutet in diesem Zusammenhang, bestehendes Wissen auf immer neue Situationen transferieren und sie damit erfolgreich bestehen zu können. Durch diese Situationspezifität ergibt sich auch Kompetenz als ein Konstrukt, das sich aus verschiedenen Lernschritten

aufbaut. „Kompetenzen können also durch Erfahrung in relevanten Anforderungssituationen erworben, durch Training oder andere äußere Interventionen beeinflusst und durch langjährige Praxis möglicherweise zur Expertise in der jeweiligen Domäne ausgebaut werden.“ (Klieme & Hartig, 2008, S. 17). So gesehen sind Trainings ein hilfreicher Baustein, um Entwicklungen voranzutreiben und innerhalb der Grenzen der vorhandenen Potentiale Kompetenzen zu verfeinern (Altmann, 2015). Im Falle des Konstrukts Empathie beziehen sich Trainings auf die Ausführung und Anwendung derselbigen. Es wird also die Nutzung der gegebenen Fähigkeiten angestrebt.

Den meisten Empathietrainings liegen Ideen der sozialen Lerntheorie von Bandura und Walters (1963) zugrunde, die besagt, dass ein großer Teil des Lernens über Beobachtung stattfindet. Aufgrund dessen sind bisherige Empathietrainings handlungsbasiert konzipiert und zielen vor allem auf Prozesse der empathischen Akkuratheit (Erkennen der Lebenswirklichkeit des Anderen) oder auf Handlungsoptionen, die im Sinne einer empathischen Verhaltensreaktion auf die Prozesse der empathischen Episode folgen (Butters, 2010). Es gibt eine Fülle an verschiedenen Trainings zur Entwicklung von Empathie, die alle sehr unterschiedlich sind. Daher gestaltet sich auch die Vergleichbarkeit als schwierig. Zusammenfassend lässt sich lediglich ableiten, dass Trainings „... allgemein eine Wirksamkeit in der Veränderung der Empathiewerte erzielen“ (Altmann, 2015, S.57). In Metaanalysen (Baker, Daniels & Greeley, 1990; Butters, 2010; Stepien & Baernstein, 2006) wurde der Versuch unternommen, dieses heterogene Feld etwas strukturierter zu beleuchten und mehr Klarheit in Aufbau und Effekte der Trainings zu bringen. Es flossen Trainingsstudien aus dem angloamerikanischen Raum ein, die verschiedene Bereiche trainierten und für verschiedenste Zielgruppen angeboten wurden. Neben Täter-Opfer-Trainings aus dem forensischen Bereich, gab es Trainings für helfende Berufsgruppen, für Mütter, für Liebespaare, für Medizinstudierende, Präventionsprogramme für Eltern, usw. Die

Stichprobengrößen der Einzelstudien waren erwartungsgemäß recht klein und die jeweiligen Studienevaluationen von unterschiedlicher Qualität (unterschiedliche Messinstrumente, wenig bis keine Kontrollgruppen-Designs, uneinheitliche Konstruktdefinitionen). In einem systematischen Review und Meta-Analyse über Empathie-Trainings für Medizin-Studierende (Stepien & Baernstein, 2006) wurde herausgefunden, dass der größte Effekt bei Kommunikations-Trainings gefunden werden konnte, die Verhaltenskomponenten von Empathie ansprachen. In einer weiteren Metaanalyse (Hill & Lent, 2006) wurden Trainingsprogramme für helfende Berufsgruppen untersucht, die vor allem auf die Verhaltenskomponente ausgerichtet waren. Es konnte festgestellt werden, dass jede untersuchte Trainingsform effektiver war als gar kein Training und multimodale Angebote besser abschnitten als unimodale Trainings.

Ein weiteres Training zur Stärkung der Selbst-Andere-Differenzierung wurde von Altmann (2015) entwickelt. Aufbauend auf der Methode der gewaltfreien Kommunikation (Rosenberg, 2016) sollte diese als Reflexionswerkzeug erlernt werden, um reine Gefühlsansteckung zu vermeiden und zwischen Eigen- und Fremdgefühl zu unterscheiden. Dieser angenommene Effekt konnte bestätigt werden. Bei der Arbeit mit Personengruppen, die typischerweise Beeinträchtigungen bei Empathie aufweisen (Narzissmus und Psychopathie), konnten Forscher zeigen, dass auch diese Gruppen empathische Akkuratheit entwickeln konnten (Hepper, Hart & Sedikides, 2014; Keysers & Gazzola, 2014). Das Problem liegt gar nicht so sehr am Können, sondern in der Motivation. In verschiedenen Studien (Hepper et al., 2014) wurden Situationen präsentiert, in denen Personen beschrieben wurden oder zu hören waren, die sich in einer misslichen Lage befanden. Die ProbandInnen reagierten erst empathisch und einfühlsam darauf, nachdem sie explizit dazu aufgefordert wurden, sich in die andere Person hineinzuversetzen. Erst dann konnten mit bildgebenden Verfahren für Empathie typische Erregungsmuster entdeckt werden, von allein konnten nur sehr schwach aktivierte Bereiche

gefunden werden. Die AutorInnen schlossen daraus, dass narzisstische Persönlichkeiten durchaus zu mitfühlendem Verhalten fähig sind, allerdings ausdrücklich zur empathischen Perspektivenübernahme aufgefordert werden müssen, da ihnen von alleine die Motivation dazu fehlt. Ähnlich verhält es sich mit psychopathischen Persönlichkeiten (Keysers & Gazzola, 2014). Auch hier gab es kaum empathische Erregungsmuster, wenn sie sich ohne Instruktion Videos ansahen, in denen Personen Schmerzen hatten. Erst wenn sie aufgefordert wurden, sich in die andere Person einzufühlen, wurden auch hier adäquate Erregungsmuster gefunden. Aus diesen Befunden lassen sich Trainingsmaßnahmen konzipieren, wobei fraglich bleibt, ob die Motivation, das Verhalten zu zeigen, auch in realistischen Settings gegeben wäre.

In einer weiteren Studie wurde untersucht, ob bei einem bestimmten Täter-Opfer Trainingsprogramm positive Veränderungen in Form von Schuld oder Empathie bei den Tätern auftraten oder eine geringere Rückfälligkeit auftrat (Jackson & Bonacker, 2006). Generell konnte das nicht nachgewiesen werden, aber Täter, die ein höheres Schuldempfinden berichteten, gaben eine höhere Empathieentwicklung an (das galt vor allem für Frauen). Außerdem gab es weniger Rückfälle als in der Kontrollgruppe in den ersten 6 Monaten nach dem Training, allerdings schien dieser Effekt nichts mit Empathie zu tun zu haben, sondern mit der Höhe des Einkommens. In einer weiteren Studie konnte gezeigt werden, dass sich bei einem Kurzzeit-Training, bei dem Zuhörkompetenzen geschult wurden, die Empathie-Werte erhöhten (Zucker, Worthington & Forsyth, 1985). Butters (2010) konnte feststellen, dass auch die Messmethoden einen Einfluss auf die Effektstärke zeigten. Aus Selbsteinschätzungen resultierten kleinere Effektstärken als aus Fremdeinschätzungen. Als Grund dafür nannte er den Faktor Soziale Erwünschtheit, so dass sich die ProbandInnen bei Fremdeinschätzungen unter Umständen besonders bemühten und es somit zu einer Überschätzung der wahren Ausprägung kam. Diese Wirkung der sozialen Erwünschtheit gilt allerdings auch für

Selbstauskunftsfragebögen, die jedoch einen geschützteren Rahmen bieten könnten, um erlebte Änderungen in Empathie anzugeben. Zudem gab es einen Einfluss von hohen Pre-Test-Scores, dem Alter (jüngere ProbandInnen profitierten mehr) und der wahrgenommenen Empathie der TrainerInnen. Es konnte über 24 Trainings hinweg eine Effektstärke von $d=1.05$ (bzw. bereinigt $d=0.93$) ermittelt werden (Butters, 2010). Hierbei hatten verschiedenste Trainingsmethoden einen Effekt (wobei dieser bei psychoedukativen Maßnahmen kleiner ausfiel) und die höchsten Empathieentwicklungen waren im späten Erwachsenenalter zu beobachten. Eine Langzeitwirkung konnte aufgrund der Befunde nicht festgestellt werden, da die Evaluationsstrategien in den meisten Fällen keine Follow-Up-Untersuchung vorsahen. Die Zielgruppe und auch die *Freiwilligkeit vs. Verpflichtung* hatten keinen Effekt auf die Ergebnisse. In zukünftiger Forschung sollte die Verbindung zwischen Empathie, sozialer Verstärkung und Motivation untersucht werden, um das Anwenden des Erlernten besser zu fördern. Zusammenfassend lässt sich sagen, dass Empathie in verschiedenen Aspekten trainierbar zu sein scheint, Trainings allerdings sehr spezifisch auf bestimmte Bereiche zugeschnitten sein müssen und eine methodisch starke Evaluation haben sollten.

2.2 Service-Learning als Lehrmethode

Service-Learning (im weiteren *SL* abgekürzt) hat entgegen dem ersten Eindruck nichts mit einem Einsatz im Dienstleistungssektor oder Einzelhandel zu tun. Es bezeichnet stattdessen ein Lehr-Lernformat, das es in Schulen und Hochschulen gibt und bei dem akademische Inhalte mit einem ehrenamtlichen Engagement außerhalb der Universität verknüpft werden. Hierbei werden in der Regel die zu besuchenden Universitätsveranstaltungen mit studienrelevanten Nachweisen honoriert (wie bspw. CPs oder Praktikumsnachweisen, o.ä.). „Unter Service Learning wird dabei eine Lehr-Lernform verstanden, bei der SchülerInnen oder Studierende Wissen und Kompetenzen in Schule und Hochschule erwerben und dieses Wissen in der Praxis zum Nutzen sozialer Belange einsetzen.“ (Reinders, 2010, S.532). Das kann beispielsweise eine Law-Clinic sein, bei der Studierende der Rechtswissenschaften unter Supervision kostenlose Rechtsberatung für geflüchtete Personen anbieten. Sie werden dabei auf die Tätigkeit in Rahmen von Lehrveranstaltungen vorbereitet, für die sie CPs erhalten und dauerhaft von Dozierenden betreut, die mit Reflexionssitzungen, sowie informellen Sprechstunden für Beratung zur Verfügung stehen. Oder Programme wie z.B. Balu und Du (Balu & Du, 2018), in dessen Rahmen Grundschulkindern (Moglis) aus sozial schwachen Familien ein Jahr lang von einem Studierenden (Balu) betreut werden, gemeinsam die Freizeit verbringen und die Balus ein offenes Ohr für die Probleme und Anliegen ihres Schützlings haben. Auch die Studierenden in diesem Projekt treffen sich regelmäßig zur Reflexion und zum Austausch und können jederzeit die Unterstützung des Dozierenden in Anspruch nehmen. Dieses Projekt läuft seit 2001 deutschlandweit an verschiedenen Hochschulstandorten. Wie bereits Reinders (2016) anmerkt, sieht sich das Lehr-Lernformat SL vielen Erwartungen gegenüber. Neben einer besseren Theorie-Praxis-Verzahnung und Erhöhung der fachlichen, sozialen, sowie politischen Kompetenz, sollen die Studierenden auch noch *etwas Gutes tun*, die Gesellschaft unterstützen und sich nebenbei noch selbst in ihrer Persönlichkeit

weiterentwickeln. „Einer dieser Faktoren ist die Förderung der Entwicklung individueller Lernstrategien und selbstgesteuerter Lernprozesse, denn: Nur Lernende, die über ihre eigenen Stärken und Schwächen ihres Lernens Bescheid wissen, sind offenbar in der Lage, ihre Lernstrategien in unterschiedlichen Kontexten auch effektiv anzuwenden.“ (Häcker et al., 2008, S.1). Hierbei muss SL den Spagat schaffen, persönliche Entwicklungsstränge mit den akademischen und kognitiven Weiterentwicklungen zu verknüpfen (Eyler & Giles, 1999). Ein entscheidendes Element von SL-Veranstaltungen stellt die Reflexion dar (Eyler & Giles, 1999; Reinders, 2016). Studierende geben in Befragungen häufig an, dass sie die Reflexion am nützlichsten fanden, um affektive und kognitive Elemente verbinden zu können, mit denen sie in der Praxis konfrontiert waren (Eyler & Giles, 1999). Aber was genau macht diese Formate besonders? Wie sollten sie optimalerweise gestaltet werden? Und welchen Nutzen haben Studierende, wenn sie teilnehmen? Mit diesen und weiteren Fragen beschäftigen sich die nächsten Kapitel.

2.2.1 Begriffsdefinition

Auch beim Begriff SL gibt es keine eindeutige Definition, sondern viele verschiedene Sichtweisen, die ein sehr heterogenes Verständnis aufzeigen. Von Veranstaltungen, bei denen einmal an einem Nachmittag ein ehrenamtliches Engagement stattfindet bis zu Veranstaltungen, die gut ins Curriculum von Hochschule und Schule implementiert sind und sich über ein bis zwei Jahre erstrecken, werden verschiedenste Formate als SL deklariert (Eyler & Giles, 1999). Ebenfalls ist die Schwerpunktlegung der Kurse auf den Service oder den Learning Teil ein Thema, das schwierig zu beantworten ist. Bestenfalls ist natürlich eine Gleichbehandlung beider Seminarteile anzustreben, aber auch hier zeigen sich immer wieder Unterschiede (Eyler & Giles, 1999; Sigmon, 1997). Je nachdem wie das Lernziel definiert ist, ergeben sich unterschiedliche Schwerpunkte in der Umsetzung. Um hier mehr Klarheit zu

erreichen, schlägt Sigmon (1996) eine Einordnung anhand der Zielsetzungen vor (siehe Tabelle 8).

Tabelle 8
Service-Learning Typisierung; nach Sigmon, 1996, 1997

Typ	Schwerpunktsetzung	Beispiel
SERVICE-learning	Service im Mittelpunkt (nur unstrukturierte, gelegentliche Reflexionsmöglichkeiten)	Ehrenamtliches Engagement kann für das Studium angerechnet werden und es finden unregelmäßige Treffen statt
service-LEARNING	Lernen im Mittelpunkt (Service ist zweitrangig)	Settings in der klinischen Ausbildung, bei der der Lerninhalt im Mittelpunkt steht und der Service additiv ist
service learning	Beide Teile sind nicht verbunden	Ehrenamtliches Engagement innerhalb der Universität, bei dem es keine akademischen Ziele gibt
SERVICE-LEARNING	Gleichberechtigung jedes Teils und aller involvierten Stakeholder	Ehrenamtliches Engagement und akademische Ziele sind verflochten und jeder lernt von jedem

Diese Heterogenität ist erschwerend, aber auch lohnend, da sich so explorieren lässt, welche Bestandteile zu welchen Ergebnissen führen können (Eyler & Giles, 1999).

SL ist auch mit Begriffen wie erfahrungsbasiertem, projektorientiertem, problembasiertem oder erforschendem Lernen assoziiert und baut auf dessen Grundprinzipien auf (Dewey, 1916; Renkl, 2009). Hierbei geht es vor allem darum, neues Wissen zu generieren und in bisher bekannte Wissenskonzepte und Modelle zu integrieren (Renkl, 2009). Wie Mayer (2004) anmerkt geht es bei derartigen Lehrveranstaltungsformaten nicht primär darum, Aktivitäten zu fördern, sondern sie sinnvoll in Kontexte einzubetten, damit dadurch neues Wissen entstehen kann. Dabei spielen die Anleitung und Instruktion eine entscheidende Rolle, da vor allem Lernende mit geringen Vorkenntnissen Prompts und auch inhaltlichen Input benötigen, um zielorientiert Wissen aufbauen zu können (Hmelo-Silver, Duncan & Chinn, 2007; Wild & Möller, 2009b). Theoretischer Hintergrund für diesen Lehr-Lernansatz ist hierbei der

Konstruktivismus (Hmelo-Silver et al., 2007). „These processes are internal to the individual—integrating knowledge (or meaning) into pre-existing schemes (assimilation) or changing the schemes to fit the environment (accommodation) (Piaget & Inhelder, 1969).“ (Young & Collin, 2004, S. 375). Damit ist gemeint, die eigene Wirklichkeit mit neuen Inhalten anzureichern, Fehlinterpretationen sowie Missverständnisse aufzudecken und damit das eigene Wissen zu erweitern bzw. anzupassen. Es wird hier auch von *situierter Kognition* gesprochen, da das Wissen in geschaffenen Situationen zum Einsatz kommt und weg von sogenanntem *trägen Wissen* führen soll (Renkl, 2009). Träges Wissen bezeichnet ein in einer Person innewohnendes Wissen, das allerdings nur in der erlernten Situation gezeigt werden kann und bei dem ein Transfer zur Anwendung in anderen Situationen nicht oder kaum vorkommt (Renkl, 2009; vgl. The Cognition and Technology Group at Vanderbilt, 1990).

SL lässt sich aufgrund seiner Spezifika noch besser vor dem Hintergrund des *sozialen Konstruktivismus* verstehen (Reinders, 2010). Hierbei wird davon ausgegangen, dass die neu erschlossenen Wissensinhalte durch soziale Beziehungen generiert werden (Young & Collin, 2004) – ein Vorgehen, das bei SL-Veranstaltungen häufig zu beobachten ist. Reinders (2016) ordnet SL in den sogenannten *Neuen Konstruktivismus* ein. Er nennt hierbei drei Lehransätze, die damit in Verbindung zu stehen scheinen: *Anchored Instruction-Ansatz*, *Cognitive Flexibility-Ansatz* und den *Cognitive Apprenticeship-Ansatz*.

Der *Anchored Instruction Ansatz* stammt von einer Gruppe der Vanderbilt-Universität in Nashville (USA) und beinhaltet eine Lernsituation, in der narrative Problemsituationen zur Bearbeitung gegeben werden (Lipowsky, 2009; Reinders, 2016; The Cognition and Technology Group at Vanderbilt, 1990). Ziel des Ansatzes sollte es sein, träges Wissen zu umgehen und Kenntnisse auf ähnliche Problembereiche anwenden zu können. Es wird eine Geschichte präsentiert (zumeist als Video), in dem Lernende viele Informationen präsentiert bekommen, die nötig sind, um am Ende ein Problem zu definieren und dieses zu lösen (The

Cognition and Technology Group at Vanderbilt, 1990). Dieser Ansatz hat mit Service-Learning gemein, dass ein spezifisches Problem bearbeitet wird, das bei Service-Learning allerdings real vorliegt und nicht nur als Modell (Reinders, 2016).

Beim *Cognitive Flexibility Ansatz* wird ein Themenbereich in seiner Gänze präsentiert und aus verschiedenen Situationen betrachtet. Dadurch sollen Gedanken und Lösungen in Gang gebracht werden, die sich auf breitere Bereiche ausweiten lassen (Reinders, 2016). Hier besteht die Gemeinsamkeit im Nutzen der Vielschichtigkeit realer Situationen, die beim Service-Learning auch gegeben ist.

Der letzte Ansatz ist der *Cognitive Apprenticeship Ansatz*, der ursprünglich aus dem Handwerksbereich stammt und auf das Aneignen kognitiver Kenntnisse übertragen wurde (Brown, Collins & Duguid, 1989; Lipowsky, 2009; Reinders, 2016). Dieser Ansatz ist sehr praxisnah und wird weitestgehend in realen Situationen vermittelt. Als Beispiel nannten Brown et al. (1989) das Erlernen neuer Worte. Sie hielten das Lernen aus Lexika mit einer einfachen Beschreibung des Wortes nicht für ausreichend, sondern betrachteten als eine geeignetere Alternative, die Worte in einen Kontext einzubetten, z.B. das Wort in einem Satz anzuwenden und dadurch die Bedeutung zu verdeutlichen. Sie begründeten das damit, dass das exakt den Prozess beschreiben würde, der bis zu diesem Zeitpunkt genutzt wurde, um Sprache zu lernen und diese Situation für den Lerngegenstand also sehr viel geeigneter wäre. Bei dieser Instruktionmethode besteht der Unterschied zu SL darin, dass bei diesem ein realer Bedarf in der Zivilgesellschaft adressiert wird und echte Interaktionen zwischen Zivilpartnern und Studierenden sowie Theorie und Praxis stattfinden (Reinders, 2016).

Als Unterscheidung zu anderen Projektseminaren oder auch den meisten Praktika lässt sich vor allem die Zielgruppe benennen: Es wird ein realer Bedarf in der Gesellschaft adressiert, der ohne das Seminar nicht gedeckt werden könnte (Reinders, 2010). Pritchard (2002) nennt als weitere Elemente zur Definition des Terms die abzugrenzenden Begriffe *Community*

Service und SL. Das Verständnis von Community Service sei eher altruistischer Natur und nicht zwangsläufig mit Lernprozessen verbunden. Ebenso bezeichneten Praktiker die Tätigkeiten eher als Community Service, während in der Literatur der Begriff SL häufiger vorkäme (Pritchard, 2002), so dass auch auf diesem Wege eine Abgrenzung möglich sei.

2.2.2 Qualitätskriterien

Da es so unterschiedliche Formen gibt, SL durchzuführen, gab es Bestrebungen, Rahmenbedingungen zu schaffen, um das Format etwas besser eingrenzen zu können. Im Jahr 1989 wurde der Prozess initiiert, konsensfähige Qualitätsstandards für SL vor allem für das amerikanische Schulsystem K-12 (Kindergarten bis High-School) zu entwickeln (Billig & Weah, 2008). Bei verschiedenen Konferenzen und landesweiten Panels von 1989 bis 2007 wurden verschiedene Bereiche besprochen und letztendlich acht als entscheidend angesehene Kriterien festgehalten: Die *K-12 Service-Learning Qualitätsstandards*. Die entscheidenden Panels in 2007 boten die Möglichkeit für Studierende, Theoretiker als auch Praktiker sich einzubringen und Vorschläge zu besprechen und fanden zwischen Juli 2006 bis Januar 2007 an 21 Standorten in den USA mit jeweils bis zu 250 Personen statt. Die Kriterien sind als Leitlinien zu verstehen. Zu jedem der Kriterien wurden außerdem Indikatoren festgehalten, die wichtige Schlüsselmerkmale des Standards kennzeichnen (siehe Anhang A1). Die Kriterien sind folgende:

1. Duration and Intensity (Dauer und Intensität)
2. Link to curriculum (Verbindung zum Curriculum)
3. Partnerships (Partnerschaften)
4. Meaningful Service (Bedeutungsvoller Service)
5. Youth Voice (Mitspracherecht)
6. Diversity (Diversität)

7. Reflection (Reflexion)
8. Progress Monitoring (Beobachten des Fortschritts)

Es wurde vorgeschlagen, dass sich die Dauer und Intensität daran bemessen sollte, dass sie ausreichend sind, um die Bedarfe der Zielgruppe in der Gesellschaft zu erkennen und identifizierte Ziele erreichen zu können. Andernfalls wäre fraglich, ob die Veranstaltung für alle Beteiligten zufriedenstellend sein könnte. Ebenfalls sollte eine Verknüpfung mit dem Curriculum der Studierenden oder SchülerInnen gegeben sein. Im Schul- bzw. Hochschulkontext wird SL als Lehr-Lern Setting genutzt, so dass akademische Lernziele zufriedenstellend bedient werden sollten, auch um die Vergabe von studienrelevanten Nachweisen zu rechtfertigen. Wie bereits in der Typologie von SL beschrieben (siehe Tabelle 8) sollte die Gewichtung optimalerweise gleichverteilt auf beide Bausteine sein. Dementsprechend kommt natürlich den Partnerschaften eine zentrale Rolle zu. Um qualitativ gute Veranstaltungen durchzuführen, wird empfohlen, die Partner als gleichwertig zu sehen und für alle Seiten gewinnbringend die gesellschaftlichen Bedarfe zu adressieren, so dass alle Seiten bestmöglich profitieren. Ein weiterer wichtiger Aspekt ist, dass der Service für die Teilnehmenden bedeutungsvoll und persönlich relevant sein sollte. Hierbei ist wichtig, dass der Service-Teil dem Alter und Entwicklungsstand angemessen ist und die Teilnehmenden motiviert sind ihn umzusetzen. Hilfreich dazu ist das Mitspracherecht der Teilnehmenden, so dass sie unter Supervision aktiv in vorbereitende, umsetzende und evaluierende Prozesse der Veranstaltung eingebunden sind. Zusätzlich sollte ein wertschätzendes Kursklima gepflegt werden, in dem Diversität in vielfältiger Form akzeptiert und respektiert wird (sowohl gegenüber der Teilnehmenden, als auch gegenüber allen anderen Partnern und Klienten in dem Projekt). Vor allem dieser Teil kann für die persönliche Entwicklung ein relevanter Faktor sein und Konfliktlöse-Kompetenzen sowie das Ablegen von Stereotypen begünstigen. Um solche Änderungen möglich zu machen, sollten vielfältige Reflexionsmöglichkeiten gegeben

werden, um Metakognitionen zu stimulieren und eine Brücke zwischen gemachten Erfahrungen und theoretischer Fundierung zu schaffen. Hier können Prozesse angeregt werden, die eigenen Einstellungen und Verantwortlichkeiten als Teil der Gesellschaft zu überdenken. Zu guter Letzt wird als ein wichtiger Bestandteil auch noch die Beobachtung und das Festhalten des Fortschritts genannt, um die festgelegten Ziele zu erreichen und den Prozess steuern zu können. Zudem können die Ergebnisse dazu dienen, die Veranstaltung weiterzuentwickeln und für Nachhaltigkeit zu sorgen. Diese Kriterien sind im ursprünglichen Kontext der K-12 Bildung entstanden, wurden allerdings auch auf den Hochschulkontext bezogen und dort angewandt. Die Nützlichkeit der Kriterien konnte bislang auch schon in einer Metaanalyse gezeigt werden (Celio, Durlak & Dymnicki, 2011). Celio et al. (2011) konnten anhand 62 Studien zeigen, dass sich die Einflüsse auf Einstellungen sich selbst gegenüber, gegenüber Schule und Lernen, zivilgesellschaftlichem Engagement, sozialen Kompetenzen und akademischen Leistungen signifikant von Nicht-SL Veranstaltungen unterschieden, sobald einer der Standards *linking to curriculum, voice, community involvement* (im Sinne einer intensiven Zusammenarbeit zwischen Universität und zivilen Partnern), sowie *reflection* angewandt wurde. Die Effekte verdoppelten sich sogar, wenn sie mit SL-Veranstaltungen verglichen wurden, die keines dieser Kriterien erfüllten (von Effektstärke .17 zu .35). Für den deutschen Raum wurden von Seifert, Zentner, Nagy, Baltes und Sliwka (2012) bereits Kriterien für den Schulkontext formuliert, an denen sich interessierte LehrerInnen orientieren konnten. Die Empfehlungen entstanden für das Netzwerk Lernen durch Engagement (Stiftung Lernen durch Engagement, 2018) und es wurden sechs Qualitätskriterien benannt, die an die K-12 Kriterien angelehnt wurden: 1. Realer Bedarf, 2. Curriculare Anbindung, 3. Reflexion, 4. Schülerpartizipation, 5. Engagement außerhalb der Schule sowie 6. Anerkennung und Abschluss. Anfang 2018 wurde nun auch vom Hochschulnetzwerk Bildung durch Verantwortung der Prozess gestartet, allgemeine

Qualitätskriterien festzuhalten, die für den deutschsprachigen Hochschulbereich von Relevanz sind (Hochschulnetzwerk BDV-Bildung durch Verantwortung, 2018a). Hier sind zehn Kriterien identifiziert worden, zu denen Empfehlungen und Anmerkungen der Hochschulnetzwerks-Mitglieder eingeholt wurden, um darauf aufbauend Tools und Instrumente zum Umsetzen und zur Qualitätssicherung vorzuschlagen oder zu entwickeln. Wie sich erkennen lässt, finden sich die K-12 Kriterien weitestgehend in den deutschen Vorschlägen wieder (siehe Tabelle 9).

Tabelle 9

Entwurf von Qualitätsstandards für SL im deutschen Hochschulbereich den K-12 Kriterien gegenübergestellt; nach Hochschulnetzwerk BDV-Bildung durch Verantwortung, 2018a

Qualitätsstandards Hochschulnetzwerk BdV	K-12 Kriterium
1. Gesellschaftlicher Bedarf	Meaningful Service
2. Definierte Ziele	Duration and Intensity
3. SL ist Bestandteil des Studiums	Link to curriculum
4. Kompetenzerwerb der Studierenden	Duration and Intensity
5. Lernen in fremden Lebenswelten	Diversity
6. Partnerschaftliche Zusammenarbeit	Partnerships/Youth Voice
7. Reflexion	Reflection
8. Begleitung von Studierenden	Reflection/Progress Monitoring
9. Qualitätssicherung	Progress Monitoring
10. Anerkennung und Abschluss	

Die deutschen Vorschläge sind umfangreicher als die originalen K-12-Kriterien und mehrere finden sich im Original in einem einzigen Kriterium widerspiegelt. Der Prozess ist allerdings noch nicht abgeschlossen und es ist nicht auszuschließen, dass sich die deutschen Vorschläge noch einmal verändern. Auch hier sind die Standards als Vorschläge zu betrachten und nicht bindend.

2.2.3 Forschungsstand im Bereich Service Learning

2.2.3.1 Hauptbefunde und Darstellung der Forschung im deutschen Raum

Die Forschung zum Bereich SL kommt vor allem aus dem angloamerikanischen Raum. Die Güte der Studien ist dabei sehr unterschiedlich, was vor allem damit zusammenhängt, dass die Studien aus den unterschiedlichsten Fächern und Fachkulturen kommen und häufig Lehrveranstaltungsevaluationen sind (vgl. Reinders, 2016). In vielen Fällen fehlt es an prä-post Designs, sowie Kontrollgruppen und auch die Konstrukte sind sehr oft unterschiedlich definiert, so dass sich eine Vergleichbarkeit schwierig gestaltet (Reinders, 2010). Es gibt auch viele Publikationen (vor allem auch im deutschsprachigen Bereich), die eher Erfahrungsberichte im Sinne einer *good practice* darstellen, in denen z.B. Teilnehmende berichten wie sehr sie die besuchte SL-Veranstaltung mögen (Bernhardt, 2008; Eyler & Giles, 1999), auf welche Bestandteile man bei der Konzeption derartiger Veranstaltungen achten sollte (Altenschmidt, 2009; Seifert et al., 2012) oder sich mit der Wirkung auf die gewählte, gesellschaftliche Zielgruppe beschäftigen (Papenfuß, Redlich, Steinhauer & Friedländer, 2015). Ebenfalls gibt es Studien über die Verbreitung von SL-Veranstaltungen in Deutschland (Backhaus-Maul & Roth, 2013) und auch Arbeiten von deutschen ForscherInnen, die allerdings ExpertInnen-Interviews zum Inhalt haben, die in den USA geführt wurden, um deren Ergebnisse für den deutschen Bildungskontext zu adaptieren (Seifert, 2011). Nichtsdestotrotz lassen sich gerade aus Metaanalysen zum Thema auch vielversprechende Ergebnisse entnehmen. In einer Metaanalyse von Novak, Markey und Allen (2007) wurden neun Studien hinsichtlich *kognitiver Ergebnisse* untersucht. Im Vergleich zu Non-SL-Kursen konnten sie signifikant bessere Ergebnisse im Themenverständnis, gelernten Kompetenzen, in der Anwendung des Gelernten, sowie des Umstrukturierens von komplexen sozialen Themen feststellen. Sie fanden mit einer durchschnittlichen Effektstärke von $d=.42$ in diesem Bereich einen kleinen Effekt. In einer weiteren Metaanalyse von Conway, Amel und Gerwien (2009)

wurden 103 Studien inkludiert und neben *akademischen Leistungen* noch *persönliche, soziale* und *gesellschaftliche Wirkungen* überprüft. Unter *akademische Leistungen* fielen bei ihnen kognitive und akademische Leistungen, die Fähigkeit Wissen anzuwenden und die Motivation zu lernen. *Persönliche Wirkungen* bezogen sich hier auf Einstellungen sich selbst gegenüber, sowie verschiedene Werte wie Selbstwert, Selbstwirksamkeit und die Planung des beruflichen Weges. Die Wirkungen im *sozialen Bereich* bezogen sich auf Einstellungen und Fähigkeiten im Umgang mit anderen. Die *gesellschaftlichen Wirkungen* als letztes Cluster zielten auf das Gefühl gesellschaftlicher Verantwortung, Teilhabe und das Eintreten für gesellschaftliche Gerechtigkeit. Die größten Effekte konnten sie bei den akademischen Leistungen ($d=.43$) und den sozialen Wirkungen hinsichtlich des Abbaus von Stereotypen ($d=.21$) finden. Auch in den Bereichen persönlicher Entwicklung und gesellschaftlicher Wirkungen wurden Effekte gefunden, die allerdings mit $d=.21$ bzw. $d=.17$ recht klein ausfielen. In dieser Studie deutete sich ebenfalls an, dass Reflexionselemente einen positiven Einfluss auf die Wirkungen hatten. Celio et al. (2011) prüften in ihrer Metaanalyse den Effekt von SL auf *Einstellungen sich selbst* sowie *Schule und Lernen gegenüber, zivilgesellschaftliches Engagement, sozialen Kompetenzen* und *akademischen Leistungen*. Das Cluster *Einstellungen sich selbst gegenüber* inkludierte Konstrukte wie Selbstwert, Selbstwirksamkeit, soziale Fähigkeiten und Gefühle von Kontrolle. Einstellungen gegenüber Schule und Lernen erhob akademische Beschäftigungen und Freude der Veranstaltung. Das Cluster *zivilgesellschaftliches Engagement* beinhaltete Ergebnisse, die mit der Zivilgesellschaft verknüpft waren wie Altruismus, gesellschaftliche Verantwortung oder politische Partizipation. Als *soziale Kompetenzen* wurden Führungsfähigkeiten, interkulturelle Kompetenz oder soziale Problemlösefähigkeiten verstanden und als *akademische Leistung* wurden Klausurergebnisse o.ä. einbezogen. In den untersuchten 62 Studien zeigte sich in allen fünf Clustern ein signifikanter Unterschied bei Teilnehmenden in SL-Veranstaltungen im Vergleich zu

Programmen ohne SL (mit dem höchsten Effekt bei den akademischen Leistungen mit einer Effektstärke von .43). Sie sehen vor allem in diesem Punkt ein bestätigendes Ergebnis für SL, da die Schulen aufgrund bildungspolitischer Bestimmungen (No Child Left Behind-Act; U.S. Department of Education, 2001) Möglichkeiten suchten, um alle SchülerInnen gut fördern zu können. Eine weitere Übersichtsarbeit wurde von Warren (2012) durchgeführt. Sie untersuchte die Wirkung von SL-Veranstaltungen auf das *studentische Lernen*. Es wurden 11 Studien einbezogen und auch hier konnte ein Effekt von $d=.33$ auf studentisches Lernen gefunden werden (verglichen mit Nicht-SL-Veranstaltungen). Die Studien orientierten sich an der Metaanalyse von Novak et al. (2007), inkludierten diese und erweiterten sie um neuere Studien. Im gleichen Jahr wurde eine weitere Metaanalyse von Yorio und Ye (2012) publiziert. Hier wurden 40 Studien in Hinblick auf *soziales Verstehen*, *persönliche Einsichten* und *kognitive Entwicklung* untersucht. Im Cluster soziales Verstehen finden sich Variablen wie Toleranz für Diversität, Entscheidungsfähigkeit bzgl. komplexer sozialer Themen, interkulturelle Kompetenz, die Motivation sozial etwas zu verändern sowie moralische und ethische Haltungen. Bei persönlichen Einsichten geht es auch wieder um die Selbstwahrnehmung hinsichtlich der eigenen Identität, Selbstwert, Selbstwirksamkeit, beruflicher Ziele, Selbsteinschätzung eigener Stärken und Schwächen sowie Durchhaltevermögen. Im Bereich der kognitiven Entwicklung flossen Fähigkeiten zur Selbstorganisation, Schreibfähigkeit, Problemlösefähigkeiten, kritisches Denken sowie akademische Leistungsindikatoren in Form von Noten und Punkten ein. Auch hier wurden positive Veränderungen in allen Bereichen gefunden, die allerdings kleine bis mittlere Effektstärken zeigen (soziales Verstehen gemittelt Hodges $g=.34$, persönliche Einsicht gemittelt Hodges $g=.28$ und kognitive Entwicklung gemittelt Hodges $g=.52$). Der Großteil der zu Grunde liegenden Studien kamen aus dem Hochschulkontext und die meisten Analysen nutzten auch unpublizierte Forschung. Trotzdem ist die Vergleichbarkeit der Studien aufgrund

der unterschiedlich definierten und erhobenen Konstrukte schwierig und die Übertragbarkeit auf das deutsche Bildungssystem eingeschränkt.

Wie bereits anfangs erwähnt, stammt der Großteil der SL-Forschung aus dem angloamerikanischen Bereich, allerdings gibt es auch einige ForscherInnen im deutschsprachigen Raum, die sich bereits mit SL-Forschung und der Wirkung des Formats auf die Studierenden beschäftigen. Gerholz, Liszt und Klingsieck (2015) konnten in einer Pilotstudie signifikante Unterschiede von Prä- zu Postzeitpunkt finden: „Die quantitativen Befunde zeigen signifikante Effekte der Veränderung über die Zeit hinsichtlich der Selbstwirksamkeit, des Selbstbildes und der Einstellung zu zivilgesellschaftlichem Engagement.“ (Gerholz et al., 2015, S. 0). Die Stärke der Veränderungen wird durch die wahrgenommene soziale Unterstützung und der Verknüpfung von Theorie und Praxis moderiert. Als Einschränkung dieser Ergebnisse nennen sie die geringe Fallzahl und die Auswahl der Instrumente (Selbstauskunft und qualitative Auswertung).

Außerdem gibt es eine Versuchsreihe von Reinders (2016), die verschiedene *kognitive* und *persönliche Entwicklungen* durch SL festhielten. Beginnend mit einer Pilotierung im Wintersemester 2007/2008 erhoben sie mit insgesamt 116 BWL-Studierenden in einem Prä-Post-Kontrollgruppen-Design Daten zehn verschiedener Variablen im Bereich *akademischer Merkmale* und *Persönlichkeitsmerkmale*. Sie erhoben Sozialklima (bezogen auf die Veranstaltung), Lernplanung (im Sinne der Selbstorganisation), Selbstregulation, Selbstwirksamkeit, subjektiver Lernerfolg, Berufsorientierung, Zielorientierung (bezogen auf das Studium), soziale Handlungswirksamkeit, Selbstbildveränderung sowie politische Partizipationsbereitschaft. Als Kontrollgruppen dienten eine Vorlesung ($n=46$), ein Lektüreseminar ($n=18$) und ein Praxisseminar ($n=20$), das als Begleitung zu einem studiumsrelevanten Praktikum entwickelt wurde. Es wurden signifikante Effekte bzgl. des Lernklimas (zum Messzeitpunkt 2 im Post-Hoc-Test), der Zielorientierung (nur zum MZP1),

des subjektiven Lernerfolgs sowie der Selbstbildveränderung gefunden. Bei allen anderen Konstrukten konnten keine Veränderungen, die auf den Veranstaltungstyp zurückzuführen sind, gezeigt werden. In einer weiteren Studie aus dem Wintersemester 2010/11 nahmen 81 Pädagogikstudierende teil. Es wurde wieder ein Prä-Post-Kontrollgruppendesign mit einer Wartekontrollgruppe zugrunde gelegt. Die Studierenden wurden zu SL verpflichtet, konnten aber wählen, ob sie zu Beginn ($n=41$) oder ein Semester später ($n=40$) an der SL-Komponente teilnehmen wollten. Im Fokus dieser Untersuchung standen vor allem immer wieder diskutierte Selektionseffekte von Studierenden, die an SL-Seminaren partizipieren (häufig vor allem Frauen und Personen, die sich ohnehin bereits sozial engagierten). Zusätzlich wurden in weiten Teilen die gleichen Konstrukte wie in der Pilotierungsstudie erfasst: Studienzufriedenheit, Sozialklima, Lernplanung, Zielorientierung, Metakognition (Nachdenken über Lernverhalten), subjektiver Lernerfolg, Reflexion (der Veranstaltungsinhalte), Praxistransfer (erlebte Wahrscheinlichkeit das Gelernte anwenden zu können) und politische Partizipationsbereitschaft. Es wurden signifikante Veränderungen in der ersten SL-Gruppe die Reflexion, den subjektiven Lernerfolg und den Praxistransfer betreffend gefunden sowie ein Effekt über die Zeit für die politische Partizipation. Alle anderen Variablen zeigten keine signifikanten Effekte. Die differenzierteste Studie wurde im Wintersemester 2012/13 bzw. 2013/14 wieder in einem Prä-Post-Kontrollgruppendesign durchgeführt. Die Seminare wurden parallelisiert und in EG sowie KG von den gleichen Dozierenden durchgeführt. Die Daten wurden mit Hilfe von Selbstauskunftsfragebögen sowie einem Lerntagebuch in den SL-Veranstaltungen erhoben. Es wurden insgesamt 14 Variablen erfasst: Metakognition, Reflexionsfähigkeit, Reflexionsphasen (fünf Subkategorien), Theorie-Praxis-Reflexion, Berufstransfer (Anwendbarkeit des Gelernten in der späteren Berufspraxis), Fachinteresse (geweckt durch die besuchte Veranstaltung), Fachkompetenz, Methodenkompetenz, Personale Kompetenz, Praxisqualität, Sozialklima, identifizierte

Lernmotivation, soziale Handlungswirksamkeit und Selbstbildveränderung. In dieser Untersuchung konnten signifikante Unterschiede im Sozialklima, dem Fachinteresse, der Reflexionsphase 1 (Beschreibung von pädagogischen Situationen), Reflexionsphase 2 (Erinnerung an Gedanken und Gefühle in einer bestimmten Situation), Reflexionsphase 3 (theoretische Einordnung zum MZP3), Reflexionsphase 4 (theoretische Bewertung), in der Fachkompetenz (auch zu MZP3), personalen Kompetenz, Berufstransfer und Selbstbildveränderung der Studierenden im SL-Seminar verglichen mit den Teilnehmenden im konventionellen Seminar ermittelt werden. Die berichteten Effekte sind allerdings in den meisten Fällen recht gering. Es lassen sich also auch hierzulande signifikante Unterschiede zwischen Teilnehmenden von SL- und konventionellen Veranstaltungsformaten finden.

2.2.4 Empathie und Service-Learning

In vielen Studien wurden immer wieder positive Effekte von SL auf verschiedenste Konstrukte berichtet. Empathie gehört zu den meistgenannten Emotionskomponenten in der SL-Forschung (Langstraat & Bowdon, 2011), wurde allerdings in nur wenigen Studien als Wirkung überprüft. Es gibt verschiedene Studien, die Empathie als Teil beispielsweise prosozialer Komponenten oder moralischer Entwicklung miterheben, allerdings nicht standardisiert und nicht explizit als eigenes Konstrukt (Batchelder & Root, 1994; Brown, 2011; Ferrari & Worrall, 2000; Govekar & Rishi, 2007; Hey, Slivovsky, Self, Widmann & Taylor, 2014; Jones & Hill, 2001; Jones & Abes, 2004; Potthoff et al., 2000; Prentice & Robinson, 2010; Soukup, 1999). Deswegen werden diese Studien hier nicht mit einfließen. Zwei Studien haben sich allerdings explizit mit dem Effekt von SL auf Empathie beschäftigt (Lundy, 2007; Wilson, 2011). In der Studie von Lundy (2007) wurde Empathie erstmalig in Zusammenhang mit SL als Wirkvariable erhoben. In der Studie wurde ein SL-Kurs in *Entwicklung über die Lebensspanne* an einer amerikanischen Universität in einem Prä-Post-

und Kontrollgruppen-Design untersucht. Die teilnehmenden 192 Studierenden konnten sich eine von drei Veranstaltungs-Optionen aussuchen: Die SL-Veranstaltung, eine Veranstaltung in deren Rahmen drei Personen interviewt werden sollten und ein Lektüreseminar, in dem verschiedene Studien gelesen und aufbereitet werden mussten. In allen Seminarformaten wurden vier multiple choice Klausuren geschrieben und Empathie wurde über das Selbstauskunftsverfahren von Mehrabian und Epstein (1972) erhoben. Im SL-Seminar leisteten die Studierenden ihr Engagement in einer Einrichtung ihrer Wahl, die in Verbindung zum Thema stand (bspw. Vorschule, Seniorenheim, betreutes Wohnen, o.ä.) und verbrachten dort jede Woche zwei Stunden für insgesamt 12 Wochen. Die Reflexion fand in geschriebener Form statt und die Teilnehmenden fertigten einen 10-12 seitigen Bericht an, in dem sie theoretische Texte mit ihren gemachten Erfahrungen verbanden und sich auf ihre Entwicklung konzentrierten. In den Klausuren schnitten die Teilnehmenden der Kurse alle gleich ab, allerdings hatten die SL-Studierenden in späteren Klausuren bessere Ergebnisse als die Studierenden der anderen beiden Optionen. Empathie betreffend konnten nach dem Kurs höhere Werte in der SL-Gruppe gefunden werden (verglichen mit den anderen Gruppen) und diese Gruppe war auch die einzige, in der es von Prä- zu Post-Messung steigende Werte gab. In beiden anderen Gruppen nahmen die Werte sogar ab. Kritisch bemerkt die Autorin den Einsatz des Selbstauskunftsfragebogen an, bei dem ein verzerrender Einfluss sozialer Erwünschtheit nicht ausgeschlossen werden kann und regt die Nutzung robusterer Verfahren an.

Ebenfalls mit dem Konstrukt Empathie beschäftigte sich Wilson (2011). Die Studie wurde an einer südkalifornischen Universität in Psychologie- und Soziologiekursen durchgeführt. In den Kursen ging es um die Rolle von Ehe und Familie in der Gesellschaft und in der Entwicklung sowie das Aufrechterhalten sozialer Beziehungen. 52 Studierende nahmen an der SL-Version des Kurses teil, bei dem 15 Stunden soziales Engagement in einer Agentur

geleistet wurden (EG) und 13 Studierende in dem Teil des Kurses, in dem ein relevantes Buch gelesen und diskutiert wurde (KG). Die Daten wurden aus vier benoteten, schriftlichen Arbeiten gewonnen, die in beiden Gruppen über die Dauer des Kurses geführt wurden und in denen es Prompts gab, um über bestimmte Themen bzw. Erfahrungen zu reflektieren. Für die Arbeiten gab es Feedback und es wurden Leitfragen zum Erlangen neuer Perspektiven, Empathie und Selbsterkenntnis gestellt. Es wurden keine Unterschiede gefunden zwischen den Gruppen bzgl. neuer Perspektiven und Selbsterkenntnis (die in beiden Gruppen recht niedrig ausfiel), nur Empathie betreffend gab es einen signifikanten Unterschied (Fisher's exact test $p=.020$). Die Autorin sieht damit als bestätigt an, dass SL einen positiven Effekt auf empathische Reaktionen hat. Sie regt allerdings für zukünftige Forschung an, ein Prä-Post-Design zu nutzen, demografische Daten zu erheben, sowie die Teilnahmemotivation festzuhalten. Da beide Studien Lehrveranstaltungsevaluationen der eigenen Veranstaltungen erheben, sind die Studien eher exploratorisch zu betrachten. Das Design in der Studie von Lundy (2007) ist stark, allerdings ist das Erhebungsinstrument für Empathie ein Selbstauskunftsfragebogen älteren Jahrgangs, für das es heutzutage andere und psychometrisch überzeugendere Alternativen gibt. Wilson (2011) wählte einen induktiven, inhaltsanalytischen Ansatz, der in Kombination mit anderen Verfahren sicherlich eine sinnvolle Testbatterie darstellen würde.

Etwas indirekter wurde Empathie durch eine Studie von Jarrell et al. (2014) in der Pflegekraftausbildung erhoben. In einem SL-Kurs nahmen Krankenpflegekräfte teil und leisteten ihr Engagement in einer Obdachlosenunterkunft und in Wohnungen von Personengruppen mit niedrigem sozio-ökonomischem Status ab. Das Thema bezog sich auf den Inhalt Armut und sollte dazu dienen, Vorurteile den entsprechenden Zielgruppen gegenüber abzubauen. Die Messungen wurden vor und nach dem 15-wöchigen Kurs über verschiedene Kohorten hinweg durchgeführt. Empathie wurde als empathische Einstellungen

implizit in den genutzten Fragebögen zu Einstellungen Armut und Pflegekräften gegenüber erhoben. Es konnte im Vergleich zur Kontrollgruppe ein deskriptiver Unterschied festgestellt werden, der allerdings nicht signifikant wurde (SL: $M=24$, $SE = .53$, Non-SL: $M=.05$, $SE = .33$). Die AutorInnen führten das auf die geringen Stichprobengrößen zurück (EG/Post: $n=25$; KG/Post: $n=51$), sehen jedoch trotzdem eine klare, positive Tendenz bezogen auf die Wirkung von SL-Veranstaltungen auf Empathie.

Eine weitere Forschungsstudie (Marshall, Lawrence, Lee Williams & Peugh, 2015) untersuchte die Wirkungen einer SL-Mentoring-Veranstaltung nur für Frauen. Es handelt sich hierbei um ein Mentoring-Programm, bei dem Studentinnen ein Jahr lang eine Schülerin in der Middle-School, die sich zu dem Zeitpunkt in einer schwierigen Lebenssituation befindet, beistehen und sie durch Einzel-, aber auch Gruppentreffen unterstützen. Die studentischen Teilnehmerinnen werden durch die Universität in Form von Beratung und Reflexionseinheiten begleitet. In der Studie wurden neben *selbst empfundener Kompetenz* (im Sinne akademischer Leistungen), *Verbundenheit* und *Autonomie* auch ein Konstrukt, das hier *Ethnocultural Empathy* genannt wird, erhoben. Es nahmen insgesamt 641 Studentinnen teil. 227 davon besuchten das SL-Programm, 414 waren in der Kontrollgruppe und besuchten keinen SL-Kurs ($n=105$) bzw. waren im vergangenen Jahr einer ehrenamtlichen Tätigkeit nachgegangen ($n=230$). Die interessierenden Variablen wurden via Selbstauskunftsfragebögen erhoben. Für die *Ethnocultural Empathy* wurde die Subskala *Empathic Feeling and Expression* (15 Items) der *Scale of Ethnocultural Empathy* (SEE; Wang et al., 2003) genutzt, ein Instrument, das kulturelle Kompetenz erfasst. Die interne Konsistenz der Subskala lag bei $\alpha=.91$. Alle Daten wurden in einem Prä-Post-Design erhoben. Die AutorInnen fanden signifikant höhere Werte in *Ethnocultural Empathy* (und auch *Autonomie*) bei SL Teilnehmerinnen, die in dem Programm eine hohe Peer-Unterstützung erlebten. Sie sehen darin einen direkten Hinweis auf die wichtige Rolle der Unterstützung und Betreuung in SL-Programmen (gerade auch das

Klima untereinander), um positive Wirkungen bei den Teilnehmenden erreichen zu können, schränken gleichzeitig aber ein, dass die Ergebnisse nur schlecht generalisierbar sind.

2.2.4.1 Rahmenbedingungen für Empathieentwicklung

In einer 2016 entstandenen Bachelorarbeit (Junghenn, 2016) wurde untersucht, welche Rahmenbedingungen eine SL-Veranstaltung haben sollte, um es den Studierenden zu ermöglichen, (unter anderem) Empathie entwickeln zu können. Hierzu wurden ExpertInnen-Interviews mit nationalen und internationalen Theoretiker- und PraktikerInnen im Bereich SL geführt. *ExpertIn* wurde dahingehend definiert, dass die in Frage kommenden Personen zu diesem Zeitpunkt an einer Universität angestellt waren und möglichst schon langjährige Erfahrungen im praktischen, aber auch theoretischen Bereich von SL haben sollten. Insgesamt wurden neun ExpertInnen interviewt; die praktische Lehrerfahrung mit dem Format lag im Schnitt bei 8.59 Jahren ($SD=3.22$ Jahre), die theoretische-forschende Betätigung im Schnitt bei 8.17 Jahren ($SD=3.49$ Jahre). Für die Befragung wurde ein Interviewleitfaden entwickelt, der den Teilnehmenden im Vorfeld zugeschickt wurde und die freien Antworten anhand der qualitativen Inhaltsanalyse nach Mayring (2015) induktiv ausgewertet. Die so gefundenen Kategorien wurden den K-12 Qualitätsstandards zugeordnet (Junghenn, 2016). Die ExpertInnen nannten auch hier die *Dauer und Intensität des Services* als entscheidend, aber auch das *Fördern von Interaktionen* zwischen allen Service-Parteien. Genauso sehen sie auch hier die *Einbindung ins Curriculum* als förderlich an und ein *Agieren auf Augenhöhe* als Kernmerkmal der Partnerschaft. In Bezug auf *bedeutungsvollen Service* halten sie vor allem für wichtig, dass Überforderungen vermieden werden sollten und bezüglich des *Mitsprachrechts*, dass eine Partizipation am Programm erlebt wird. Ebenfalls wird zum Entwickeln von empathie-relevanten Facetten der *Diversitätsaspekt* als hilfreich gesehen. Vor allem bei der Arbeit mit KlientInnen, die weniger privilegiert sind und auch unter zu

Hilfenahme von kooperativen Methoden als Teil der Veranstaltung, um unterschiedliche Perspektiven Kennenlernen zu können. Natürlich spielt auch hier *Reflexion* eine bedeutende Rolle, ebenso die *Betreuung*, *Beratung* und die *Öffentlichkeit des SL-Projektes* als Methoden zum Beobachten des Fortschritts (Junghenn, 2016).

2.2.5 Service-Learning in der Lehramtsausbildung

Wie bereits beschrieben, kann SL vielfältige Wirkungen auf die teilnehmenden Studierenden haben. Aber welche spezifischen Wirkungen konnten bei Lehramtsstudierenden festgestellt werden? Die Lehramtsausbildung bildet einen passenden Rahmen, um SL-Veranstaltungen anbieten zu können, auch und gerade um eine weitere Verzahnung von Theorie und Praxis möglich zu machen (Wade, 1995). Die Forschung in diesem Bereich stammt erneut aus dem angloamerikanischen Raum, so dass die Übertragung auf das deutsche LehrerInnen-Ausbildungssystem natürlich nur bedingt möglich ist, allerdings Anhaltspunkte liefert, an denen es sich orientieren lässt. Generell gilt, dass SL in der Lehramtsausbildung auch hier reale Bedarfe in der Gesellschaft adressieren sollte, besonders um sich von reinen Praktika im Schulkontext abgrenzen zu können (Erickson & Anderson, 2012). SL ist für angehende LehrerInnen ein hilfreiches Lehrveranstaltungsformat, das ihnen als didaktische Methode an die Hand gegeben wird (Erickson & Anderson, 2012). Zudem kann es die moralische Entwicklung und Sensibilisierung für eine SchülerInnen-zugewandte Lehrweise unterstützen (Root, 2012). An der Universität von Iowa gibt es ein Programm für GrundschullehrerInnen, in denen sie lernen, SL für SchülerInnen an Schulen anzubieten (Wade, 1995). In einer Evaluationsstudie konnte die Autorin feststellen, dass direkter Service in Form von Mentoring o.ä. die größte Wirkung mit sich brachte und dass die Durchführenden darauf gefasst sein sollten, mit den Emotionen von SchülerInnen umzugehen (Wade, 1995). Hier zeigte sich auch noch einmal die Wichtigkeit von Reflexion, um die Erfahrungen in einen größeren Kontext

einzubetten und über Moral und gesellschaftliche Mitwirkung zu sprechen. Zudem bot das Programm den Studierenden einen guten Eindruck der strukturellen Schwierigkeiten, die mit derartigen Formaten einhergingen (Implementation in das Curriculum, das Herstellen eines Kontaktes zu zivilen PartnerInnen, Akzeptanz im Kollegium, usw.). Die Erfahrungen des Programms ermutigten die teilnehmenden Studierenden, trotzdem SL in ihre spätere Lehrtätigkeit mit einzubeziehen. Nichtsdestotrotz konnte in einer weiteren Studie gezeigt werden, dass LehrerInnen, die ein langes SL-Projekt während des Studiums besucht hatten, in ihrem ersten Jahr genauso viele SL-Projekte umsetzten wie AbsolventInnen, die kein SL im Studium besucht hatten (George et al., 1995; zitiert nach Root, 2012). Sie begründeten das mit dem hohen Zeitaufwand von SL, der im ersten praktischen Jahr nur schwer aufzubringen sei. Weitere Studien zeigten als Effekte eine Erhöhung des *Selbstwertes*, der *Selbstwirksamkeit*, eine größere *Verpflichtung und Verbindlichkeit dem Lehrberuf* gegenüber sowie eine höhere *Anteilnahme gegenüber anderen* (Root, 2012). In einem anderen Programm, in dem die Studierenden Viert- und Sechstklässler mit niedrigen sozio-ökonomischen Hintergründen unterstützten, konnten die AutorInnen eine Steigerung der *Diversitäts-Sensibilität* und eine größere Offenheit der Studierenden, sich mit sozialkritischen Fragen auseinanderzusetzen, feststellen (Baldwin, Buchanan & Rudisill, 2007). Ebenfalls wurden sie in diesem Projekt mit ihren eigenen Einstellungen und Vorurteilen konfrontiert, mit denen sie sich kritisch auseinandersetzten. Ähnliche Resultate wurden in einer weiteren explorativen Studie von Zygmunt et al. (2018) gefunden. In einem SL-Kurs begleiteten Studierende Familien in verschiedenen Gemeinden und konnten auch hier ihre *Einstellungen* hinterfragen, *Diversitätssensibilität* entwickeln und erlebten gesteigerte *Verbindung zu anderen*.

2.3 Lehramtsausbildung und Lehrberuf

2.3.1 Aufbau der Ausbildung in Deutschland

Die Lehramtsausbildung in Deutschland ist mehrschrittig und findet an Universitäten oder an pädagogischen Hochschulen statt, sowie während der Praxisphasen und später dem Vorbereitungsdienst auch an Schulen. Sie ist in der Ausgestaltung der Rahmenbedingungen und Inhalte in weiten Teilen Sache der Länder. Die Vorgaben der jeweiligen Kultusministerien sind Richtlinien, die von den Universitäten spezifisch umgesetzt werden und daher sogar im Bundesland selbst von Universität zu Universität variieren. Seit 1948 gibt es eine ständige Konferenz der KultusministerInnen in Deutschland, an der die KultusministerInnen bzw. BildungssenatorInnen der Länder teilnehmen und verschiedene Maßnahmen vereinbaren, um eine Qualitätssicherung und zumindest eine gewisse Standardisierung deutscher Hochschulabschlüsse gewährleisten zu können (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland., 2004).

2.3.2 Standards in der Lehrerbildung

Im Rahmen der Kultusministerienkonferenz und den beschlossenen Maßnahmen wurden sogenannte Bildungsstandards vereinbart – sowohl für die Kompetenzerreichung der Schülerschaft, als auch Standards für die LehrerInnenausbildung (Kultusministerkonferenz, 2014; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland., 2004). Folgende Punkte wurden als essentiell für den Beruf der Lehrkraft eingestuft:

1. **Bildung und Erziehung** (Begründung und Reflexion von Bildung und Erziehung in institutionellen Prozessen)
2. **Beruf und Rolle des Lehrers** (Lehrerprofessionalität; Berufsfeld als Lernaufgabe; Umgang mit berufsbezogenen Konflikt- und Entscheidungssituationen)

3. **Didaktik und Methodik** (Gestaltung von Unterricht und Lernumgebungen)
4. **Lernen, Entwicklung und Sozialisation** (Lernprozesse von Kindern und Jugendlichen innerhalb und außerhalb von Schule)
5. **Leistungs- und Lernmotivation** (Motivationale Grundlagen der Lern-, Leistungs- und Kompetenzentwicklung)
6. **Differenzierung, Integration und Förderung** (Diversität und Heterogenität als Bedingungen von Schule und Unterricht)
7. **Diagnostik, Beurteilung und Beratung** (Diagnose und Förderung individueller Lernprozesse; Leistungsmessungen und Leistungsbeurteilungen)
8. **Kommunikation** (Kommunikation, Interaktion und Konfliktbewältigung als grundlegende Elemente der Lehr- und Erziehungstätigkeit)
9. **Medienbildung** (Umgang mit Medien unter konzeptionellen, didaktischen und praktischen Aspekten)
10. **Schulentwicklung** (Struktur und Geschichte des Bildungssystems; Strukturen und Entwicklung des Bildungssystems und Entwicklung der einzelnen Schule)
11. **Bildungsforschung** (Ziele und Methoden der Bildungsforschung; Interpretation und Anwendung ihrer Ergebnisse) (Kultusministerkonferenz, 2014, S. 5)

Aus diesen inhaltlichen Bereichen wurden Kompetenzbereiche abgeleitet, für die Standards formuliert wurden. Die Kompetenzbereiche umfassen: 1. Unterrichten, 2. Erziehen, 3. Beurteilen und 4. Innovieren. Der erste Bereich *Unterrichten* weist LehrerInnen als ExpertInnen für Lehren und Lernen aus und setzt sich aus drei Kernkompetenzen zusammen. Der zweite Bereich nennt *Erziehen* als Aufgabe von LehrerInnen und gliedert sich ebenfalls in drei Teilkompetenzen auf. Beim *Beurteilen* geht es um die Aufgabe von Lehrkräften gerecht und verantwortungsbewusst zu beurteilen, sowie objektiv und SchülerInnen-orientiert zu beraten. Dieser Bereich unterteilt sich in zwei Kompetenzen. Der letzte Kompetenzbereich *Innovieren*

soll Lehrkräfte befähigen sich und ihre Fähigkeiten beständig weiterzuentwickeln. In diesem Bereich gibt es auch wieder drei Kompetenzen, die es umzusetzen gilt (alle Kompetenzbereiche und Kompetenzen siehe Tabelle 10).

Tabelle 10

Übersicht der Standards für Lehrerbildung; adaptiert nach Kultusministerkonferenz, 2014

Kompetenzbereich	Kompetenzen
1. Unterrichten	<ol style="list-style-type: none"> 1. Unterrichtsplanung (unter Berücksichtigung individueller Lernvoraussetzungen, sowie Entwicklungsstand; sachlich und fachlich korrekt) 2. Schaffen geeigneter Lehr-Lern-Settings (Motivation der SchülerInnen; Unterstützung beim Erkennen von Zusammenhängen und Anwenden von Gelerntem) 3. Förderung der Fähigkeit zu selbstbestimmtem Lernen und Arbeiten
2. Erziehen	<ol style="list-style-type: none"> 1. Diversitätssensibilität (Kennen der sozialen und kulturellen Lebensbedingungen, eventueller Benachteiligungen, Beeinträchtigungen und Hindernisse; Einflussnahme auf individuelle Entwicklung) 2. Werte- und Normvermittlung (Wertschätzung und Anerkennung von Diversität; Unterstützung von selbstbestimmtem Urteilen und Handeln) 3. Vermittlung von Lösungsansätzen bei Schwierigkeiten und Konflikten
3. Beurteilen	<ol style="list-style-type: none"> 1. Diagnostik (Lernvoraussetzungen und Lernprozesse; gezielte Förderung; Lernenden- und Elternberatung) 2. Lernstandsermittlung (Erfassung der Leistungsentwicklung; Beurteilung des Lernens und Leistungen anhand nachvollziehbarer Beurteilungsmaßstäbe)
4. Innovieren	<ol style="list-style-type: none"> 1. Verantwortungs- und Pflichtbewusstsein (Bewusstsein über besondere Anforderungen des Berufs als ein öffentliches Amt) 2. Verständnis des Berufs als ständige Lernaufgabe 3. Beteiligung an Planung und Umsetzung schulischer Projekte und Vorhaben

2.3.3 Lehramtsstudium in Hessen und an der Goethe-Universität

Da die vorliegenden Untersuchungen mit Lehramtsstudierenden an der Johann-Wolfgang-Goethe Universität in Frankfurt durchgeführt wurden, beschränken sich die folgenden Ausführungen auf die Ausbildungsspezifika in Hessen bzw. der betreffenden Universität. Die erste Phase der Ausbildung findet an den Universitäten statt und wird von Theorie (durchsetzt von unterschiedlichen Praxisphasen) dominiert. Sie endet mit der ersten Staatsprüfung vor der Hessischen Lehrkräfteakademie und geht anschließend über in den Vorbereitungsdienst

(ehemals: Referendariat), der vor allem praxisorientiert ist (mit begleitenden Seminaren zur Vertiefung und Supervision) und an einer Schule absolviert wird (Kultusministerkonferenz, 2014). Dem folgt dann die zweite Staatsprüfung. In der ersten Phase (Universität) findet eine schulformbezogene Ausbildung statt, die in der Regel 6 – 8 Semester dauert:

1. Lehramt an Grundschulen (L1)
2. Lehramt an Haupt- und Realschulen (L2)
3. Lehramt an Gymnasien (L3)
4. Lehramt an Förderschulen (L5)
5. Studiengang Wirtschaftspädagogik (keine Staatsprüfung, sondern Bachelor/Master)

Alle Lehramtsstudierenden studieren mindestens zwei Fächer und besuchen zum Anfang ihres Studiums schwerpunktmäßig Vorlesungen und Seminare in den Bildungswissenschaften (ABL - Akademie für Bildungsforschung und Lehrerbildung, 2018a). Dieser Studienbereich teilt sich in sechs Module auf und vermittelt die Grundlagen für pädagogisch-praktische Aufgaben des späteren Berufsfeldes. Die Inhalte orientieren sich an den Standards für die Lehrerbildung der Kultusministerkonferenz (2014) bzw. den daraus abgeleiteten Kompetenzbereichen. Im Modul BW-A geht es um eine *Einführung in die Bildungswissenschaften*, BW-B bis BW-E umfassen die im vorigen Kapitel beschriebenen Kompetenzbereiche *Unterrichten, Erziehen, Beurteilen* und *Innovieren* und im letzten Modul BW-F werden Praxisphasen absolviert, die von theorie-vermittelnden Seminaren begleitet werden (entweder als schulpraktische Studien oder in L3 als Praxissemester; ABL - Akademie für Bildungsforschung und Lehrerbildung, 2018a). Für das bildungswissenschaftliche Cluster werden je nach Studiengang 50 bis 60 CPs benötigt, um mit noch weiteren nachgewiesenen CPs in anderen Bereichen (Fachdidaktiken, Fachwissenschaften) die Staatsprüfungen antreten zu dürfen. Diese Prüfungen finden in schriftlicher und mündlicher Form statt und werden vor Dozierenden der Fachbereiche und VertreterInnen der Hessischen Lehrkräfteakademie

abgelegt (Hessische Lehrkräfteakademie, 2018; HMWK, 2018). Danach schließt der 21-monatige pädagogische Vorbereitungsdienst an, der mit der zweiten Staatsprüfung und dem Erhalt der Lehrbefähigung schließt (Hessische Lehrkräfteakademie, 2018).

2.3.4 Service-Learning im Lehramtsstudium an der Goethe-Universität

Das Curriculum für Lehramtsstudierende an der Goethe-Universität ist recht frei gestaltbar, so dass neben konventionellen Seminaren und Vorlesungen auch praxisorientierte Veranstaltungen wie SL Platz finden können (ABL - Akademie für Bildungsforschung und Lehrerbildung, 2018a). Nicht immer ist es einfach in die bestehenden Regularien für Lehrveranstaltungen SL sinnvoll einzugliedern, doch gibt es bereits Bestrebungen seitens der ABL Anpassungen am Curriculum vorzunehmen, womit eine Implementation erleichtert werden soll (ABL - Akademie für Bildungsforschung und Lehrerbildung, 2018b). Seit 2016 gibt es auch in der ABL eine Ansprechpartnerin für SL, die neben der Arbeitsstelle SL (als zentrale Einrichtung der Universität zur Beratung von Dozierenden und Studierenden) speziell für Lehrveranstaltungen im Rahmen der LehrerInnenbildung für Anfragen und Beratung zur Verfügung steht (ABL - Akademie für Bildungsforschung und Lehrerbildung, 2018c). Hiermit soll der Einsatz innovativer Veranstaltungsformen gefördert werden. Neben dem bereits beschriebenen positiven Nutzen für die Teilnehmenden selbst bietet das Veranstaltungsformat auf der Metaebene eine Bereicherung für den didaktischen Methodenkoffer (Erickson & Anderson, 2012). SL ist auch ein im Schulkontext anwendbares didaktisches Konzept (vgl. Stiftung Lernen durch Engagement, 2018), das natürlich auch in diesem Kontext verschiedenste Kompetenzen fördern kann. Somit bietet der Besuch einer SL-Veranstaltung im Studium einen mehrfachen Nutzen für die angehenden Lehrkräfte. An der Goethe-Universität sind die meisten SL-Seminare im bildungswissenschaftlichen Cluster angesiedelt und an vielen Stellen ist die Differenzierung zwischen *normalen* praxisorientierten- und SL-

Veranstaltungen nicht ganz einfach zu klären, da die Zielgruppe in beiden Fällen eine SchülerInnenpopulation ist. SL unterscheidet sich lediglich darin, dass Subpopulationen adressiert werden und häufig SchülerInnen mit besonderen Förderbedarfen oder mit schwierigem sozialem Hintergrund Unterstützung bekommen. Seit über 10 Jahren werden solche Praxisprojekte, die sich unter dem Begriff SL eingruppiert lassen, von der Arbeitsstelle für Diversität und Unterrichtsentwicklung - Didaktische Werkstatt im Fachbereich Erziehungswissenschaften angeboten (Didaktische Werkstatt, 2018). Hier gibt es ein buntes Portfolio von ca. 11 verschiedenen Projekten, die bestimmte Schülerschaften unterstützen und diese Praxistätigkeiten mit Seminaren begleiten, in denen theoretische Inhalte vermittelt und Reflexionsmöglichkeiten bereitgestellt werden. Die Seminare sind meistens für verschiedene Lehramtsfächer geöffnet, für ein komplettes Schuljahr konzipiert und haben in der Regel 10 bis 15 Teilnehmende. Für das Engagement in diesen Veranstaltungen und um den Anreiz für eine ein-jährige Teilnahme zu erhöhen, bekommen die Studierenden eine Vergütung für die praktische Tätigkeit in Höhe von 10€/Stunde. Die Besonderheit der Praxisprojekte besteht zusätzlich darin, dass Lehramts- und Erziehungswissenschaftsstudierende gemeinsam teilnehmen und mit diesem Ansatz eine interdisziplinäre Zusammenarbeit trainiert werden kann. Ebenfalls interdisziplinär und vom Konzept her einjährig aufgestellt ist die Veranstaltung von Dr. Anne Seifert, bei der es um Migration und Teilhabe geht und die aus verschiedenen Kleinprojekten besteht, sowie das TELLUS-Projekt, das für L3-Studierende als Praxissemester angerechnet werden kann. Hier arbeiten Studierende verschiedenster Richtungen zusammen, um SchülerInnen in Sprachförderprogrammen an beruflichen Schulen zu unterstützen. Weitere Programme sind für alle Lehramtsrichtungen geöffnet und auf ein oder zwei Semester ausgelegt (siehe Anhang D1).

2.3.5 Professional Vision

Professional Vision (oder *professionelle Wahrnehmung*) bezeichnet die wissensbasierte und spezielle Wahrnehmung aus dem Blickwinkel einer bestimmten Profession bzw. Berufsbildes heraus (Goodwin, 1996; Seidel, Blomberg & Stürmer, 2010). Es entsteht durch das Aneignen von Wissen in diesem Bereich, das sich von Handlungswissen anderer Tätigkeitsbereiche unterscheidet und eine spezifische Bedeutung für diesen Beruf hat, so dass Situationen und Dinge auf eine ganz bestimmte Art wahrgenommen werden (Goodwin, 1996). Goodwin (1996) beschreibt diese Prozesse generell als sozial erlernte Abläufe und Vorgehen, die vor dem Hintergrund eines Berufsbildes gewachsen sind und sozial situierte Ereignisse darstellen (vgl. auch Putnam & Borko, 2000). Goodwin (1996) beschreibt es folgendermaßen:

All vision is perspectival and lodged within endogenous communities of practice. An archaeologist and a farmer see quite different phenomena in the same patch of dirt (for example, soil that will support particular kinds of crops versus stains, features, and artifacts that provide evidence for earlier human activity at this spot). (S.606)

Er sieht diesen Vorgang als Erlernen eines Kodierungsschemas, mit dessen Hilfe Situationen und Wahrnehmungen betrachtet und beurteilt werden. Dieser Aspekt liegt ihm zufolge auch wissenschaftlichem Arbeiten zugrunde, da so das Bewerten und Interpretieren bestimmter Sachverhalte durch mehrere Personen erleichtert wird und dadurch ähnliche Ergebnisse geliefert werden können. Hierbei spielen gesteuerte Aufmerksamkeits- und Informationsverarbeitungsprozesse eine entscheidende Rolle (Seidel et al., 2010), natürlich auch im LehrerInnenberuf. "Professionelle Wahrnehmung wird entsprechend in zwei Komponenten unterteilt: (1) noticing –Identifikation relevanter Situationen und Ereignisse im Unterrichtsgeschehen; (2) knowledge-based reasoning –wissensgesteuerte Verarbeitung identifizierter Situationen und Ereignisse." (Seidel et al., 2010, S. 297). Diese Fähigkeiten lassen sich über Novizen, Anfänger mit Erfahrung, sowie ExpertInnen gut diskriminieren. In

einer Studie von Sabers, Cushing und Berliner (1991) wurden LehrerInnen, die einer dieser drei Erfahrungscluster zugeordnet werden konnten, an drei Monitoren parallel verschiedene Klassensituationen gezeigt, die sie beobachten sollten, um im Anschluss Fragen bezüglich der beobachteten Instruktions- und Classroom-Management-Techniken zu beantworten. Sie stellten fest, dass die ExpertInnen erwartungskonform sehr viel differenzierter in der Lage waren, die Situationen gleichzeitig zu beobachten, elaborierter zu verstehen und Rückschlüsse zu ziehen als die beiden anderen Gruppen (Palmer, Stough, Burdenski & Gonzales, 2005; Sabers et al., 1991). Professionelle Kompetenzen scheinen sich also von reinem Wissen über Lehrvielfalt zu einer anwendbaren Fähigkeit, wie und wann bestimmte Methoden zum Einsatz kommen, zu entwickeln und auch die Wahrnehmung entsprechender Situationen zu schärfen (Sabers et al., 1991; Seidel et al., 2010). Professional Vision stellt somit einen Indikator dar, ob sich konzeptuelles Wissen angeeignet wurde und inwiefern sowie in welchem Maße es bereits auf Situationen im Unterricht angewandt werden kann (Stürmer, Seidel & Schäfer, 2013). Aus wissenschaftlicher Perspektive handelt es sich bei diesen Kompetenzen um Aspekte *pädagogisch-psychologischen Wissens* wie Zielorientierung, Lernbegleitung oder –atmosphäre. Pädagogisch-psychologisches Wissen bezeichnet das Wissen, das LehrerInnen benötigen, um Lehr-Lern-Situationen fächerübergreifend optimal gestalten zu können und das neben fachspezifischem Wissen äußerst wichtig für das Unterrichten ist (vgl. Voss, Kunter & Baumert, 2011). Pädagogisch-psychologisches Wissen beinhaltet deklaratives und prozedurales Wissen in Themenbereichen wie Classroom Management, Lehrmethoden, Beurteilung von Klassenzimmeraktivitäten und Lernprozessen oder auch den Umgang mit der Diversität der Schülerschaft wie individuelle Eigenheiten einzelner SchülerInnen, die Einfluss auf das Unterrichtsgeschehen und den Rest der Klasse haben. Um dies alles tun zu können, benötigen LehrerInnen vor allem die Fähigkeit diese Besonderheiten wahrnehmen zu können und ein breites Repertoire an Methoden und Lerntools, die je nach individuellem Lernstand

der SchülerInnen eingesetzt werden können (Seidel et al., 2010; Voss et al., 2011). Dafür ist natürlich praktische Erfahrung hilfreich, die während der universitären Ausbildung zumindest in Teilen und sehr verstärkt im zweiten Ausbildungsteil (dem Vorbereitungsdienst) zum Tragen kommt (ABL - Akademie für Bildungsforschung und Lehrerbildung, 2018a; Hessische Lehrkräfteakademie, 2018). Um die Kenntnisse der angehenden LehrerInnen trotzdem so gut wie möglich bereits in der universitären Ausbildungsphase zu schulen, kommen in Kursen häufig Videoaufnahmen zum Einsatz (Stürmer et al., 2013). Dieses Werkzeug kann sehr nützlich sein, um Theorie und Praxis zu verbinden und bringt Studierende näher an die praktische Tätigkeit, ohne dass sie selbst vor einer Klasse stehen müssen. Besonders Studierende mit wenig Kenntnissen profitieren von solchen didaktischen Maßnahmen (Stürmer et al., 2013). Goodwin (1996) sieht in Videoaufnahmen eine Verdeutlichung von Lehrinhalten, die durch die visuelle Darstellung größere Bedeutung bekommen als nur allein durch die theoretische Vermittlung. Außerdem hält er das Aneignen professioneller Wahrnehmung nur für möglich, wenn der Bereich nicht rein kognitiv, sondern zusätzlich durch situiertes Lernen in einem passenden Setting erschlossen wird. Insofern kann auch SL ein geeignetes Lehr-Lernformat darstellen, um sich erforderliche Kompetenzen aneignen zu können.

Zudem spielen Feedbackprozesse eine wichtige Rolle, da die Lehrenden dadurch eine Rückmeldung der Schülerschaft bekommen, ob der Lernprozess mit Hilfe der angewandten Methoden erfolgreich gewesen ist oder Anpassung benötigt wird (Helmke & Lenske, 2013). Feedbackprozesse können in der professionellen Wahrnehmung auch unterstützend genutzt werden, um Instruktionen und Lehr-Lern Settings anzupassen und somit die Perspektive der Lernenden besser zu verstehen (Sherin, 2002). Instrumente zur Erfassung der Ausprägung von Professional Vision sind auch primär videobasiert (Seidel et al., 2010; Stürmer et al., 2013; Voss et al., 2011).

2.3.6 Empathie und Lehrberuf

Es wird also von LehrerInnen erwartet, dass sie den Unterricht managen können, den Lernstand ihrer Klasse beurteilen, flexibel auf individuelle Besonderheiten ihrer SchülerInnen eingehen und adäquate Lehr-Lern Methoden wählen können, um so die Lernprozesse bestmöglich zu steuern. Wie kann Empathie diese professionelle Wahrnehmung unterstützen und bereichern und damit eine wertvolle Kompetenz im Unterricht darstellen? Unstrittig ist, dass LehrerInnen einen Einfluss auf die Lernleistungen ihrer SchülerInnen haben (vgl. Kunter & Pohlmann, 2009). Vor allem am Anfang ihrer Tätigkeit ist es für neue LehrerInnen herausfordernd, die Perspektive ihrer SchülerInnen einzunehmen, um geeignete Lehr-Lernmaßnahmen vorzuschlagen und die Lernenden mit ihren gewählten Maßnahmen wirklich zu erreichen (Decety & Ickes, 2011). Im Sinne der professionellen Wahrnehmung ist allerdings gerade das unerlässlich, da diese Informationen im Sinne eines formativen Feedbacks Aufschluss über den Lernstand geben und ggf. zu veranlassende Förderungsmaßnahmen bedingen (Drummond, 1993). „A child with low self-esteem, for example, may need extra support and structure. A teacher’s understanding of the emotional state of the child can be as important to learning as any other factor.” (Cooper, 2011, S. 37). In den 50er Jahren des letzten Jahrhunderts entwickelte Carl Rogers den Ansatz der klientenzentrierten Psychotherapie, bei der *Achtung – Wertschätzung – Wärme, Einfühlung – Verständnis* sowie *Aufrichtigkeit – Echtheit* des Therapeuten als Kernkompetenzen gelungener psychotherapeutischer Prozesse angesehen werden (Rogers, 1957). In verschiedenen Untersuchungen konnte gezeigt werden, dass die Symptome der PatientInnen von TherapeutInnen mit diesen Haltungen entweder erheblich reduziert wurden oder sogar nicht mehr nachweisbar waren (Rogers & Carmichael, 1951; Rogers & Dymond, 1954). Diese Haltungen wurden und werden als synonym für empathisches Verhalten bzw. empathische Haltung gesehen (Rogers, 1957; Tausch & Tausch, 1998). Reinhard und Anne-Marie Tausch

(2008) übertrugen diese Ansätze auf den Schulkontext, obwohl LehrerInnen natürlich keine TherapeutInnen sind und das Setting einer Schulklasse sowie die Zielsetzung im Schulkontext von einer Eins-zu-Eins-Beratungssituation verschieden sind (Cooper, 2011). Die Tauschsahen trotzdem die gleichen Haltungen auch für diesen Berufsbereich als entscheidend an und erweiterten sie noch um die Komponente *fördernde - nicht dirigierende Einzeltätigkeiten*, die sie als logische Schlussfolgerungen der ersten drei Bereiche sahen. Der wichtige Einfluss von empathischen Lehrkräften konnte auch schon in der Forschung gezeigt werden und bekräftigte die Richtigkeit der vier Haltungen. Combs et al. (1969) fanden heraus, dass als effektiv eingeschätzte LehrerInnen besser darin waren, die Perspektive der SchülerInnen zu verstehen und größeres Interesse an Personen als an materiellen Gütern zeigten als ineffektiver eingeschätzte KollegInnen (zitiert nach Kravas, 1975). Zudem gibt es Anhaltspunkte, dass es Zusammenhänge zwischen dem empathischen Verständnis der Lehrkräfte und der Akzeptanz durch die SchülerInnen gibt, wobei auch eine Übereinstimmung in Haltung und Verhalten der Lehrkraft eine Rolle spielt (Emmerling, 1961; zitiert nach Kravas, 1975). Zudem wurden LehrerInnen, die von ihren SchülerInnen als empathischer eingeschätzt wurden, auch als bessere LehrerInnen wahrgenommen (Dixon & Morse, 1961). Zudem profitieren auch die SchülerInnen von ihren empathischen Lehrkräften. Aspy (1969) konnte zeigen, dass SchülerInnen von empathischen, respektvollen und authentischen Grundschullehrkräften signifikant höhere Werte in verschiedenen Bereichen eines Lesefähigkeitstests erreichten als diejenigen, deren LehrerInnen niedrigere Fähigkeiten hatten (zitiert nach Kravas, 1975). Ebenfalls konnte in Studien gezeigt werden, dass SchülerInnen, die sich von ihren LehrerInnen angenommen und verstanden fühlten, bessere Unterrichtsbeiträge beitrugen (höhere kognitive Verarbeitungsebenen), weniger Angst hatten sowie motivierter und offener waren (vgl. Tausch & Tausch, 1998). Für die Studie wurden Tonaufnahmen aufgenommen und die Lehrkräfte von Beobachtern anhand der personenzentrierten Haltungen eingeschätzt. Die

SchülerInnen wurden auch in Kategorien eingeteilt und gaben zusätzlich per Selbstauskunftsfragebögen an, wie ihre Gefühle bzgl. des Unterrichts waren. Ebenfalls zeigten Studien, dass SchülerInnen von empathischen LehrerInnen angstfreier im Unterricht waren, hier ebenfalls die Denkvorgänge auf höheren kognitiven Ebenen stattfanden. Außerdem wurden sie als selbstständiger eingeschätzt, sie zeigten eine größere Arbeitsmotivation, höhere Selbstzufriedenheit, es wurde ein besserer Klassenzusammenhalt beobachtet, vertrauensvollere Lehrer-Schüler-Beziehungen gepflegt, wodurch auch häufiger persönliche Dinge im Unterricht erzählt wurden. Weiterhin konnte gezeigt werden, dass auch die Selbstachtung und damit auch die Achtung vor anderen gefördert werden konnte (Tausch, 2008). Ähnliche Ergebnisse wurden auch in einer weiteren Studie gefunden (Höder, Tausch & Weber, 1976). Hier zeigten die SchülerInnen von empathischeren LehrerInnen wieder qualitativ höherwertige Unterrichtsbeiträge, selbstständigeres Denken und mehr unterrichtsbezogene Kommunikation. Tausch und Tausch (1998) erklären die Ergebnisse damit, dass die empathischeren LehrerInnen adaptiveres Verhalten zeigen würden und häufiger ein echter Kontakt mit den SchülerInnen stattfände. Über verschiedene Studien hinweg konnte zusätzlich gezeigt werden, dass empathischere Lehrkräfte angemessener auf die Gefühle der SchülerInnen reagierten, mehr Diskussionen zuließen, häufiger *echte*, authentisch interessierte Gespräche mit den SchülerInnen führten und Schülerwünschen mehr Platz einräumten (Dubs, 1995). Das Erkennen und Verstehen der Gefühle der SchülerInnen kann als ein wichtiger Faktor für die emotionale Gesundheit der Lernenden gesehen werden (Moustakas, 1966; zitiert nach Kravas, 1975). Es zeigt sich also, dass LehrerInnen, die die Gefühle ihrer SchülerInnen verstehen, damit einen wertvollen Beitrag zum Klassen- und Unterrichtsklima beitragen und damit das Lernen unterstützen (Kravas, 1975). Neben expliziten Lernzielen und Kohärenz, sowie Unterstützung durch die LehrerInnen stellt ein positives Klassenklima darüber hinaus einen wichtigen Faktor für Lernerfolg dar (vgl. Stürmer

et al., 2013). Ebenfalls kann Empathie ein Werkzeug sein, um sowohl Eltern, Kinder, als auch die Gesellschaft zu verstehen (Brown, 1996; zitiert nach Cooper, 2011).

Allerdings bringt empathisches Verhalten im Klassenzimmer auch seine Tücken und Einschränkungen mit sich. Cooper (2011) beschreibt die Problematiken folgendermaßen:

The possibility of deep personal relationships in classes of thirty is much less practicable. The concept of entering every private individual's perceptual world is not realistic for anyone, much less for the class teacher in crowded state schools, who, though aiming to know individuals, will inevitably have to prioritize their time and energy while having to manage the whole class and deliver the prescribed curriculum.

(S.35)

Imitation als Aspekt von Empathie fördert auch kognitive Verarbeitungsstile, die empfänglicher für die soziale Umwelt machen, aber auch ein besseres Erinnerungsvermögen für räumliche Positionen fördern (Decety & Ickes, 2011). Empathie kann allerdings auch als aversiv empfunden werden, wenn bspw. die Emotionsansteckung von belastenden Emotionen vordergründig wird und keine Abgrenzung mehr stattfindet, die Belastung also zur eigenen wird (Batson, 1997). Altmann (2015) beschreibt dies als *empathischen Kurzschluss*. Dieses Phänomen wird auch in Verbindung mit Burnout gesehen und das Wissen darum, kann den betreffenden Personen schon helfen, alternative Verhaltensweisen zu nutzen und präventiv zu agieren (vgl. Altmann, 2015). Als unterstützende Methodik können hier reflexive Prozesse gesehen werden, die helfen, das gebildete mentale Modell der Lebenswirklichkeit des anderen mit weiteren Informationen anzureichern sowie abzugleichen und eine emotionale Abgrenzung möglich zu machen. Dies geschieht optimalerweise in der Kommunikation mit dem anderen, bei der erlebte Emotionen und Wahrnehmungen geschildert werden und die andere Person rückmelden kann, ob diese der Realität entsprechen. Dieser Prozess kann

helfen, um persönliches Unwohlsein abzubauen und die Selbst-Andere-Differenzierung zu schärfen (Altmann, 2013).

2.4 Forschungsleitende Annahmen

Empathie ist ein Prozess, der uns wie bereits erwähnt in Teilen angeboren ist, allerdings auch im sozialen Kontext erlernt wird (Bischof-Köhler, 2009; Zahn-Waxler et al., 2001). Es ist ein wichtiges Element, um andere zu verstehen und auf dieser Basis zufriedenstellende, soziale Beziehungen zu führen (Dimberg, 1990; Rameson & Lieberman, 2009). Die Erforschung des Konstruktes hat in der Psychologie eine über hundert-jährige Tradition, aber trotzdem gibt es keine allgemein verbindliche Definition, auf die man sich bislang einigen konnte (vgl. Altmann, 2013). Im Gegenteil scheint sich das Konstrukt von Empathie immer der jeweiligen Messmethodik anzugleichen und dementsprechend immer wieder in Veränderung zu sein. Unstrittig ist zwischenzeitlich allerdings, dass es sich um ein multidimensionales Konstrukt handelt, das üblicherweise aus kognitiven und affektiven Komponenten besteht, und bei dem es einer Unterscheidung zwischen dem *Selbst* und *den Anderen* bedarf (Decety & Ickes, 2011). Letzteres ist entscheidend, um es von Konstrukten wie Gefühlsansteckung und Mitleid abgrenzen zu können. Um empathische Prozesse entstehen zu lassen, braucht es in sensu oder vivo andere Personen, deren Lebenswirklichkeit mental erschlossen werden kann (Preston & Waal, 2002). Somit ist Empathie ein komplexer Prozess, der abhängig von anderen Menschen ist. Als grundlegendes Modell kann Empathie als drei-schrittiger Kommunikationsprozess verstanden werden (Kunyk & Olson, 2001). Hierbei werden Aspekte der anderen Person wahrgenommen, die erhaltenen Informationen verbal oder non-verbal zurückgemeldet, das Gegenüber nimmt diese Information auf und kann dann eine Rückmeldung geben, ob das Verstandene stimmt (Barrett-Lennard, 1981). Für diese Verstehensprozesse bieten Imitationsprozesse eine Grundlage, um Emotionen zu dekodieren und zu verstehen und diese mit weiteren Informationen über die Situation des anderen in einen Kontext zu betten. Gleichzeitig sorgen diese Imitationsprozesse auch für ein eigenes Mitschwingen (vermutlich auch, um die Emotionen des Gegenübers zu begreifen), so dass in gewissem Maße eine

abgeschwächte Gefühlsansteckung stattfindet, da im Verarbeitungsprozess bewusst bleibt, dass das Gefühlsauftreten unmittelbar dem Erleben der anderen Person geschuldet ist. Hier spricht man von *empathischer Akkuratheit* und *empathischem Mitschwingen* (Decety & Ickes, 2011; Rogers, 1957). Im weiteren Verlauf kann dieses Verstehen zu prosozialem Verhalten führen, um die andere Person in ihrer Situation zu unterstützen. In Anlehnung an das Empathie-Prozessmodell von Altmann (2015) wird in dieser Arbeit ein ähnliches Empathiemodell zugrunde gelegt. Im ersten Schritt geht es um das Wahrnehmen der emotionalen Situation des Gegenübers. Hierbei fließen verschiedene Informationen mit ein, die durch Aufmerksamkeitsprozesse wahrgenommen werden, um im nächsten Schritt eine mentale Repräsentation der Erlebenswelt des Gegenübers zu generieren. Diese Repräsentation wird mit den Informationen und eigenen Erfahrungen abgeglichen und löst das emotionale Mitschwingen aus. Letzteres wird als mittelfristig stabiler Persönlichkeitsaspekt oder *Fähigkeit* von Empathie angesehen, alle anderen Aspekte sind potentiell trainierbar und können als *Fertigkeit* betrachtet werden (Altmann, 2015). Bisherige Verfahren zur Messung von Empathie stützen sich weitestgehend auf Selbstauskunft und versuchen je nach Modell, kognitive oder affektive Anteile zu erfassen (vgl. Kapitel 2.1.3). Dieses Vorgehen bringt neben den üblichen Problematiken der Selbstauskunft (bspw. soziale Erwünschtheit; Schmidt-Atzert & Amelang, 2012) auch die Einschränkung mit sich, dass den Daten eine objektive Komponente fehlt. In der Arbeit mit Kindern wird seit jeher versucht im Bereich der Empathieforschung mit Geschichten und Bildern zu arbeiten. Bei der Picture-Story-Method (Feshbach & Roe, 1968) beispielsweise geben die Kinder an wie sie sich selbst als Protagonist einer erzählten Geschichte fühlen würden bzw. wie es der Person aus der Geschichte gehen würde. Da die Aufgaben allerdings auf Geschichten basierten, wurde kritisch hinterfragt, ob derartige fiktive Situationen geeignet seien, um Empathie auszulösen (Eisenberg & Miller, 1987). In der Forschung mit Erwachsenen verfolgte man den Ansatz *echte* Filmsequenzen zu

produzieren, die vermutlich anders wirken als durch SchauspielerInnen gestellte Situationen und die zur Messung von Empathie nach bestimmten Gesichtspunkten bewertet werden sollen (Richter, 2009). Häufig werden auch Interaktionen aufgenommen, zu denen die Protagonisten dann nachträglich zu ihren Gedanken und Gefühlen befragt werden (Decety & Ickes, 2011). An diesen Antworten können dann Einschätzungen anderer Personen validiert und auf Richtigkeit geprüft werden, um damit vor allem Fähigkeiten zu empathischer Akkuratheit einschätzen zu können. Eine Standardisierung bei diesem Vorgehen kann erreicht werden, indem die aufgenommenen Interaktionen von geschulten Ratern anhand unterschiedlicher Konstrukte eingeschätzt werden (Richter, 2009). Die videobasierten Messverfahren sind allerdings häufig situations- sowie themenspezifisch, so dass es bislang noch kein Instrument gibt, das sich situations- bzw. zielgruppenunabhängig einsetzen lässt. Sieht man Empathie als Kompetenz an, verwundert das wenig, da etwaige Anteile inhaltsspezifisch ausgebildet werden (Klieme & Hartig, 2008). Betrachtet man Lehrkräfte benötigen breit gefächerte Kompetenzen um ihren Beruf erfolgreich ausüben und die von der Kultusministerkonferenz geforderten Standards umsetzen zu können (Kultusministerkonferenz, 2014). Neben fachlichem Wissen der unterrichteten Fächer gehört auch pädagogisch-psychologisches Wissen dazu, das in breitem Maße spezifisches Wissen über Unterrichtsgeschehen beinhaltet (Seidel et al., 2010; Voss et al., 2011). Diese Fähigkeiten werden häufig über Videoaufnahmen vermittelt und auch gemessen (Sabers et al., 1991; Seidel et al., 2010; Stürmer et al., 2013), mit denen man in der Lage ist, in komplexen Situationen Handlungswissen und -kompetenzen zu erfassen. Für LehrerInnen hat Empathie einen hilfreichen Einfluss auf das Unterrichtsgeschehen, da es unterstützend für die professionelle Wahrnehmung ist, um Informationen über den aktuellen Lernstand der Schülerschaft zu bekommen und ggf. Maßnahmen zur Veränderung zu ergreifen (Drummond, 1993). Zusätzlich kann sich empathisches Verhalten der Lehrkraft positiv auf das Unterrichtsklima und die Leistungen der

SchülerInnen auswirken (Aspy, 1969; Tausch & Tausch, 1998). Insofern handelt es sich bei Empathie um ein hochgradig relevantes Konstrukt bei Lehrkräften.

Ein erstes Ziel der vorliegenden Arbeit ist es, ein möglichst realitätsnahes Instrument auf Basis von Videosequenzen zu entwickeln, um durch die dort präsentierten sozialen Interaktionen empathische Fähigkeiten bei angehenden Lehrkräften objektiver erfassen zu können als mit klassischen Selbstauskunftsfragebögen. Grundlage sollen reale Unterrichtssituationen mit emotional gefärbten Inhalten sein, anhand derer sich empathische Reaktionen auslösen lassen, die es dann auf verschiedenen Dimensionen zu erfassen gilt.

SL ist wie bereits erwähnt ein Lehr-Lernformat, das bereits seit den 1940er Jahren existiert und das in den USA entstand. Als wichtigen Bestandteil beinhaltet es ein ehrenamtliches Engagement, das einen realen Bedarf in der Gesellschaft adressieren soll (Reinders, 2010). Durch dieses Engagement am Menschen wird prosoziales Handeln angewandt und dadurch die Interaktionen mit anderen geschult. Durch die Verzahnung mit einer begleitenden Lehrveranstaltung an der Universität wird die Möglichkeit zum Aufbau theoretisch, fachlichen Wissens und zur Reflexion und Einordnung der gemachten Erfahrungen gegeben (Eyler & Giles, 1999). Im Vordergrund dieser Lehrkonzepte steht, neue Inhalte mit bereits bestehenden Wissenskonzepten und mentalen Schemata zu verknüpfen, um so neues Wissen entstehen zu lassen (Mayer, 2004; Renkl, 2009). Grundannahme ist hierbei die Theorie des Konstruktivismus, der davon ausgeht, dass die Konstruktion von Wissen maßgeblich durch die eigene Wirklichkeit geprägt ist und stark beeinflusst wird von zuvor gemachten Erfahrungen oder bereits entstandenen kognitiven Schemata (Young & Collin, 2004). Im sozialen Konstruktivismus steht in diesem Zusammenhang auch besonders der Einfluss sozialer Interaktionen im Vordergrund. Damit SL wirken kann, wurden im Verlauf der letzten 20 Jahre Qualitätskriterien erarbeitet, die ein optimales Lernen in diesen Veranstaltungsformen ermöglichen sollen (Billig & Weah, 2008). Hierbei stehen vor allem

Aspekte wie das Arbeiten auf Augenhöhe und ausreichende Zeit zur Bewältigung der möglichst persönlich relevanten und als wichtig empfundenen Aufgaben im Vordergrund. Besonders kommt auch der in der Begleitveranstaltung stattfindenden Reflexion eine wichtige Rolle zu, um Lernen überhaupt möglich zu machen (Eyler & Giles, 1999; Reinders, 2016). Da es verschiedene Aspekte gibt, die in SL-Veranstaltungen zum Tragen kommen, sind die Forschungsergebnisse vielfältig und nicht immer einheitlich (Celio et al., 2011; Conway et al., 2009; Novak et al., 2007; Yorio & Ye, 2012), so dass eine Vergleichbarkeit schwierig und Effektstärken heterogen sind. Wenn man in der bestehenden SL-Forschung einen Blick auf das Konstrukt *Empathie* wirft, fällt auf, dass es häufig implizit als Teil von moralischer Entwicklung oder prosozialem Verhalten miterhoben wird (Batchelder & Root, 1994; Brown, 2011; Ferrari & Worrall, 2000; Govekar & Rishi, 2007; Hey et al., 2014; Jones & Hill, 2001; Jones & Abes, 2004; Potthoff et al., 2000; Prentice & Robinson, 2010; Soukup, 1999). Da prosoziales Verhalten in den meisten Definitionen Empathie nachgelagert und kein Bestandteil davon ist, fehlt hierbei die definatorische Trennschärfe. Einzig zwei Studien beschäftigen sich explizit mit Empathie in einem sozialwissenschaftlichen Kontext und konnten einen Effekt finden (Lundy, 2007; Wilson, 2011).

Aufgrund der Tatsache, dass sich die Forschung bislang nur selten in diesem Zusammenhang mit Empathie beschäftigt hat und das auch nur mit einer anderen Zielgruppe sowie im angloamerikanischen Raum, der wenig repräsentativ für Lehramtsstudierende in Deutschland ist, wird in dieser Arbeit ein exploratorisches Vorgehen gewählt. Es wird die forschungsleitende Annahme zugrunde gelegt, dass es einen Unterschied in Empathie-relevanten Facetten gibt, wenn man Lehramts-Studierende in SL-Veranstaltungen mit Studierenden vergleicht, die keine SL-Veranstaltung besuchen. Dieser Unterschied dürfte dahin gehend bestehen, dass SL-Studierende von der Veranstaltungsform diesbezüglich mehr profitieren dürften und höhere Werte in Empathie-relevanten Aspekten aufweisen. Zur

Messung wird eine Testbatterie eingesetzt, in deren Rahmen auch das neu entwickelte Empathie-Testverfahren im Praxiseinsatz überprüft wird.

Kapitel 3: Testentwicklungsschritte

Da die bisherigen gängigen Testverfahren nicht umfassend und objektiv Empathie erfassen und für Lehramtsstudierende ein situations-spezifisches Verfahren entstehen soll, wurde für die vorliegende Studie ein eigenes Testverfahren entwickelt. Das neue Verfahren soll auf objektive Weise verschiedene Facetten von Empathie abbilden. Neben der Selbstauskunft zu eigenem Gefühlserleben und dem Einschätzen der Emotionen anderer, werden auch Gesichtsaufnahmen gemacht, die mit einer Face-Reader Software analysiert werden und den wichtigen Aspekt des *Mitschwingens* messen sollen. Bei der Entwicklung des Verfahrens wurde sich an der Struktur und Entwicklung des *Observers* orientiert (Seidel et al., 2010). Der Observer ist ein videobasiertes Verfahren, das für Lehramtsstudierende entwickelt wurde, um professionelle Wahrnehmung standardisiert und systematisch abbilden zu können. Zum Einsatz kommen verschiedene Videosequenzen mit Unterrichtssituationen, die verschiedene Unterrichtskomponenten zum Inhalt haben und von den Studierenden eingeschätzt werden sollen. Aufbauend auf diesem Entwicklungskonzept wurden für den vorliegenden Vignettentest auch videobasierte Unterrichtssequenzen gesucht. Zur Sichtung passender Clips wurde Stimulusmaterial aus der Paedocs-Datenbank (mittlerweile *Forschungsdatenzentrum Bildung*) des *Deutschen Instituts für Pädagogische Forschung – DIPF* genutzt, die für das Forschungsvorhaben nach Antragsstellung zur Verfügung gestellt wurden. Alle gesichteten Videoaufnahmen stammten aus den Studien „Pythagoras: Unterrichtsqualität, Lernverhalten und mathematisches Verständnis in verschiedenen Unterrichtskulturen“ (Lipowsky & Klieme, 2006), „ Audiovisuelle Aufzeichnungen von Schulunterricht in der Bundesrepublik Deutschland“ (Schluß & Jehle, 2013), „VERA - Gute Unterrichtspraxis“ (Helmke, 2008), sowie „Nicht-intendierte Effekte Neuer Steuerung im Schulsystem - Theoretische Konzeptualisierung und Instrumenten-Entwicklung zur empirischen Erfassung (NeFo)“ (Bellmann, 2014). Insgesamt wurden 23 jeweils ungefähr 2-stündige Unterrichtsvideos gesichtet und 29 Videosequenzen zwischen 00:30 Minuten und 1:20 Minuten ausgewählt, die

eine affektiv gefärbte, soziale Interaktion zum Inhalt hatten. Die ausgewählten Videosequenzen stammten final aus den Studien „Pythagoras: Unterrichtsqualität, Lernverhalten und mathematisches Verständnis in verschiedenen Unterrichtskulturen“ und „VERA - Gute Unterrichtspraxis“ und wurden aus den originären Unterrichtsaufnahmen mit dem Videoschnittprogramm Power Director Ultimate v12 (Cyber Link) extrahiert. Zur weiteren Entwicklung der finalen Testversionen folgten weitere Sondierungs- und Anpassungsschleifen, die mit Hilfe von Expertenbefragungen realisiert wurden. Die Entwicklungsschritte werden im Folgenden erläutert.

3.1 Erster Testentwicklungsschritt: Befragung der Empathie-ExpertInnen

Um die Güte der ausgewählten Videosequenzen festzustellen und die Anzahl auf diejenigen, die im Sinne des Testes am besten funktionieren, reduzieren zu können, wurde eine ExpertInnenbefragung durchgeführt. Das Ziel war, konkrete Informationen über die Vignetten bzw. die zu sehenden Emotionen zu sammeln und die Anzahl der Vignetten zu reduzieren. Dafür sollten die in den Videosequenzen gezeigten Emotionen eingeschätzt und auf ihr Potential Empathie auszulösen bewertet werden. Die Befragungen wurden online über Unipark von Questback (<https://www.unipark.com/>) umgesetzt.

3.1.1 Interne Pilotierung

Bevor die Befragung mit den ExpertInnen stattfinden konnte, gab es eine interne Pilotierungsrunde, bei der sechs studentischen Hilfskräften alle 29 Vignetten mit den geplanten Fragen vorgelegt wurden. Der Arbeitsauftrag lautete, alle Fragen zu beantworten und eine Rückmeldung über die Dauer und eventuelle aufgetretene technische Problematiken zu geben. Alle teilnehmenden Personen brauchten mehr als 60 Minuten zur Beantwortung, so

dass die Vignetten für die eigentliche Befragung aufgeteilt und auf drei Testversionen mit jeweils 16 Videosequenzen verteilt wurden, um vor allem die Bearbeitungszeit zu reduzieren und so eine höhere Partizipation zu erreichen. Außerdem wurden die Fragen angepasst, da teilweise fehlinterpretiert wurde auf was geachtet werden sollte. Die Antwortmöglichkeiten wurden auch noch einmal aufgrund der heterogenen Antworten und der damit verbundenen Auswertungsschwierigkeiten überarbeitet. Bspw. wurde die teilweise offene Frage „Konnten Sie bei dem Mädchen Emotionen erkennen? Ja/Nein, wenn ja: Welche Emotionen waren das?“ in eine Rating-Skala konvertiert („In welchem Maße konnten Sie hauptsächlich die folgenden Emotionen bei dem Mädchen beobachten? Freude/ Ärger/ Angst/ Traurigkeit/ Überraschung/ Ekel; wie würden Sie die Emotion nennen?“ mit 1=*gar nicht* bis 6=*sehr stark*). Zusätzlich wurden Beschreibungen der Videoclips hinzugefügt, ebenso wie die Frage nach einem empathischen Verhalten in dieser Situation, wenn die zuschauende Person im Klassenzimmer gewesen wäre. Die Vignetten, die die studentischen Hilfskräfte in der Mehrzahl als potentiell Empathie auslösen zu können bewerteten, wurden in der Expertenbefragung in allen erstellten Testversionen gezeigt. Technische Probleme traten bei der internen Pilotierungsrunde nicht auf.

3.1.2 Stichprobe

Um geeignete ExpertInnen zu identifizieren wurde eine Internetsuche nach anerkannten, d. h. bei den jeweiligen Dachverbänden registrierten, *TrainerInnen emotionaler Kompetenzen* nach Berking (TEK; <http://www.tekonline.info/>), *TrainerInnen für gewaltfreie Kommunikation* (GFK, <https://www.fachverband-gfk.org/>), sowie PsychologInnen, die sich mit Empathie (-messung) und/oder Emotionsforschung im deutschsprachigen Raum beschäftigen, durchgeführt. Das TEK-Training ist ein auf aktuellen Forschungsbefunden basierendes emotionales Training, das trainieren soll, mit belastenden Gefühlen nutzbringend umzugehen

(Berking, 2015). Die GFK ist ein Kommunikations- und Handlungskonzept nach Rosenberg (2016), das auf der klientenzentrierten Gesprächspsychotherapie nach Rogers (1957) aufbaut. Hierbei geht es darum, die eigenen und die Bedürfnisse der Anderen zu verstehen und sich einzufühlen, um auf dieser Basis zu konstruktiven Lösungen zu kommen. Das Konzept der *Empathie* (vor allem im Sinne der *empathischen Akkuratheit*) ist ein Kernmerkmal dieses Konzeptes.

Insgesamt konnten 75 Personen identifiziert werden, die im Juli 2016 per Email angeschrieben wurden (siehe Anhang A2). Im August 2016 wurden weitere PsychologInnen angefragt mit der Bitte um Streuung der Email-Anfrage. Zusätzlich gab es eine Erinnerungsmail an 74 Personen mit der erneuten Bitte um Teilnahme. Final wurden 86 Personen angeschrieben und eingeladen, die laut der Recherche verschiedenen Berufsgruppen angehörten. Die ExpertInnen konnten sich zurückmelden, wenn sie über Ergebnisse informiert werden wollten. Im September 2016 gab es eine zusätzliche Befragungsrunde bestehend aus weiteren fünf PsychologInnen, denen nur die Testversion 1 vorgelegt wurde, da einige Vignetten bis zu diesem Zeitpunkt nur von zwei ExpertInnen bearbeitet und deswegen für eine sinnvolle Auswertung weitere Informationen benötigt wurden. Final konnten die Daten von 27 Personen genutzt werden. Die Expertise und ausgeübten Berufe der teilnehmenden ExpertInnen setzten sich unterschiedlich zusammen (siehe Abbildung 3; Mehrfachnennungen waren möglich).

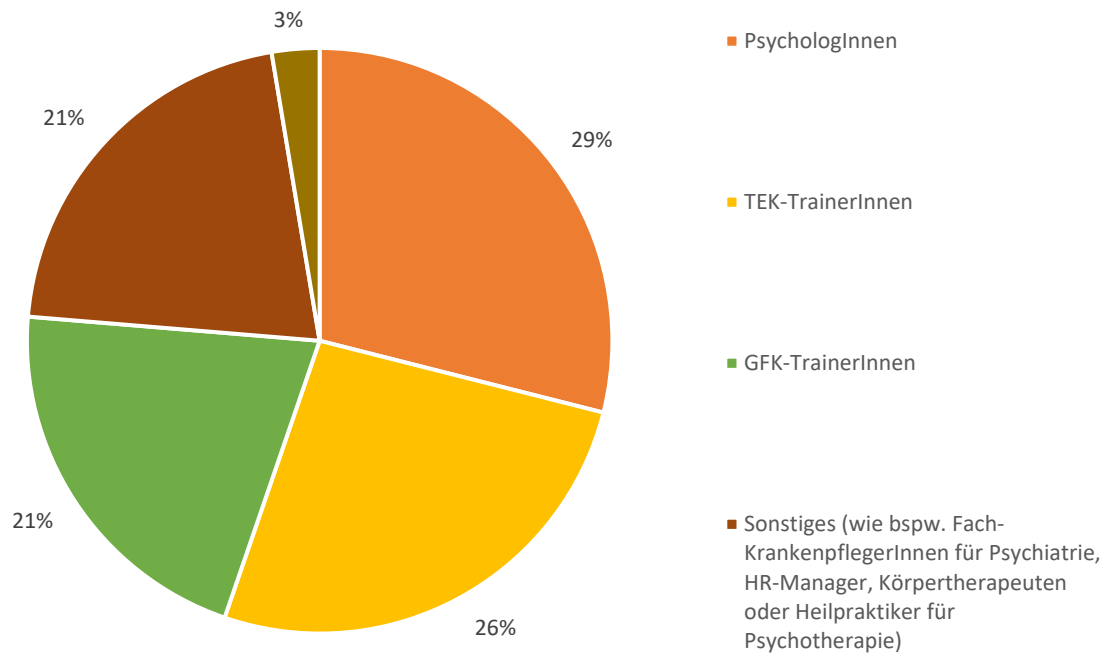


Abbildung 3. Darstellung der Berufsgruppen der teilnehmenden EmpathieexpertInnen; Bei der Angabe waren Mehrfachnennungen möglich, bspw. konnten TEK-TrainerInnen auch PsychologInnen sein u. ä.; insgesamt gab es $N=38$ Nennungen.

Unter den Teilnehmenden waren 19 Frauen und fünf Männer mit einem Durchschnittsalter von 39.58 Jahren ($SD=11.31$ Jahre). Die demografischen Angaben von drei Personen fehlen, da sie in den ersten beiden Wochen des Befragungszeitraums teilnahmen und den Test vorzeitig abbrachen, so dass die Daten nicht angegeben wurden. Danach wurde die Seite dem Vignettentest vorangestellt. Die Daten wurden allerdings trotz fehlender Demografie genutzt, da durch die Anschreiben sichergestellt war, dass es sich um die gewünschten ExpertInnen handelte und die Informationen damit nutzbar waren.

3.1.3 Aufbau

Die insgesamt 29 Vignetten wurden auf drei Testversionen mit jeweils 16 Vignetten aufgeteilt. Die Auswertung erfolgte auf Vignettenebene, so dass auch nicht vollständig bearbeitete Testversionen berücksichtigt werden konnten. Um möglichst viele Informationen pro Vignette

zu erhalten kamen in den verschiedenen Versionen teilweise die gleichen Vignetten vor (siehe Tabelle 11).

Tabelle 11
Aufbau Testversionen der Empathie-ExpertInnen

Anzahl Items	Testversion 1	Testversion 2	Testversion 3
1	Vignette 4	Vignette 4	Vignette 4
2	Vignette 9	Vignette 9	Vignette 9
3	Vignette 10	Vignette 10	Vignette 10
4	Vignette 21	Vignette 21	Vignette 21
5	Vignette 24	Vignette 24	Vignette 24
6	Vignette 23	Vignette 23	Vignette 23
7	Vignette 2	Vignette 12	Vignette 1
8	Vignette 5	Vignette 13	Vignette 6
9	Vignette 7	Vignette 14	Vignette 8
10	Vignette 11	Vignette 15	Vignette 16
11	Vignette 15	Vignette 18	Vignette 17
12	Vignette 16	Vignette 20	Vignette 19
13	Vignette 19	Vignette 25	Vignette 22
14	Vignette 20	Vignette 27	Vignette 26
15	Vignette 25	Vignette 28	Vignette 27
16	Vignette 30	Vignette 29	Vignette 29

Anmerkung. Die *Anzahl Items* 1 – 6 beinhalten Vignetten, die in allen Testversionen vorkamen, die restlichen Vignetten kamen teilweise in den anderen Testversionen vor.

Die Vignetten wurden randomisiert gezeigt, um den Einfluss von Ermüdungseffekten zu minimieren. Nach einer Begrüßung und der Instruktion generierten die ProbandInnen einen Code, der dazu diente bei technischen Problematiken die verschiedenen Testversionen kombinieren zu können. So konnten die Datensätze problemlos zusammengefügt bzw. gelöscht werden. Als nächstes folgte ein Hinweis auf die Tonqualität, die aufgrund der Aufnahmequalität der groß angelegten Unterrichtsstudien sehr variierte und trotz akustischer Normalisierung im Vorfeld immer noch schwankten. Zudem wurde aufgrund der Rückmeldungen aus der internen Pilotierung (siehe Kapitel 3.1.1) darauf hingewiesen, dass Schilder um den Hals der SchülerInnen keine Bewertungskomponente hatten, sondern nur der

Zuordnung im Forschungskontext der Ursprungsstudien dienten. Im nächsten Schritt wurde darum gebeten, eine kurze Definition des eigenen Verständnisses von Empathie zu geben, damit bei der Auswertung sichergestellt werden konnte, dass dieses deckungsgleich mit dem Verständnis in dieser Studie ist. Danach folgte der Vignettentest. Für jede Vignette gab es eine kurze Erläuterung des Inhalts, danach konnte sie so häufig wie gewünscht angeschaut werden. Danach wurden acht Fragen gestellt, die der Einschätzung dienten, ob im jeweiligen Videoclip Emotionen zu sehen waren und Empathie beim Anschauen ausgelöst wurde (siehe Anhang A3).

Bei der Frage nach der Intensität der gesehenen Emotionen wurde sich an den Basisemotionen nach Ekman (1992) orientiert. Frage nach empathischem Verhalten in der Situation wurde offen und breiter formuliert (empathisches Verhalten in dieser Situation bei Anwesenheit im Klassenzimmer), um möglichst diverse Informationen/Antworten zu erhalten, aus denen in einem nächsten Schritt geeignete Antwortalternativen zu professionellem, empathischen LehrerInnenverhalten gebildet werden sollten. Das führte zu einigen Rückmeldungen, dass unklar blieb, als welche Person bzw. in welcher Rolle man eine Antwort hätte geben sollen.

3.1.4 Datenbereinigung

Der Datensatz wurde mit IBM SPSS Statistics Version 24 (*SPSS Statistics*) ausgewertet. Zur Bereinigung wurden Missings definiert, die standardmäßig in SPSS vorgegeben wurden (-99 bei fehlenden Werten, -77 bei quantitativen bzw. -66 bei Zeichenvariablen, wenn Items gar nicht präsentiert wurden). Zusätzlich wurden die Daten in einer Excel-Datei chronologisch aufbereitet um die Freitextantworten übersichtlich auswerten zu können (Excel 2013). Teilgenommen und irgendeine Eingabe gemacht (z.B. den Code) haben 36 Personen, keine Vignette bearbeitet haben davon neun Personen und drei Personen haben nur eine einzige Vignette beantwortet, so dass es verwertbare Daten von insgesamt 27 Personen gibt. Bei der

Rating-Skala zur Einordnung der Emotion anhand der beobachteten Intensität wurden Datensätze mit 0en auf 1 rekodiert, sofern zumindest eine der sechs Emotionen eingeschätzt wurde (mit 1=*gar nicht vorhanden*). Da Unipark nicht vorhandene Intensitäten aufgrund technischer Gegebenheiten mit 1 kodierte, wurde der Datensatz entsprechend dieser Kodierung weitergeführt.

3.1.5 Ergebnisse

3.1.5.1 Vignettenauswahl

Als Hauptindikator zur Bewertung der Güte einer Vignette, wurde die Auswertung der Häufigkeiten zur Frage: „Würden Sie sagen, dass diese Videosequenz dazu geeignet ist, empathische Empfindungen beim Zuschauenden auszulösen?“ herangezogen. Bei einer häufigeren Ja-Angabe (>51%) schienen die Vignetten geeignet und wurden zur weiteren Nutzung ausgewählt. (siehe Tabelle 12; Übersicht aller Vignetten: siehe Anhang A4).

Tabelle 12

Von den ExpertInnen eingeschätzte Empathie-auslösende Vignetten

Empathieauslösend?	N	Ja-Antworten		Nein-Antworten	
		Anzahl	Prozentualer Anteil	Anzahl	Prozentualer Anteil
Vignette 6	8	6	75.0 %	2	25.0 %
Vignette 8	9	5	62.5 %	3	37.5 %
Vignette 9	18	11	61.1 %	7	38.9 %
Vignette 10	16	11	68.6 %	5	31.3 %
Vignette 14	6	4	66.7 %	2	33.3 %
Vignette 21	18	16	88.9 %	2	11.1 %
Vignette 25	13	9	69.2 %	4	30.8 %
Vignette 27	15	8	53.3 %	7	46.7 %

Nach der Erstauswahl der Vignetten wurden noch weitere Vignetten ausgewählt, um sie für weitere Entwicklungsschritte mit einzubeziehen. Als Auswahlkriterien dienten hierbei noch die Mittelwerte der Intensitätsantworten, die Freitextantworten der Gefühlsqualität und die

übereinstimmenden Nennungen der ExpertInnen ebenfalls bei der Angabe zur Gefühlsqualität (siehe Tabelle 13).

Tabelle 13

Auswahlkriterien für nicht Empathie-auslösende Vignetten zur Verwendung bei der weiteren Testentwicklung, Befragung Empathie-ExpertInnen

	Anzahl Antworten <i>N</i>	Empathie- auslösend		Gefühlsqualitäten	Gefühlsintensität <i>MW</i>		Anmerkungen
		Ja (<i>n</i>)	Nein (<i>n</i>)		Min.- Angabe	Max.- Angabe	
Vignette 2	7	3	4	Divers, aber Mehrfachnennungen	1	4	
Vignette 11	6	1	5	Kompakt mit Mehrfachnennung	1	4	
Vignette 19	13	3	10	Sehr divers, mit vielen Mehrfachnennungen	1	4	Freude als Mehrfachnennung
Vignette 24	18	4	14	Sehr divers, mit Mehrfachnennungen und vielen Einzelnennungen	1	4	Freude als häufigste Nennung, aber sehr divers
Vignette 28	7	2	5	Sehr divers, mit Mehrfachnennungen	1	4	Frust und Ärger als häufigste Nennung
Vignette 30	6	1	5	Divers mit einer Mehrfachnennung	1	3	

Anmerkung. Gefühlsintensität: 1=Gar nicht, 3=schwach, 4=mittelmäßig.

Es wurde auch Vignetten weiterhin im Pool belassen, die weniger als 25.0 % (Vignette 11, 19 und 24) als Empathie-auslösend eingeschätzt wurden, wenn in einer eigentlich nicht Empathie-auslösenden Vignette relativ hohe gemittelte Gefühlsintensitäten (ab 4=*mittelmäßig*) eingeschätzt wurden (Ausnahme: Vignette 30). Das betraf allerdings nicht die Vignetten, bei denen in den Kommentaren rückgemeldet wurde, dass das Gesicht der entsprechenden SchülerInnen schlecht zu sehen gewesen sei. In diesen Fällen wurde auf den Einsatz verzichtet. Zudem sprach für den weiteren Einsatz der Vignette, dass es darüber hinaus gleiche Mehrfachnennungen der gesehenen Gefühlsqualität gab (Auswertung aller Vignetten siehe Anhang A5).

Neben der Vignettenreduktion war ein weiteres Ziel dieser Expertenbefragung Informationen zu sammeln, um ein Scoring zu entwickeln, das eine Einteilung und Bewertung in *richtige* und *falsche* Antworten der Vignetten möglich macht. Dafür wurden die Intensitäts-Angaben der sechs Basisemotionen (Freude, Ärger, Angst, Traurigkeit, Überraschung, Ekel) gemittelt und auf ganze Zahlen gerundet. Pro korrekt eingeschätzter Intensität gibt es zukünftig einen Punkt, für jede Einschätzung +/- eine Standardabweichung noch 0.5 Punkte (vgl. Anhang A6). Als Subskalenwert *Emotionsintensität* wird der Summenwert aus den vergebenen Punkten gebildet.

Die Antworten zu der Frage „Wie würden Sie die Emotion nennen?“ wurden nach Häufigkeiten ausgezählt und anhand der Wortbedeutungen zusammengefasst. Hierfür wurde eine Internetrecherche der Bedeutungen und Synonyme der Wörter durchgeführt (Duden) und überprüft, ob es sich um einen Gefühlsausdruck handelt oder etwas anderes bezeichnet. Als relevant wurden nur die Nennungen betrachtet, die mehr als einmal genannt wurden. Durch die Auswertung der Emotionsantworten entstand eine Liste von 97 potentiellen Emotionswörtern, die nochmals von der Autorin und einer weiteren Psychologin eingeordnet wurden, ob es sich hierbei unserer Einschätzung nach um eine Emotion handelte oder eher nicht (siehe Anhang A7). Die Interrater-Reliabilität nach Cohens Kappa war nicht zufriedenstellend (Kappa=.34, 95%, Konfidenzintervall). Da Kappa von -1 bis +1 variiert, sind Werte ab etwa 0.5 annehmbar (Wirtz & Caspar, 2002), was in diesem Fall nicht gegeben war. Aufgrund dieses Ergebnisses wurde die Liste für die weitere Verwendung eingekürzt. Bei Übereinstimmung wurden die Worte als Emotion eingestuft und flossen entsprechend in die Entwicklung des Testes mit ein (entweder als Antwort oder als Distraktor), alle anderen Wörter wurden gestrichen und ausgeschlossen.

Aufgrund der Freitext-Anmerkungen der ExpertInnen wurden die Beschreibungen einiger Videosequenzen überarbeitet, um suggestive Darstellungen zu vermeiden.

Die Antworten auf die Frage: „Stellen Sie sich vor Sie wären im Klassenzimmer gewesen. Wie würden Sie sich dem Mädchen/Jungen gegenüber empathisch verhalten?“, wurden genutzt, um Verhaltensoptionen zu generieren, die für die Lehrerrolle passten, um sie als Antwortoptionen in der nächsten Expertenbefragung mit LehrerInnen einfließen zu lassen (siehe Anhang A8). Hierbei wurde versucht, mehrfach gegebene Antworten zu finden und in einer Antwort zusammenzufassen, was bei allen Vignetten gelang.

3.1.5.2 Intra-Class-Correlation

Zur Berechnung und Einschätzung der generellen Beurteilerübereinstimmung wurde die Intra-Class-Correlation (ICC) berechnet. Es handelt sich um ein Verfahren, das zur Einschätzung der Interrater-Reliabilität genutzt werden kann und im Rahmen der allgemeinen Testtheorie Anwendung findet, um den Grad der Übereinstimmung zwischen verschiedenen Beobachtern wiederzugeben (Wirtz & Kutschmann, 2007). Diese Analyse beruht auf dem varianzanalytischen Modell und lässt sich angelehnt an die Produkt-Moment-Korrelation beurteilen. Bei der ICC wird der Abstand jedes einzelnen Wertes zum Mittelwert bewertet. Sie kommt zum Einsatz, da es die ICC ermöglicht, die Übereinstimmung mehrerer Rater über ein Merkmal hinweg einzuschätzen. Im Gegensatz zu anderen Maßen wie beispielsweise Cohens κ (ein Korrelationsmaß bei mehr als zwei Fällen) müssen hier nicht die Zusammenhänge zwischen Rating-Paaren einzeln berechnet werden, sondern die Einschätzung des Zusammenhangs kann mit jedem beliebigen Rater erfolgen.

Wie bei jedem anderen Verfahren gibt es auch hierbei einige Grundannahmen zu beachten. Zuerst einmal sollten die Daten Intervallskalenniveau aufweisen. Intervallskalenniveau kann angenommen werden, wenn der Abstand zwischen nebeneinanderliegenden Punkten immer derselbe ist (Wirtz & Caspar, 2002). Einige der Items der vorliegenden Befragung sind lediglich nominalskaliert (Ja/Nein), aber da dichotome Daten einen Sonderfall darstellen in

dem Sinne, dass es nur zwei Skalenpunkte gibt und durch diese einfache Distanz *Äquidistanz* angenommen werden kann, können sie im Rahmen des allgemeinen linearen Modells wie intervallskalierte Daten behandelt werden (Moosbrugger, 2002; Wirtz & Caspar, 2002). In der Regel gibt es in diesem Fall auch eine inhärente Rangordnung. Die Werte der ICC können zwischen -1 und +1 liegen. Die Interpretation der Werte orientiert sich an anderen Reliabilitätsmaßen, die zwischen 0 bis +1 liegen, so dass Negativ-Werte wie 0 interpretiert werden können. Alle numerischen Items ($N=9$) wurden in die Analyse einbezogen und z -standardisiert, um eine Vergleichbarkeit der unterschiedlichen Skalen der Items möglich zu machen und die Bedingung einer gleichen Metrik zu erfüllen.

Eine weitere Bedingung ist, dass es in jedem bewerteten Fall gleich viele Beurteiler geben muss. In der vorliegenden Befragung wird die Beurteilerübereinstimmung deswegen pro Vignette bewertet und als Einzelfall betrachtet. Da es damit pro Vignette konsistent für jeden Fall die gleichen Rater gab und es sich bei den teilnehmenden Personen um eine Stichprobe der Population *ExpertInnen für Empathie und emotionale Kompetenzen* handelt, wird die ICC als zweiseitig, zufällig durchgeführt.

Es wurden die Vignetten angeschaut, die aufgrund der Antworten auf die Frage „Empathieauslösend?“ bereits ausgewählt und in die ExpertInnenbefragung für LehrerInnen eingeflossen sind. Die Reliabilitäten lagen bei allen Vignetten zwischen .65 und .96 ($p < .05$; vgl. Anhang A9). Bei nahezu allen Vignetten gibt es hohe bis sehr hohe Übereinstimmungen zwischen den Beurteilern, so dass die Vignetten für die weitere ExpertInnenbefragung mit LehrerInnen genutzt werden können.

3.2 Zweiter Testentwicklungsschritt: Expertenbefragung mit LehrerInnen

Der Vignettentest soll neben der Fähigkeit Emotionen akkurat zu erkennen und einzuschätzen auch professionelles Lehrerverhalten abbilden. Deswegen wurde in einem nächsten Schritt eine weitere Befragung mit ExpertInnen (in diesem Fall LehrerInnen) durchgeführt. Ein primäres Ziel dieser ExpertInnenbefragung sollte es sein, die Antwortmöglichkeiten der Frage nach einem professionellen und empathischem Lehrkraft-Verhalten in den gezeigten Situationen einzuschätzen und in eine Rangreihenfolge zu bringen bzw. die Antwortoptionen um eine bessere Alternative zu ergänzen. Zudem wurde überprüft ob und wie die Fragen zur Emotion und Intensitätseinschätzung funktionierten. Auch diese Befragung wurde wieder über Unipark umgesetzt und auch hier alle in Frage kommenden ExpertInnen per Email initiativ angeschrieben (siehe Anhang B1).

3.2.1 Stichprobe

Im November 2016 wurden 25 LehrerInnen und Bekannte angeschrieben mit der Bitte, die Nachricht an deutschsprachige LehrerInnen ab dem Vorbereitungsdienst/Referendariat weiterzuleiten oder auch selbst teilzunehmen. Die Ansprache erfolgte über Email, WhatsApp und Facebook. Im Januar 2017 wurden zusätzlich 50 Schulen im Rhein-Main-Gebiet mit der Bitte die Anfrage im Kollegium weiterzuleiten, initiativ angeschrieben. Ausgenommen wurden Grundschulen, da die meisten Vignetten Jugendliche SchülerInnen zeigen und die LehrerInnen Erfahrungen mit der entsprechenden Zielgruppe haben sollten. Insgesamt haben 75 Personen auf die Umfrage zugegriffen und 37 Personen Daten hinterlassen. Davon konnten 36 Datensätze genutzt werden. Die Daten waren von 19 Frauen und 17 Männern mit einem Altersdurchschnitt von 49.08 Jahre ($SD=12.31$). Der jüngste Teilnehmende war 29 Jahre alt, der älteste Teilnehmende 73 Jahre. Es gab 31 LehrerInnen, die zum Zeitpunkt der Befragung

noch aktiv im Schuldienst waren und vier LehrerInnen, die bereits im Ruhestand waren. Die Lehrerfahrung der ExpertInnen variierte von 1-2 Jahren bis zu über 15 Jahren Erfahrung. Der Großteil der Teilnehmenden lehrte bereits seit über 15 Jahren im Schuldienst (siehe Anhang B2). Da die meisten Anschreiben an Schulen im Rhein-Main-Gebiet stattfanden, unterrichteten die meisten Teilnehmenden an hessischen Schulen. Die unterrichteten Fächer der teilnehmenden LehrerInnen befanden sich vor allem im sprachlich-literarisch-künstlerischen Aufgabenfeld. Jede LehrerIn gab an, mehrere Fächer zu unterrichten (siehe Abbildung 4).

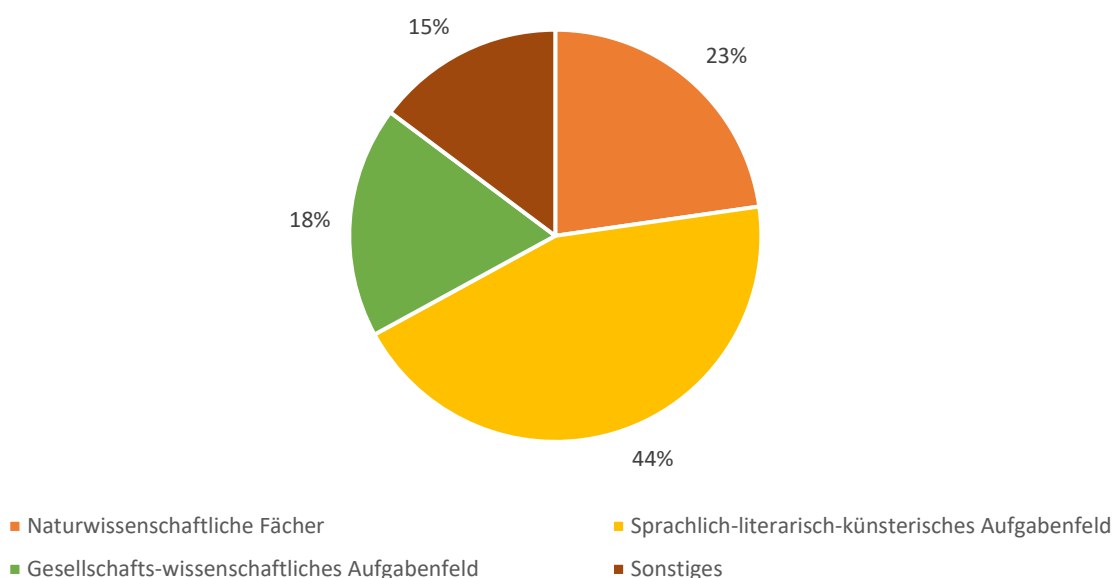


Abbildung 4. Unterrichtete Fächer der LehrerInnen; Alle LehrerInnen gaben mehrere Fächer an; Naturwissenschaftliche Fächer=Mathe, Physik, Informatik/EDV, Chemie, Biologie; Sprachlich-literarisch-künstlerisches Aufgabenfeld=Deutsch, Englisch, Französisch, Latein, Kunst, Musik, Darstellendes Spiel/Tanz, kreatives Gestalten; Gesellschaftswissenschaftliches Aufgabenfeld=Geographie/Erdkunde, Kath. Religion, Geschichte, Philosophie/Ethik, ev. Religion, Soziales Lernen, Gesellschaftslehre/Politik, Kommunikation; Sonstiges=Sport, Wirtschaft, Bürowirtschaft, Rechnungswesen, Arbeitslehre.

3.2.3 Testaufbau

Für diese ExpertInnenbefragung wurden die in der vorherigen Befragung ausgewählten vierzehn Vignetten genutzt, die auf zwei Testversionen mit jeweils neun Vignetten verteilt wurden. Vier Items waren in beiden Testversionen vorhanden (siehe Tabelle 14).

Tabelle 14

Aufbau Testversionen der LehrerInnen-Befragung

Anzahl der Items	Testversion 1	Testversion 2
1	Vignette 2	Vignette 2
2	Vignette 6	Vignette 6
3	Vignette 21	Vignette 21
4	Vignette 25	Vignette 25
5	Vignette 8	Vignette 10
6	Vignette 9	Vignette 24
7	Vignette 11	Vignette 27
8	Vignette 14	Vignette 28
9	Vignette 19	Vignette 30

Anmerkung. Die Anzahl der Items 1 – 4 beinhalten Vignetten, die in allen Testversionen vorkamen, die restlichen Vignetten waren teilweise in beiden Versionen vorhanden.

Die gesammelten Antworten wurden in einem finalen Datensatz zusammengefasst und nicht getrennt nach Testversionen ausgewertet (außer für die Trennschärfebestimmung).

Die Items zu jeder Vignette wurden angepasst, um die Ergebnisse der vorherigen ExpertInnenbefragung zu berücksichtigen. Insgesamt gab es sieben Fragen pro Vignette, wovon es für vier Fragen Ratingskalen und vorgegebene Antwortmöglichkeiten gab (siehe Anhang B3).

Neben diesen Fragen wurden vorneweg demografische Daten erhoben und eine Definition von Empathie gegeben, die dem Verständnis dieser Arbeit zugrunde liegt. Die Definition sollte dazu dienen zu verdeutlichen was mit Empathie gemeint ist, um eine einheitliche Verständnisgrundlage zu schaffen. Die Teilnehmenden wurden, erhoben über eine 4-stufige Likert Skala von *gar nicht wichtig* (1) bis *Sehr wichtig* (4), daraufhin dazu befragt, wie wichtig empathisches Verhalten SchülerInnen gegenüber sei und ob solches Verhalten als

professionell einzuschätzen sei (dichotome Antwortmöglichkeiten: *Ja/Nein*). Dreizehn Mal wurden Vignetten der Testversion 1 bearbeitet und 23 Mal die Vignetten der Testversion 2.

3.2.3 Datenaufbereitung

Die Daten wurden mit Hilfe von Excel 2013 (*Office 2013*) und IBM SPSS 23 bzw. 24 (*SPSS Statistics*) bearbeitet. Fehlende Werte (Missings) wurden wieder mit -77, -66 oder -99 kodiert. Bei der Frage nach der Intensität der beobachteten Emotionen anhand der sechs Basisemotionen wurden (sofern überhaupt eine Emotionsintensität angegeben wurde) ausgegebene 0en als 1 (=gar nicht) bewertet und entsprechend umkodiert. Die ersten beiden Items waren in ihrer Form so wie sie später auch in die fertigen Tests einfließen sollten, so dass anhand dieser Fragen die Schwierigkeit und Trennschärfe der Items mit dem bisherigen Scoring betrachtet werden konnte. Die Ergebnisse der Intensitätsfrage flossen zusätzlich noch ins Scoring und somit in das *richtige* Antwortmuster ein, um die Aussagen der EmpathieexpertInnen noch zu erweitern.

Bei Frage eins wurde für jede *richtig* gegebene Antwort ein Punkt vergeben und die Summe dieser Werte als Itemscore gebildet (weiterhin als *Escore* – Emotions-Dekodierungs-Score - bezeichnet). Bei Item zwei wurde pro richtig gewähltem Wert auch ein Punkt vergeben und für Werte im Bereich +/- 1 Standardabweichung 0.5 Punkte (Übersicht der Scorings siehe Anhang B4). Auch aus diesen Punkten wurde ein Summenscore gebildet (weiterhin als *IScore* – Intensitäts-Score – bezeichnet).

Bei der Frage nach dem professionellem Lehrerverhalten wurde der Mittelwert jeder Antwort ermittelt und dadurch eine Rangreihenfolge gebildet. Das Scoring wurde dahingehend festgelegt, dass es pro richtig eingeschätztem Platz der Antwort 0.5 Punkte gibt. Die zusätzlichen Verhaltensoptionen wurden gesammelt und analysiert.

3.2.4 Ergebnisse

Zu Beginn sollte ermittelt werden, für wie wichtig die Teilnehmenden Empathie gegenüber SchülerInnen hielten. Der Mittelwert dieser Frage lag bei 3.75 ($SD=.44$; 1=*gar nicht wichtig* bis 4=*sehr wichtig*). Ebenfalls hielten 35 Personen empathisches Verhalten für professionelles Lehrerverhalten, eine Person verneinte das.

Zur Einschätzung der Güte der Items und des Scorings des Vignettentests wurden für die beiden schon in der Originalform vorliegenden Items eins und zwei Schwierigkeits- und Trennschärfeanalysen vorgenommen. Die Itemschwierigkeit gibt Auskunft darüber wie einfach bzw. schwierig es ist, ein vorliegendes Item richtig zu lösen (Moosbrugger & Kelava, 2007). Die Werte liegen zwischen 0 und 100 und es wird empfohlen Items zu präferieren, die im mittelschweren Bereich liegen. Allerdings können auch leichte oder schwere Items in einem Test eingesetzt werden. Da es sich in dieser Stichprobe um eine ExpertInnenbefragung mit LehrerInnen handelt, werden auch leichte Items akzeptiert, da die interessierende Stichprobe aus Lehramtsstudierenden besteht (also weniger Erfahrung und Wissen angenommen wird) und leichte Items in der Zielstichprobe unter Umständen schwieriger zu lösen sind.

In diesem Fall lagen die Schwierigkeits-Werte zwischen 38.46 (Escore Vignette 11) und 85.29 (Escore Vignette 2; siehe Abbildung 5 und Anhang B5).

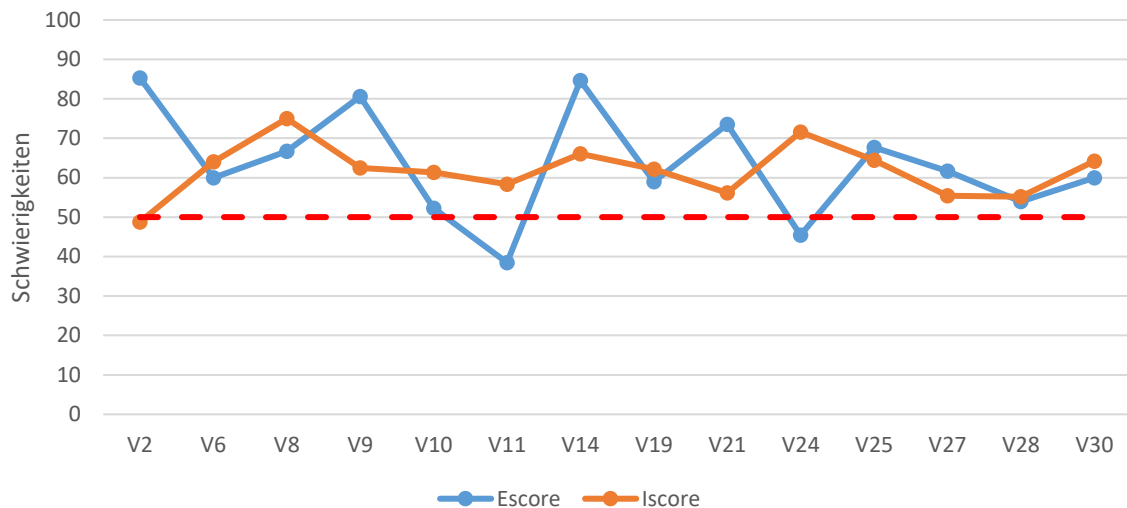


Abbildung 5. Schwierigkeiten von Items 1 (Escore=Emotionsdekodierungsscore) und 2 (Iscore=Intensitätsscore) über alle Vignetten hinweg in der LehrerInnenbefragung; V=Vignetten.

Neben der Itemschwierigkeit sollte allerdings auch die Itemtrennschärfe in die Auswahl der Items mit einfließen. Die Itemtrennschärfe gibt Auskunft darüber wie effektiv ein Item zwischen Personen mit unterschiedlicher Merkmalsausprägung differenzieren kann (Moosbrugger & Kelava, 2007). Die Trennschärfe kann Werte zwischen -1 und 1 annehmen, wobei negative Werte auf nicht umkodierte, invertierte Items hinweisen oder aber zeigen, dass ein Item unter Umständen nicht richtig verstanden wurde und somit für methodisch-inhaltliche Problematiken sprechen. Geeignete Items sollten zwischen .40 und .70 liegen. Zur Berechnung wird überprüft, wie hoch ein bestimmtes Item mit dem um das Item bereinigten Testwert eines Probanden korreliert (*part-whole-correction*). Die Bereinigung des Testwertes um den Itemwert sorgt dafür, dass die Werte die tatsächliche Trennschärfe nicht überschätzen (siehe Abbildung 6).

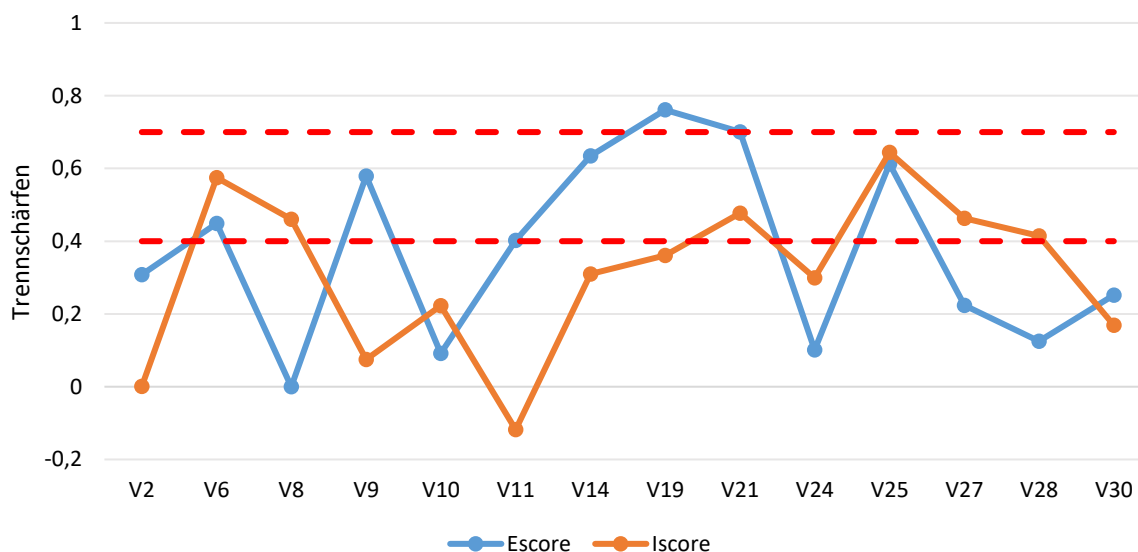


Abbildung 6. Trennschärfen von Item 1 und 2 über alle Vignetten hinweg in der LehrerInnenbefragung; V=Vignetten; Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score.

Die Trennschärfen lagen teilweise im Bereich zwischen .40 und .70, in einigen Fällen außerhalb und in einem Fall sogar im negativen Wertebereich.

Bei der Frage nach dem empathischen LehrerInnen-Verhalten wurden die am meisten genannten Aspekte aggregiert und in einer weiteren Antwortoption berücksichtigt (siehe Anhang B6). Dadurch entstanden vier verschiedene Antwortoptionen, die nun anhand der bewerteten Wichtigkeit in eine Rangreihenfolge gebracht werden können.

3.2.5 Auswertung und Implikation für die Weiterentwicklung des Vignettentestes

Nach der Auswertung auch dieser ExpertInnen-Befragung, die das Ziel hatte, weitere Informationen über die Vignetten und das Scoring zu sammeln (vor allem über die Verhaltensantworten) konnte der Itempool weiter ausgedünnt werden. Da die Schwierigkeitswerte des Escores teilweise etwas leicht waren, wurden die Distraktoren der betreffenden ersten Frage angepasst und nur noch Emotionsbegriffe genutzt, die in einer der im Pool befindlichen Vignetten vorkamen. Dadurch standen nun insgesamt noch 20

Emotionsbeschreibungen zur Verfügung, die so eingesetzt wurden, dass möglichst keine gegensätzlichen Emotionen als Antwortoption zur Auswahl standen (siehe Anhang B7). Zudem wurden die Videovignetten noch einmal reduziert. Aufbauend auf den Angaben der ersten ExpertInnenbefragung und der dort getroffenen Aufteilung in Empathie-auslösende und nicht-auslösende Vignetten, wurde darüber entschieden, welche weniger Empathie-auslösenden Vignetten für die weitere Entwicklung beibehalten werden sollten. Da keine dieser Vignetten im Vergleich mit allen anderen besonders gute Werte aufweisen konnte, wurde die Einschätzung des Empathie-Auslösens höher gewichtet und entsprechend eingeschätzte Items wurden in die weitere Entwicklung einbezogen (siehe Tabelle 15).

Tabelle 15
Übersicht Vignettenkennwerte LehrerInnenbefragung

Vignette	Escore		Iscore	
	Schwierigkeit	Trennschärfe	Schwierigkeit	Trennschärfe
V2	85.29	.31	48.77	.00
V6	60.00	.45	64.05	.58
V8	66.67	.00	75.00	.46
V9	80.56	.58	62.50	.08
V10	52.27	.09	61.36	.22
V11	38.46	.40	58.33	-.12
V14	84.62	.64	66.03	.31
V19	58.97	.76	62.18	.36
V21	73.53	.70	56.13	.48
V24	45.45	.10	71.59	.30
V25	67.68	.61	64.39	.64
V27	61.67	.22	55.42	.46
V28	53.97	.13	55.16	.42
V30	60.00	.25	64.17	.17

Anmerkungen. Escore=Summenwert von Item 1 (Emotions-Dekodierungs-Score);
Iscore=Summenwert von Item 2 (Intensitäts-Score).

Zur Erstellung der nächsten Testversion wurde versucht Vignetten mit ähnlichen Schwierigkeits-Werten zu paaren, um eine Aufteilung in Prä- und Post-Test möglich zu machen und zwei Testversionen mit einem ähnlichen Schwierigkeitsdurchschnitt zu

gewährleisten. Dazu wurden die Werte des E- und Iscore jeder in Frage kommenden Vignette gemittelt und der Mittelwert pro Testversion genommen (siehe Tabelle 16).

Tabelle 16

Zusammensetzung Testversionen nächster Entwicklungsschritt

Testversionen		Schwierigkeit <i>MW</i>	
1 (Prä)	2 (Post)	1 (Prä)	2 (Post)
V6	V6	62.02	62.02
V8	V9	70.83	71.53
V10	V21	56.82	64.83
V14	V2	75.32	67.03
V27	V25	58.54	66.04
V30	V30	62.08	62.08
<i>MW</i> Testschwierigkeit		64.27	65.59

Anmerkung. *MW*=Mittelwert; *V*=Vignette.

Vignette 6, die in allen Werten gute Ergebnisse brachte und Vignette 30 wurden sowohl im Prä-, als auch im Post-Test genutzt. Sie sollten dazu dienen eventuelle Veränderungen bzw. Entwicklungen besser nachvollziehen zu können.

Zur Absicherung des Scorings mit der neuen Antwortoption der Verhaltensfrage gab es eine weitere LehrerInnenbefragung Anfang 2018. Es nahmen fünf LehrerInnen teil, denen pro Vignette nur die Verhaltensantworten vorgelegt wurden, um sie beginnend mit der empathischsten Option noch einmal einschätzen zu lassen. Demografie-Abfrage, Empathiedefinition und Frage nach der Wichtigkeit von empathischem Verhalten für LehrerInnen waren mit den entsprechenden Teilen in der vorherigen LehrerInnen-Befragung identisch. Es nahmen vier weibliche und ein männlicher Proband mit einem Durchschnittsalter von 32.6 Jahren ($SD=1.52$ Jahre) teil. Vier der fünf ProbandInnen waren aktiv im hessischen Schuldienst und hatten 1-2 Jahre (2 Personen) bzw. 6-10 Jahre praktische Erfahrung (3 Personen). Vier Personen hielten den empathischen Umgang mit den SchülerInnen für sehr wichtig, eine Person für wichtig und alle waren sich einig, dass empathisches Verhalten

professionellem LehrerInnen-Verhalten entspricht, so dass die Einstellung der LehrerInnen in den beiden Befragungsrunden als sehr ähnlich bewertet werden kann.

3.3 Dritter Testentwicklungsschritt: Validierung mit Studierenden

Im nächsten Schritt wurden die ausgewählten Videovignetten einer studentischen Stichprobe vorgelegt. Die Daten wurden im Rahmen einer Bachelorarbeit erhoben und der Vignettentest mit weiteren Verfahren online über Unipark (Questback) bearbeitet. Die genauen Ergebnisse sind im Folgenden aufgeführt.

3.3.1 Stichprobe

Für die Validierung wurde die im Folgenden beschriebene Testbatterie online (*Unipark*) an einer studentischen Stichprobe der Goethe-Universität Frankfurt erhoben. Die Daten wurden im Rahmen einer Bachelorarbeit erhoben, die sich mit Empathie-Unterschieden in verschiedenen Studienfächern beschäftigte (Konopka, 2017).

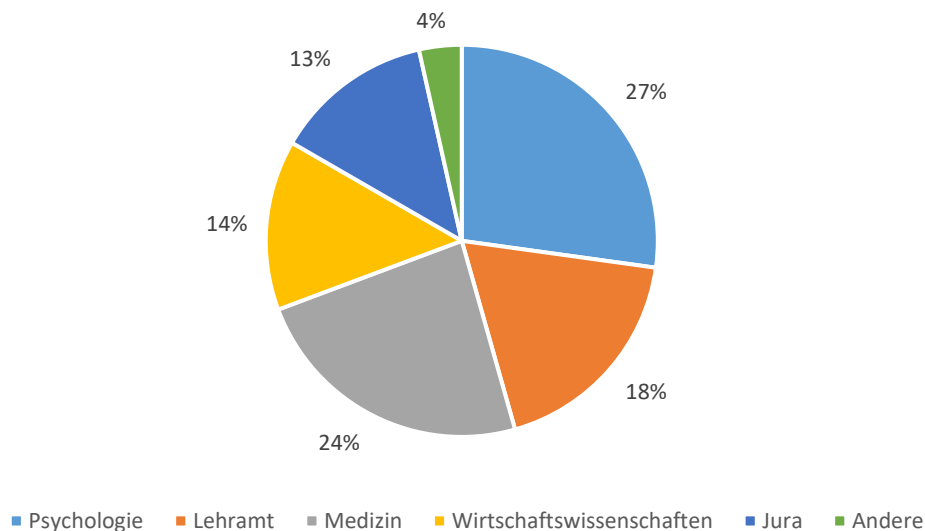


Abbildung 7. Prozentuale Verteilung der Studienfächer der teilnehmenden Studierenden der Validierung; Andere= Studienfächer Zahnmedizin, Germanistik/Anglistik, Wirtschaftspädagogik und Ethnologie.

Insgesamt griffen 1341 Personen auf die Umfrage zu, 212 Personen hinterließen verwertbare Daten. Der Datensatz wurde um all jene ProbandInnen bereinigt, die im Vignettentest weniger

als fünf der sechs Vignetten und zusätzlich noch den *Culture Fair Test (CFT)* bearbeitet hatten. Somit blieb ein finaler Datensatz von 112 Personen. Der Altersschnitt betrug 21,93 Jahre ($SD=4.68$ Jahre) und es nahmen 97 Frauen, sowie 16 Männer teil. Die Verteilung auf Studienfächer war gemischt (siehe Abbildung 7). Die meisten Studierenden waren in niedrigeren Semestern, da diese Zielgruppe den Fokus der Bachelorarbeit darstellte (siehe Abbildung 8). Schaut man sich die Teilnehmenden getrennt nach Testversions-Bearbeitung an gibt es bezüglich der BIG Five Dimensionen keine signifikanten Unterschiede (T-Test, zweiseitig; Extraversion: $t(110)=.41, p=.686$; Verträglichkeit: $t(110)=.18, p=.859$; Gewissenhaftigkeit: $t(110)=1.04, p=.303$; Neurotizismus: $t(110)=-.30, p=.762$; Offenheit: $t(110)=.51, p=.610$), ebenfalls nicht beim CFT-Score ($t(110)=-1.00, p=.322$). Auch Alter- ($t(110)=1.62, p=.108$), Geschlechts- ($t(110)=.54, p=.593$) und Studienfachverteilungen (Psychologie: $t(110)=-.42, p=.673$; Lehramt: $t(110)=-.24, p=.811$; Medizin: $t(110)=-.66, p=.512$; WiWi: $t(110)=.54, p=.593$; Jura: $t(110)=.28, p=.784$; Andere: $t(110)=1.01, p=.313$) sind in beiden Gruppen ähnlich.

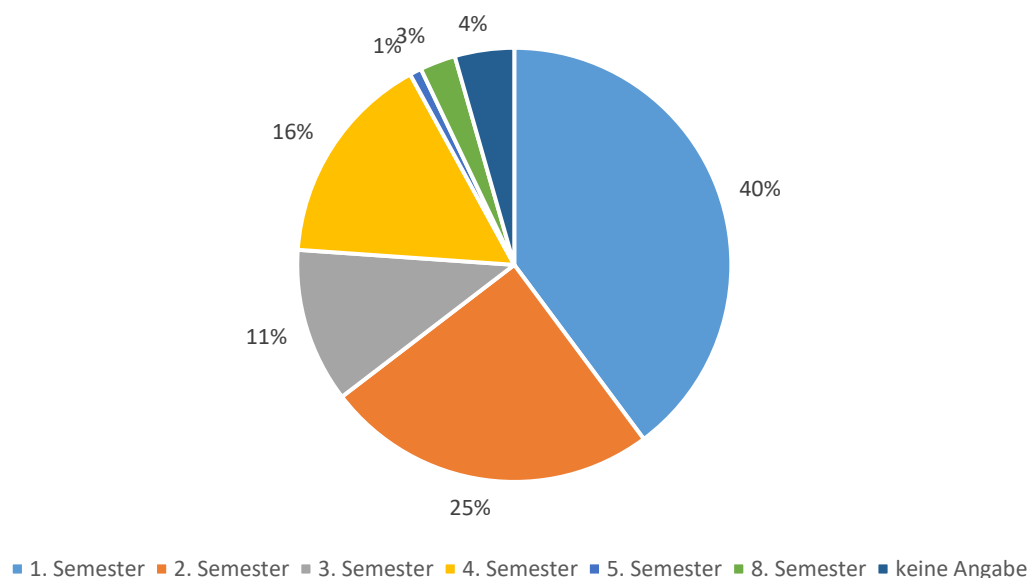


Abbildung 8. Fachsemester der Teilnehmenden der Validierung.

3.3.2 Instrumente

3.3.2.1 Essener Fragebogen zur Erfassung des subjektiven Empathiekonzepts – ESEK

Als erstes Testverfahren wurde der *Essener Fragebogen zur Erfassung des subjektiven Empathiekonzeptes (ESEK)* eingesetzt (Palenga & Schönefeld, 2014). Der Originalfragebogen umfasst 38 Items und unterteilt sich in zwei Skalen mit je zwei Subskalen, die Aussagen auf Gruppenebene zulassen. Die Skalen fragen zum einen allgemein nach dem subjektiven Verständnis von Empathie und zum anderen nach einem berufsbezogenen Konzept. In jeder Skala werden verschiedene Aussagen zu Empathie gemacht und die Probanden sollen angeben, welche Aussagen ihrer Meinung nach zum Konzept Empathie dazugehören (Skala I) bzw. welche Erwartungen sie bezüglich eines empathischen Umgangs mit anderen an sich selbst haben (Skala II). Die Items beider Skalen sind gleich, sie unterscheiden sich lediglich durch die Instruktion. Die Einschätzung erfolgt via einer Likert-Skala von *Gar nicht* (1; Skala I) bzw. *Keine Erwartung* (1; Skala II) bis *Sehr stark* (6; Skala I) bzw. *Sehr hohe Erwartung* (6; Skala II). Die beiden Skalen können getrennt voneinander eingesetzt werden. Items sind beispielsweise „Mitleid haben“ (Item 1), „Erst einmal einen Ratschlag geben können“ (Item 6) oder „Mich genauso fühlen wie mein Gegenüber“ (Item 7). Jede Skala unterteilt sich noch einmal in zwei weitere Subskalen: in ein funktionales und ein dysfunktionales Empathieverständnis (allgemeine Skala I) bzw. –erwartung (berufsbezogene Skala II). Hiermit soll ermittelt werden, ob die ProbandInnen ein gesundes ausgewogenes Verständnis von Empathie mitbringen oder dysfunktionale Anteile überwiegen. Ein Beispiel für das funktionale Verständnis wäre eine hohe Zustimmung der Aussage: „Zwischen meinen eigenen Gefühlen und denen meines Gegenübers unterscheiden.“ Ein Beispiel für ein eher dysfunktionales Verständnis wäre eine hohe Zustimmung für „Immer dafür sorgen, dass es meinem Gegenüber wieder besser geht.“ Ursprünglich wurde der berufsspezifische Teil des Verfahrens für Pflegekräfte entwickelt und anhand einer solchen Stichprobe psychometrisch

überprüft. Hierbei wird angenommen, dass sich vor allem bei überwiegenden dysfunktionalen Anteilen das Burnout-Risiko erhöhen kann (Palenga, 2014) und diese Werte optimalerweise niedrig ausfallen sollten. Die Überprüfung an den beiden Stichproben zeigte gute interne Konsistenzen (siehe Tabelle 17). Ebenso konnten mittels einer Faktorenanalyse die beiden Subskalen als Faktoren bestätigt werden (Palenga, 2014).

Tabelle 17

Interne Konsistenzen des ESEK

Skala	Item-anzahl	Cronbachs α	
		Stichprobe I – stationäre Pflegekräfte ^a	Stichprobe II – ambulante Pflegekräfte ^b
Funktionales Empathieverständnis	9	.81	.78
Dysfunktionales Empathieverständnis	10	.86	.80
Funktionale Empathieerwartung	9	.86	.86
Dysfunktionale Empathieerwartung	10	.89	.81

Anmerkungen. ^a $N = 155$, ^b $N = 57$.

Die internen Konsistenzen in der Validierungsstichprobe dieser Arbeit ($N=112$) für die Skala des Empathieverständnisses lagen mit $\alpha=.63$ (funktionale Erwartung) bzw. $\alpha=.74$ (dysfunktionale Erwartung) etwas niedriger als bei der Pflegekräfte –Stichprobe. Es werden auf Vignetten- oder Testwertebene positive Korrelationen mit der funktionalen Dimension und negative Korrelationen mit der dysfunktionalen Dimension erwartet.

3.3.2.2 Toronto Empathy Questionnaire – TEQ

Der Toronto-Empathy Questionnaire - TEQ; (Spreng et al., 2009) ist ein Selbstauskunftsfragebogen mit 16 Items. Die Items werden auf einer 5-stufigen Likert Skala von *nie* (1) bis *immer* (5) beantwortet. Das Verfahren wurde aus einer Vielzahl bislang

bestehender Instrumente zur Messung von Empathie entwickelt und bereits im Theorieteil ausführlich erläutert (vgl. Kapitel 3.1.3.5).

3.3.2.3 Interpersonal Reactivity Index – IRI

Der Interpersonal Reactivity Index (Davis, 1980, 1983) ist ein Selbstauskunftsfragebogen zur Erfassung von Empathiefacetten. Er bedient sich als einer der ersten Verfahren einer multidimensionalen Sichtweise auf Empathie, enthält 28 Items (neun davon in invertierter Form) und gliedert sich in vier Subskalen: Perspective Taking (PT), Empathic Concern (EC), Fantasy (FS) und Personal Distress (PD). Die Items werden auf einer 5-stufigen Likert Skala von *beschreibt mich überhaupt nicht* (1) bis *beschreibt mich sehr gut* (5) bewertet. Auch dieses Instrument wurde bereits ausführlich erläutert (vgl. Kapitel 2.1.3.4). Die internen Konsistenzen der hier vorliegenden Validierungsstudie ($N=112$) lagen bei $\alpha=.82$ (FS), $\alpha=.71$ (PT), $\alpha=.80$ (EC) und $\alpha=.77$ (PD) und somit in den meisten Subskalen sogar höher als in den bisherigen psychometrischen Überprüfungen. Es werden positive Korrelationen mit PT, FS und EC erwartet, sowie keine oder negative Korrelationen mit PD.

3.3.2.4 Culture Fair Test (CFT) – Untertest: Matrizen

Der *CFT*³ ist ein sprachfreies Messverfahren zum Erfassen von fluider Intelligenz nach Cattell in der Überarbeitung von Weiß (2006) und besteht aus vier Subskalen mit insgesamt 92 Items. Fluide Intelligenz bezeichnet allgemeine Anteile des Denkens, die als biologische Disposition angenommen werden (Händel, 2018). Im Verfahren selbst werden Figuren gezeigt, die fortgeführt, eingeordnet, komplettiert und anhand derer sprachunabhängig

³ © Hogrefe Verlag GmbH & Co. KG, Göttingen

Schlussfolgerungen gezogen werden sollen (Amelang & Schmidt-Atzert, 2006). Das Verfahren wurde entwickelt, um eine Messung von Intelligenz zu gewährleisten, die frei ist von kulturabhängigen Aspekten wie Sprache und Bildung (Cattell & Cattell, 1959). Der Test besteht aus zwei Teilen und liegt in zwei parallelen Versionen vor, die sich dahingehend unterscheiden, dass die Reihenfolge der Aufgaben variiert, so dass in Gruppentestungen Abschreiben vermieden werden kann. Der komplette Test lässt sich als Langform (Teil I und II) oder als Kurzform (nur Teil I) durchführen und sowohl in Einzel-, als auch in Gruppensettings testen. Der Test kann für Personen von 8 Jahren 5 Monaten bis 80 Jahren angewandt werden. Hier in dieser Validierungstestung kam nur der Matrizen-Subtest des Teils A mit 15 nach Schwierigkeit geordneten Items zum Einsatz, der auch anderweitig schon alleinig zur Testung logisch-abstrakten Denkens eingesetzt wurde (Lorenz, 2011). Zu Beginn wurden drei Items als Beispiel gegeben (siehe Anhang C1), um in die Aufgabe einzuführen, zu denen es ein unmittelbares Feedback gab, ob sie richtig gelöst wurden. Die Bearbeitungszeit war begrenzt und betrug 4 Minuten. Die Reliabilität wurde für den kompletten ersten Teil mit $\alpha=.92$ angegeben. Die Reliabilität in dieser Validierung lag für den Matrizen-Subtest-Teil mit $\alpha=.57$ in einem Bereich, der für Leistungstests zwischen $\alpha=.80$ und $\alpha=.90$ liegen sollte (Moosbrugger & Kelava, 2007). Deutlich niedriger auch als die Gesamtreliabilität, die allerdings noch andere Teile beinhaltet. Es werden keine Zusammenhänge mit den Vignetten- bzw. Testscores erwartet.

3.3.2.5 Big Five Inventory (BFI-K)

Der *BFI-K* ist ein Persönlichkeitsfragebogen zur Erfassung von Persönlichkeitseigenschaften (Rammstedt & John, 2005). Er basiert auf dem Big Five Inventory, das sich am Fünf-Faktoren-Modell der Persönlichkeit orientiert (vgl. John & Srivastava, 1999) und besteht aus 21 Items. Die Items werden auf einer 5-stufigen Likert Skala von *sehr unzutreffend* (1) bis *sehr*

zutreffend (5) eingeschätzt (siehe Anhang C2). Die Originalversion des Big Five Inventory war im Bereich der Messung des Fünf-Faktoren-Modells das bisher ökonomischste Instrument mit 44 Items (John, Donahue & Kentle, 1991) und setzt sich aus den Subdimensionen Extraversion, Verträglichkeit, Gewissenhaftigkeit, Neurotizismus und Offenheit für neue Erfahrungen zusammen. Ursprünglich wurde das Modell mit Hilfe des lexikalischen Ansatzes entwickelt, bei dem Wörterbücher einer Sprache systematisch nach Eigenschaftsworten durchsucht, die Ergebnisse in einem mehrstufigen Prozess reduziert und durch verschiedene Einschätzungsverfahren, sowie –analysen auf eine überschaubare Menge geschrumpft wurden (Asendorpf, 2007).

Die Itemselektion der Kurzversion beruhte auf psychometrischen und inhaltlichen Überlegungen und orientierte sich an der Zusammenstellung der 44-Item-Version. Die Kurzversion wurde mit zwei Stichproben validiert, die Reliabilitäten sind in Tabelle 18 aufgelistet.

Tabelle 18

Interne Konsistenzen des BFI

Subskala	Cronbachs α		
	BFI-K Stichprobe I ^a	BFI-K Stichprobe II ^b	Originalversion BFI ^c
Extraversion	.86	.81	.86
Verträglichkeit	.64	.50	.74
Gewissenhaftigkeit	.70	.69	.84
Neurotizismus	.74	.77	.85
Offenheit für neue Erfahrungen	.66	.70	.75

Anmerkungen. Übernommen aus Rammstedt & John 2005. ^a $N=459$, studentische Stichprobe, $M= 25.8$ Jahre, $SD= 5.7$); ^{b+c} $N=391$, studentische Stichprobe, $M= 24.1$ Jahre, $SD= 7.7$.

Die internen Konsistenzen der Vignettentest-Validierung sind in dieser studentischen Stichprobe ($N=112$) ähnlich der BFI-K Stichprobe. Es werden keine bis leicht positive Werte

der Emotionsmaße mit Neurotizismus und Verträglichkeit erwartet, sonst werden keine bestimmten Zusammenhänge mit dem Vignettentest erwartet.

3.3.3 Datenaufbereitung

Zur Datenaufbereitung wurden zuerst die Missings bearbeitet und dann die einzelnen Test-Scores gebildet. Im ESEK gab es keine Missings, so dass direkt die Summenscores der beiden Subskalen (funktional/dysfunktional) gebildet werden konnten – maximal 54 Punkte in der funktionalen und 60 in der dysfunktionalen Dimension. Für den IRI wurden die invertierten Items umgepolt und der Mittelwert der Items jeder Subskala erstellt. Missings gab es keine. Beim Bilden des Testscores des TEQ ist ein fehlendes Item zulässig, so dass der Testscore als Mittelwert aus mindestens 15 Items gebildet werden kann (auch hier wurden invertierte Items erst umkodiert). Missings gab es bei dieser Stichprobe auch hier nicht, so dass alle Werte vorlagen und der Testscore ohne Einschränkung formiert werden konnte. Ein ähnliches Vorgehen gab es beim Bilden der BIF-K Subdimensionen (umkodieren invertierter Items und Mittelwert der Subskalen bilden). Beim CFT gab es ein paar fehlende Werte (wenn die Zeit z.B. abgelaufen war), die beim Umkodieren der Items einbezogen wurden und bei dem richtige Antworten mit 1 und falsche bzw. fehlende Antworten mit 0 gescored wurden. Der Gesamtwert wurde als Summe dieser Scorings gebildet und konnte maximal den Wert 15 erreichen. Für den Vignettentest wurden drei Scores gebildet: Pro Vignette ein Emotionserkennungs-Score (Escore), ein Intensitäts-Score (Iscore), und ein Verhaltens-Score (Vscore), die jeweils über die sechs Vignetten einer Testversion aufsummiert wurden (GesamtEscore, GesamtIscore, GesamtVscore). Wenn die Testversion oder Vignette nicht angeschaut wurde, wurden Missings in den Daten entweder mit -99, mit -77 bzw. -66 kodiert. Alternativ wurde der fehlende Wert durch 1 ersetzt, wenn bspw. bei der Frage nach der Emotionsintensität auf einer der sechs Basisemotionen ein Wert ≥ 1 abgegeben wurde. In

letzterem Fall konnte davon ausgegangen werden, dass bestimmte Emotionen in der angegebenen Intensität gesehen wurden und die weiteren nicht, so dass die Studierenden durch die Nicht-Angabe das Nicht-Vorhandensein meinten. Zusätzlich wurden Antwortmuster (z.B. nur 0en bei Frage eins oder 1en bei Frage zwei) durch Missings ausgetauscht. Dieses Muster entstand, wenn weitergeklickt wurde, ohne eine Antwort zu hinterlassen (was Missings/nicht gegebenen Antworten entspricht). Beim Einstufen der Verhaltensantworten wurden nicht gegebene Antworten mit 0 kodiert, um beim anschließenden Scoring auch 0en (=nicht richtig) vergeben zu können. Bei dem Video der Vignette 6 in der Testversion 2 zeigte es sich, dass es viele Abspielprobleme vermutlich technischer Art gab, so dass nur ein Bruchteil der gegebenen Antworten genutzt werden konnte. Das war erkennbar an den Missingwerten auf der Variablen „Abspielen der Vignette 6“ und den hinterlassenen Freitext-Kommentaren, die in der Mehrzahl auf technische Schwierigkeiten hinwiesen. Deswegen wurde diese Vignette von den weiteren Auswertungen ausgeschlossen. Im Vignettentest gab es auch die Möglichkeit bei Angabe der Gefühlsqualität eine „Andere“-Möglichkeit zu markieren und ein weiteres Gefühl zu nennen, das beobachtet wurde (siehe Tabelle 19).

Tabelle 19
Scoring des Vignettentests

Item	Frage	Antwortmöglichkeit	Scoring
1	Welche Emotionen konnten Sie hauptsächlich bei dem Schüler sehen?	6 Antwortoptionen (Mehrfachnennung möglich); Ergänzung eines eigenen Gefühls möglich	1 Punkt/richtiger Antwort → Bilden eines Summenscores
2	In welcher Intensität haben Sie die folgenden Emotionen hauptsächlich erkennen können?	6 Basisemotionen und Ratingskala von 1(gar nicht) bis 6 (sehr stark)	1 Punkt/richtiger Antwort (=MW der ExpertInnenbefragung); 0.5 Punkte für Werte +/- 1 SD → Bilden eines Summenscores
3	Wenn Sie in der Situation als Lehrperson dabei gewesen wären, welches Verhalten wäre professionell und empathisch gewesen?	4 Antwortoptionen, die von 1= <i>bestes Verhalten</i> bis 4= <i>am wenigsten günstiges Verhalten</i> und die Möglichkeit ein alternatives Verhalten zu benennen, reichen	0.5 Punkte/richtiger Antwort (=MW der LehrerInnenbefragung ohne das angepasste Scoring aus Anfang 2018)

Die Anzeige-Reihenfolge der Vignetten und der Antwortoptionen pro Item erfolgte randomisiert, ebenso wie die Zuteilung zu den beiden Testversionen. Zum Bestimmen der folgenden Analyseverfahren wurden die Daten via *Kolmogorov-Smirnov-Test (KS-Test)* auf Normalverteilung geprüft. Der KS-Test ist ein Verfahren für metrische Daten, das die empirische Stichprobe mit einer Normalverteilung gleichen Mittelwertes und Standardabweichung vergleicht (Field, 2013). Signifikante Ergebnisse weisen auf eine nicht gegebene Normalverteilung hin. Für nahezu alle Vignetten kann die Normalverteilungsannahme nicht bestätigt werden (siehe Anhang C3), so dass im weiteren Verlauf nicht-parametrische Analysemethoden zum Einsatz kommen.

3.3.4 Ergebnisse

Für die folgenden Korrelationsberechnungen wird *Kendalls Tau* als Non-parametrisches Korrelationsverfahren eingesetzt. Es handelt sich um ein Rangkorrelationsverfahren, bei dem erst Ränge gebildet werden, auf die dann *Pearsons Korrelation* angewandt wird (Field, 2013). *Kendalls Tau* ist wie *Spearman's Rho* robust gegenüber Verletzungen der Vorannahmen (wie Normalverteilungsverletzung), sowie Ausreißern und vor allem bei kleineren Stichproben geeignet, die hier aufgrund der zwei Testversionen vorliegen. Der Wert bewegt sich zwischen -1 und +1 und kann ähnlich gedeutet werden wie *Pearsons Korrelationskoeffizient* r (+/-0.1 ist ein kleiner Effekt, +/-0.3 ist ein mittlerer Effekt und +/-0.5 ist ein großer Effekt). Die Ergebnisse der signifikanten Korrelationen mit Empathie-relevanten Maßen sind in

zu sehen (der Übersichtlichkeit halber sind nur Variablen angegeben, bei denen signifikante Korrelationen gefunden werden konnten).

Tabelle 20

Korrelationstabelle Vignetten und Empathie-relevante Maße (Validierung)

Vignettenwert	ESEK funk (N)	ESEK Dysfunk (N)	IRI PT (N)	IRI FS (N)	IRI EC (N)	IRI PD (N)	TEQ (N)
Escore V2	.24* (55)						
Escore V6T1	.22* (56)						
Iscore V6 T1				-.20* (32)			
Iscore V6 T2				.28* (32)			
Escore V9					.27* (54)		.31** (54)
Vscore V9			.27* (54)				
Iscore V10	.21* (56)						
Escore V14		-.29* (56)					
Vscore V14				.40** (56)			.22* (56)
Escore V21		-.23* (56)	.28* (56)			-.22* (56)	
Escore V25	.23* (56)			.25* (56)	.32** (56)		.23* (56)
Iscore V25	.21* (56)						.20* (56)
Vscore V30T2					.26* (55)		.32** (55)
GesamtVVT1				.26* (51)			
GesamtEscoreT2	.24* (54)		.27* (54)	.21* (54)	.29** (54)		.23* (54)
GesamtVVT2	.21* (54)						

Anmerkungen. * $p < .05$, ** $p < .01$; ESEKfunk.=ESEK, Subskala funktional; ESEKDysfunk.=ESEK, Subskala Dysfunktional; IRI PT=IRI, Subskala Perspective Taking; IRI FS=IRI, Subskala Fantasy; IRI EC=IRI, Subskala Empathic Concern; IRI PD=IRI, Subskala Personal Distress; V=Vignette; T=Testversion (bei doppelt eingesetzten Vignetten); Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score; Vscore=Verhaltensscore.

Des weiteren finden sich in Tabelle 21 die Korrelationen mit den weiteren eingesetzten Testverfahren zur Validierung. Der Übersichtlichkeit halber sind auch hier nur signifikante Korrelationen abgetragen.

Tabelle 21

Korrelationstabelle Vignetten und BFI-K sowie CFT (Validierung)

Variable	BFI-K E (N)	BFI-K V (N)	BFI-K G (N)	BFI-K N (N)	BFI-K O (N)	CFT (N)
Escore V2		-.35** (55)				.25* (55)
Escore V6T1				-.23* (56)		
Vscore V6 T1					.24* (56)	
Iscore V6 T2						.30* (32)
Escore V9				.26* (54)		
Iscore V9						.25* (54)
Vscore V10			.33** (51)		.23* (51)	
Escore V25				.28* (56)		
Escore V27	.26* (56)					
Vscore V27					.24* (56)	
GesamtVVT1			.21* (51)		.33** (51)	
GesamtEscoreT2				.27** (54)		
GesamtVVT2				.24* (54)		

Anmerkungen. * $p < .05$, ** $p < .01$; BFI-K E= BFI-K Extraversion, BFI-K V=BFI-K Verträglichkeit, BFI-K G=BFI-K Gewissenhaftigkeit, BFI-K N=BFI-K Neurotizismus, BFI-K O=BFI-K Offenheit, CFT=Culture Fair Subtest Matrizen; V=Vignette; Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score; Vscore=Verhaltensscore.

Ebenfalls wurden die internen Konsistenzen der Testversionen (Cronbachs Alpha) überprüft, die wie folgt aussehen (Tabelle 22):

Tabelle 22

Vignettentest reliabilitäten, interne Konsistenzen (Validierung)

	Cronbachs α	
	Testversion 1 (N)	Testversion 2 (N)
GesamtEscore	.61 (56)	.71 (54)
GesamtIscore	.66 (56)	.76 (54)
GesamtVscore	.53 (51)	.64 (54)

Anmerkungen. Testversion 1= V6, V8, V10, V14, V27, V30 (V=Vignette); Testversion 2= V2, V9, V21, V25, V30 (V=Vignette); N=ProbandInnenanzahl; GesamtEscore= Gesamtsummenscore des Emotionsdekodierungs-Scores; GesamtIscore= Gesamtsummenscore des Intensitäts-Scores; GesamtVscore= Gesamtsummenscore des Verhaltens-Scores.

3.3.4.1 Itemanalysen

Da nach der LehrerInnenbefragung nun auch eine Antwortschablone für die Verhaltensantworten vorlag, wurden die Items erneut psychometrisch untersucht. Die Itemschwierigkeiten bewegten sich bei den meisten Items im Bereich zwischen 50-65. Das Item zur Einschätzung der Gefühlsqualität hatte in fast allen Vignetten Werte >70 (siehe Abbildung 9).

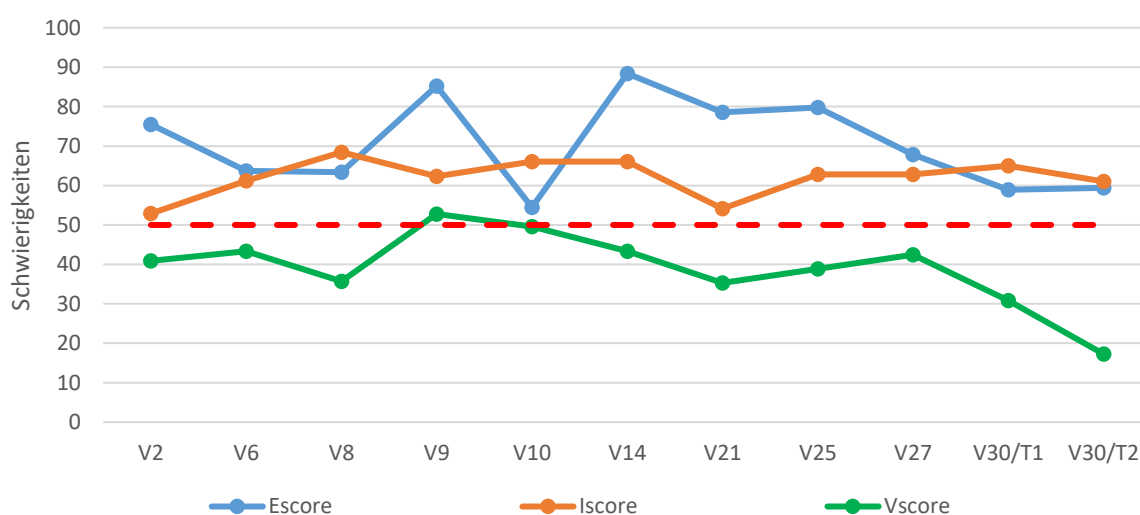


Abbildung 9. Itemschwierigkeiten (Validierung); N=56, außer V2 (n=55), V9 (n=55), V30/T2 (n=55); V=Vignetten; Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score; Vscore=Verhaltensscore.

Die Trennschärfen bewegen sich für fast alle Items zwischen .30 und .70 (siehe Abbildung 10).

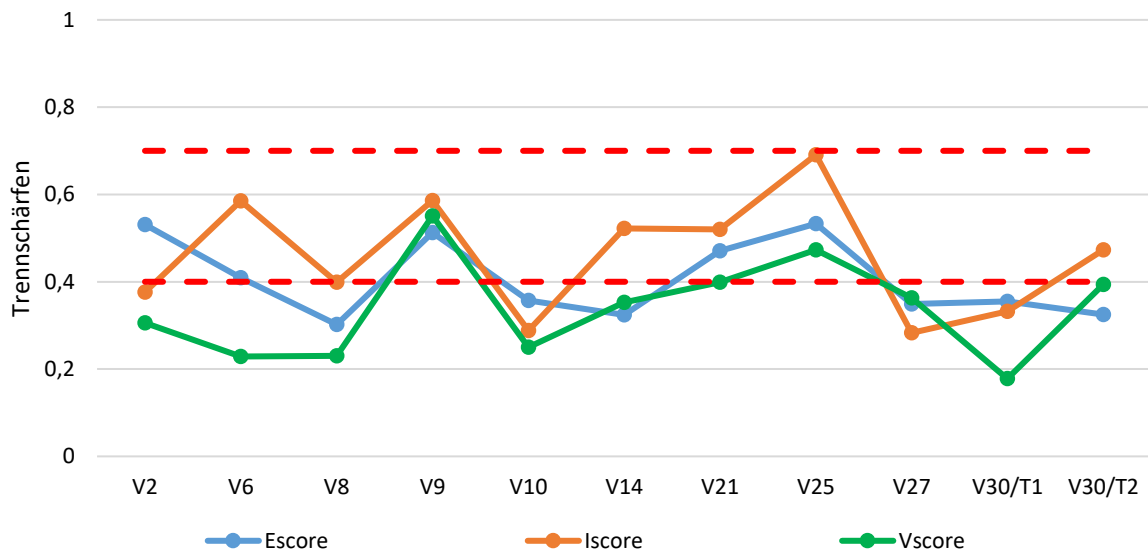


Abbildung 10. Trennschärfen (Validierung); $N=56$, außer V2 ($n=55$), V9 ($n=55$), V30/T2 ($n=55$); V=Vignetten; Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score; Vscore=Verhaltensscore.

Schaut man sich die Psychometrie der Items nur für die Lehramtsstudierenden an, zeigen sich die gefundenen Effekte in extremerer Form: Besonders die Werte des Escore sind über alle Items hinweg $>.70$ bzw. $.80$ (siehe Abbildung 11).

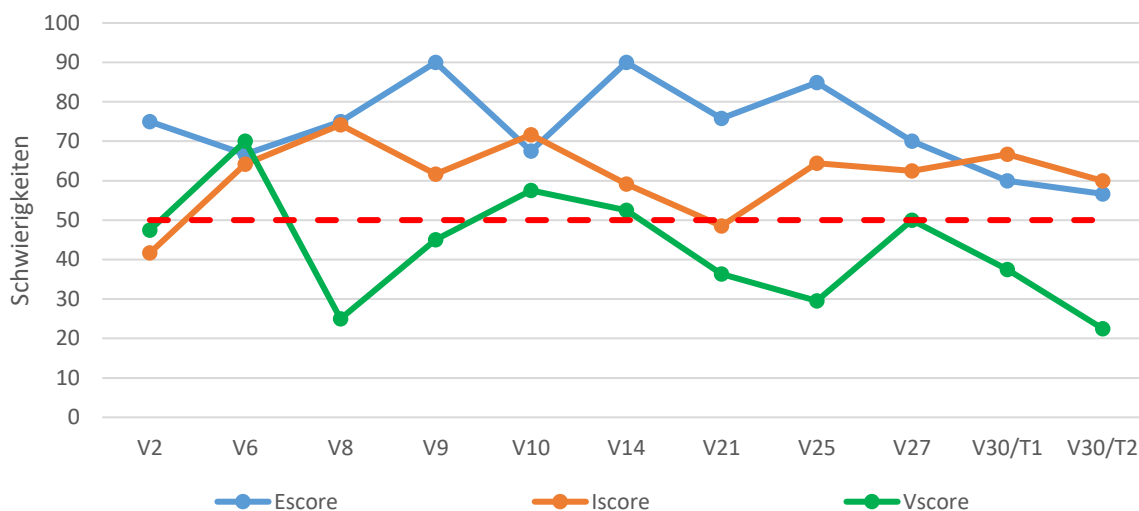


Abbildung 11. Schwierigkeiten, nur Lehramtsstudierende (Validierung); N=10, außer V21 und V25 (n=11); V=Vignetten; Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score; Vscore=Verhaltensscore.

Auch bezüglich der Trennschärfen zeigt sich ein anderes Bild. Es zeigen sich negative Trennschärfen und sehr diverse Werte (siehe Abbildung 12).

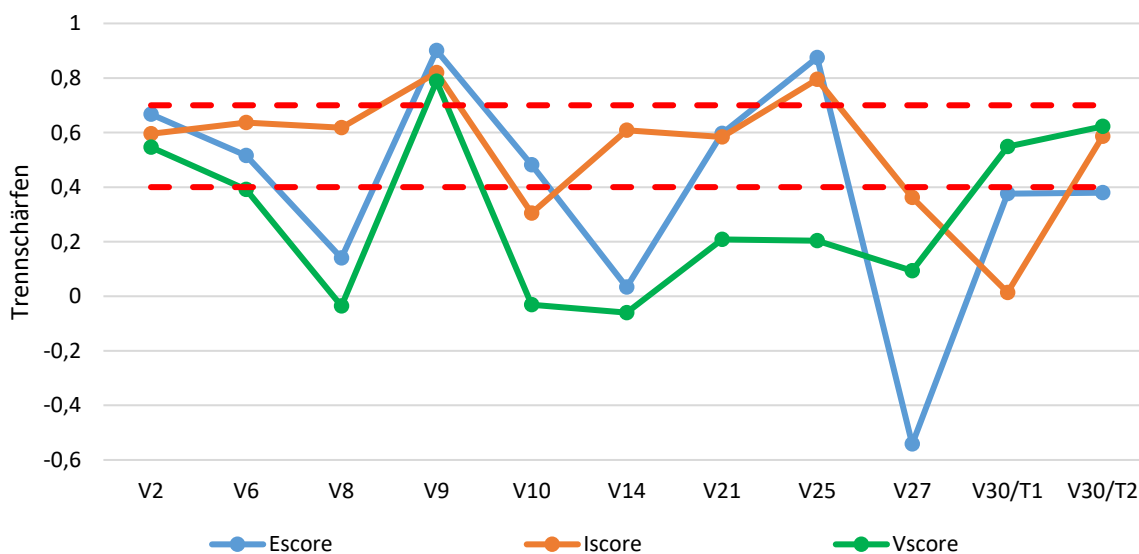


Abbildung 12. Trennschärfen nur Lehramtsstudierende (Validierung); V=Vignetten; Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score; Vscore=Verhaltensscore.

Die extremen Werte zeigen sich vor allem bei Vignetten der Testversion 1, was sich auch bei den internen Konsistenzen zeigt, deren Werte in dieser Version geringer ausfallen als in der Testversion 2 und der Gesamtstichprobe (siehe Tabelle 23).

Tabelle 23

Vignettentest reliabilitäten, interne Konsistenzen (Validierung, nur Lehramtsstudierende)

	Cronbachs α	
	Testversion 1 (N)	Testversion 2 (N)
GesamtEscore	.40 (10)	.85 (10)
GesamtIscore	.67 (10)	.85 (10)
GesamtVscore	.27 (10)	.70 (10)

Anmerkungen. Testversion 1= V6, V8, V10, V14, V27 (V=Vignette), V30; Testversion 2= V2, V9, V21, V25, V30 (V=Vignette); GesamtEscore=Gesamtsummenscore des Emotionsdekodierungs-Scores; GesamtIscore=Gesamtsummenscore des Intensitäts-Scores; GesamtVscore=Gesamtsummenscore des Verhaltens-Scores.

3.3.5 Auswertung und Implikation für die weiteren Testversionen

Die Validierung an der studentischen Stichprobe konnte zeigen, dass die ausgewählten Vignetten zumindest in Teilen Empathiefacetten abbilden. Es zeigten sich kleine bis mittlere Korrelationen mit Empathie-relevanten Maßen, die außer in Vignette V8 und V27 in allen Vignetten zu finden waren (vgl. Tabelle 20). Die Korrelationen sind zwar eher klein, aber da es Hinweise darauf gibt, dass Selbstauskunftsfragebögen und performanzbasierte Instrumente nur eingeschränkt miteinander korrelieren, können diese Zusammenhänge trotzdem als Indiz dafür gewertet werden, dass Empathie-relevante Facetten erfasst werden (vgl. Richter, 2009; Decety & Ickes, 2011).

Ebenfalls fanden sich Zusammenhänge zwischen den Vignetten und dem BFI-K, die auf der Dimension *Neurotizismus* erwartungskonform mit keinen oder kleinen positiven Zusammenhängen (bis auf V6 mit einem negativen Zusammenhang) und auch mit *Verträglichkeit* einen negativen Zusammenhang (bei V2) auftraten.

Auch die Reliabilitäten der beiden Testversionen lagen mit Werten zwischen $\alpha=.27$ bis $\alpha=.85$ im Vergleich zu anderen Testarten teilweise etwas niedrig - z.B. Leistungstests mit einem Cronbachs Alpha von .8 bis .9 oder Persönlichkeitstests mit annehmbaren internen Konsistenzen von $\alpha=.7$ bis $\alpha=.8$ (Moosbrugger & Kelava, 2007). Dies gilt vor allem für die Werte der Testversion 1, die am niedrigsten ausfielen. Die internen Konsistenzen in Testversion 2 waren in einem akzeptablen Bereich.

Die Schwierigkeiten der Vignetten lagen alle im angestrebten mittelschweren Bereich (gemittelte Werte pro Vignette) und auch die gemittelten Trennschärfen zeigten, dass fast alle Vignetten eine gute Diskriminationsfähigkeit besitzen und Personen verschiedener Ausprägungen gut unterscheiden können. Lediglich die Vignetten V8, V10, V27 und V30 liegen unter der .4-Grenze. Bei den Einzelwerten der Vignetten (Escore, Iscore, Vscore) ist zu erkennen, dass die Frage nach der Emotionsqualität bei einigen Items die Tendenz hat zu leicht

zu sein, so dass im nächsten Entwicklungsschritt das Scoring angepasst wird und pro falsch gegebener Antwort 0.5 Punkte abgezogen werden. Im Gegensatz dazu ist die Schwierigkeit der Verhaltensantwort etwas schwer, so dass auch hier das Scoring angepasst wird und nun pro richtig gewählter Position 1 Punkt und bei einer Platzierung innerhalb $\pm 1/2$ Standardabweichung 0.5 Punkte vergeben werden. Das Scoring wurde bei der nächsten Erhebungsrunde in Präsenz getestet, um sie mit der interessierenden Zielgruppe zu pilotieren. Bei der Validierung zeigte sich mit einer kleinen Teilstichprobe, dass Lehramtsstudierende teilweise andere Werte bei den Vignetten erzielten als Studierende anderer Fächer (siehe Tabelle 24).

Tabelle 24

Gemittelte Trennschärfen, nur Lehramtsstudierende (N=10)

Vignette	Schwierigkeit (gemittelt)	Trennschärfe (gemittelt)
V2	54.72	0.60
V6	66.94	0.52
V8	58.06	0.24
V9	65.56	0.84
V10	65.56	0.25
V14	67.22	0.19
V21	53.54	0.46
V25	59.60	0.63
V27	60.83	-0.03
V30	50.56	0.42

Mit der nächsten Testung soll untersucht werden, ob sich diese Tendenz weiterhin zeigt oder bei der kleinen Stichprobengröße nur ein Stichprobenartefakt der Validierung vorlag. Für den nächsten Testentwicklungsschritt sollen die Reliabilitäten erhöht werden, so dass die Testversionen anhand der gemittelten Schwierigkeitswerte angepasst und neue Testzusammensetzungen gebildet werden (siehe Tabelle 25).

Tabelle 25

Testzusammensetzungen für die Testung im Sommersemester 2017

Testversion		Schwierigkeiten		Trennschärfe	
1	2	1	2	1	2
V6	V6	54.54	54.54	0.52	0.52
V9	V8	62.58	54.39	0.84	0.24
V10	V9	55.02	62.58	0.25	0.84
V14	V21	61.94	54.50	0.19	0.46
V21	V25	54.50	59.60	0.46	0.63
V30	V27	49.06	55.77	0.42	-0.03
Ø-Schwierigkeit		56.27	56.90	0.45	0.44

Anmerkung. V=Vignette.

Zudem wurden bei der nächsten Untersuchung zwei Messzeitpunkte angesetzt, um die Veränderungssensitivität der Tests zu untersuchen. Als noch fehlende Facette wurden nun auch Videoaufnahmen der Gesichter der Teilnehmenden während des Vignettentests aufgenommen, um anschließend mit einer Face-Reader-Software auf gezeigte Emotionen analysiert zu werden und gemeinsam mit anderen Items die Facette des *Mitschwingens* abzubilden. Im weiteren Verlauf soll die Implementation dieser Datenquelle geprüft werden. Der Vignettentest wird wieder eingebettet in einer Testbatterie und mit verschiedenen Verfahren zur Emotionsregulationsfähigkeit und Empathie präsentiert. Das neue Scoring kommt zur Anwendung und es werden in diesem Zusammenhang die Gütekriterien Schwierigkeit und Trennschärfe nochmals überprüft.

3.4 Vierter Testentwicklungsschritt: Pilotierung des finalen Testungsaufbaus mit Lehramtsstudierenden

Die Daten im Sommersemester 2017 wurden erhoben, um die bisherigen Ergebnisse in der Testentwicklung umzusetzen und das Testequipment für die Präsenzmessungen zu pilotieren. Es wurde versucht die internen Konsistenzen des Vignettentests zu steigern, indem der Test für Messzeitpunkt 1 und Messzeitpunkt 2 aus anderen Vignetten zusammengesetzt wurde als in der Validierungserhebung (siehe Tabelle 25) und das Scoring verändert wurde. Zudem wurden die Testungen nun in Präsenz unter Realbedingungen durchgeführt und während des Vignettentest-Teils Aufnahmen von den Gesichtern der ProbandInnen gemacht. Das Video wurde mit einem entsprechenden Hinweis nach Ausfüllen eines ersten anderen Instrumentes gestartet und auf der Seite nach der letzten Vignette auch mit entsprechendem Hinweis automatisch wieder gestoppt. Das Testungsequipment bestand aus einer selbst entwickelten, montierten Röhrenlampe (Lichtfarbe: kaltweiß), Tablets von Samsung mit Kamera und Android-System, einem Tablet Ständer für die geeignete Aufnahmeposition des Gesichtes sowie Headsets als Kopfhörer, um den Ton der Videoclips verstehen zu können. Mit diesem Aufbau konnten pro Lampe vier Personen so ausgeleuchtet werden, dass die Aufnahmen eine ausreichende Helligkeit besaßen, um von der FaceReader7 Software (Noldus) *gut* analysiert werden zu können. *Gut* bezieht sich hier darauf, dass es der Software bei diesen Aufnahmen möglich ist, das Gesicht der ProbandInnen durchweg zu erkennen, um die Analyse lückenlos leisten zu können (Aufbau: siehe Abbildung 13).



Abbildung 13. Aufbau der Testung in Präsenz ab dem Sommersemester 2017.

Das Programm Face-Reader arbeitet mit Algorithmen, für die in Studien nachgewiesen werden konnte, dass sie eine hohe Trefferquote haben, was das Erkennen von Emotionen betrifft. Laut einer validierenden Studie von Lewinski, den Uyl und Butler (2014) liegt die Trefferquote der Software bei 88% - bei Menschen liegt der Anteil richtig eingeschätzter Emotionen je nach Studie zwischen 60% bis 90%. Dementsprechend kann davon ausgegangen werden, dass die Auswertungen der Software aussagekräftig und mit einer hohen Wahrscheinlichkeit korrekt sind.

Die Testbatterie wurde über LimeSurvey programmiert und auf den Laptop gespielt. LimeSurvey ist ein Open-Source-Umfrage-Tool, das in der Programmiersprache *php* geschrieben ist und digitale Umfragen ermöglicht.

Während der Testung selbst wurde im Raum mit Hilfe eines zusätzlichen Routers ein Netzwerk aufgebaut, über das die Tablets auf die Umfrage zugreifen konnten. Die Fragebogendaten wurden dann als Text-Datei auf dem Laptop gespeichert, die

Videoaufnahmen auf den Tablets. Die Zuordnung von Fragebogendaten und Videoaufnahmen war mit Hilfe der IP-Adresse und des Testungszeitpunktes möglich – so ließ sich der ProbandInnen-Code ermitteln, um die Daten später zusammenführen zu können. Um diese Informationen zur Hand zu haben, wurde während der Testungen ein Protokoll geführt (siehe Anhang C4). Neben Tabletnummern und zugehöriger IP-Adresse gab es hier die Möglichkeit, besondere Vorkommnisse zu vermerken, die unter Umständen Auswirkungen auf die Aufnahmen haben könnten (z.B. laute Geräusche oder ähnliches, bei denen sich die Mimik der ProbandInnen vor Schreck entsprechend verziehen könnte). Nach erfolgter Testung wurden die Videos auf den Rechner hochgeladen und auf einen gesonderten und nur für wenige Personen freigegebenen Bereich des Servers der Abteilung abgelegt. Die Rohdaten der anderen Instrumente wurden auf einem separaten Server gesammelt. Der Name der Ordner der Videos wurde gebildet aus der IP-Adresse, der Routernummer (da bei Testungen mit Gruppengrößen > 20 zwei parallele Netzwerke über unterschiedliche Router aufgebaut wurden, um eine schnelle und verzögerungsfreie Verarbeitungs- bzw. Abspielgeschwindigkeit der Videos gewährleisten zu können) und die Tabletnummer, so dass sich die benötigten Aufnahmen leicht finden ließen. Als Übersicht wurde ein Dokument angefertigt, das die Testungsdaten, die Codes und die Zuordnung der Prä-Post Videodaten ermöglichte.

3.4.1 Stichprobe

Die Messung wurde längsschnittlich an zwei Messzeitpunkten und mit der interessierenden Zielgruppe (Lehramtsstudierenden) durchgeführt. Um möglichst nah am Setting des geplanten Praxis-Einsatzes in SL-Veranstaltungen zu bleiben, wurden die ProbandInnen schon jetzt in drei SL Veranstaltungen (zwei 1-semesterige und eine 2-semesterige) und drei klassischen Seminaren rekrutiert. Die Testungen fanden für die SL-Studierenden an gesonderten Terminen und für die Teilnehmenden aus dem klassischen Seminarformat direkt in den

Seminarsitzungen statt. Alle bekamen Anschreiben per Email mit Erinnerungen für die gewählten bzw. festgelegten Testungstermine.

Die ProbandInnen zum Prä-Messzeitpunkt ($N=70$) waren im Schnitt 25.15 Jahre alt ($SD=4.77$ Jahre) und es nahmen 49 Frauen und 18 Männer teil (3x keine Angabe). Vor Beginn der Testung wurde allen Teilnehmenden transparent gemacht, dass während eines bestimmten Testteils Aufnahmen von ihren Gesichtern angefertigt würden und sie sich gegen eine Teilnahme entscheiden könnten, da ihre Partizipation freiwillig wäre. Es waren alle weiterhin mit der Teilnahme einverstanden. Als Anreiz wurden für alle Teilnehmenden am zweiten Messzeitpunkt fünf 50 €-Gutscheine verlost (wahlweise bei verschiedenen Anbietern, alle GewinnerInnen entschieden sich für Amazon.de). Zum Post-Messzeitpunkt am Ende des Sommersemesters 2017 nahmen nur noch 41 Personen teil (im Schnitt 25.0 Jahre, $SD=5.1$ Jahre), davon 30 Frauen und 10 Männer. Dies entspricht einem Dropout von 41.6%.

Die an beiden Messzeitpunkten teilnehmenden Studierenden studierten vor allem die Lehramtsstudiengänge L2, L3 sowie L5 (siehe Anhang C5) und folgende Fächer (siehe Abbildung 14):

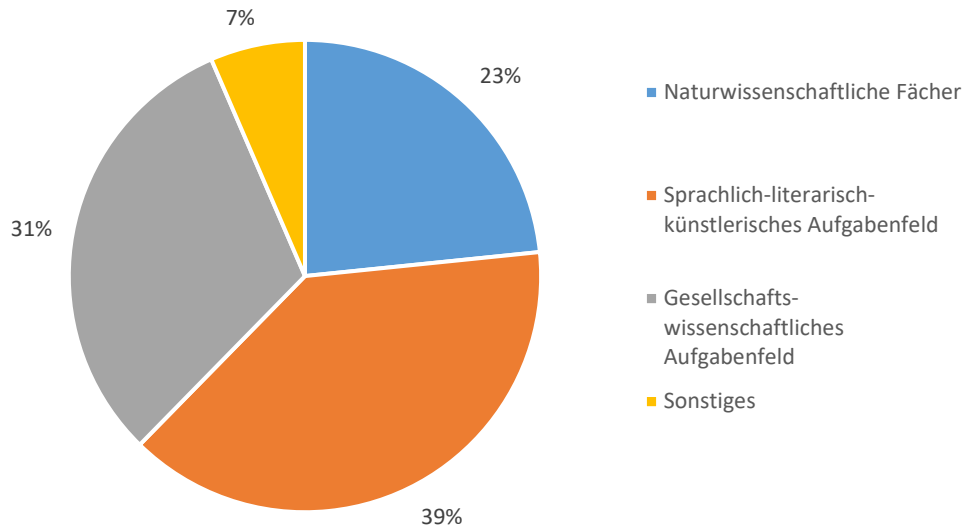


Abbildung 14. Studierte Fächer der teilnehmenden Studierenden der Pilotierung im SoSe17; jeder Studierende nannte mehrere Fächer; Naturwissenschaftliche Fächer: Mathe, Physik, Chemie, Biologie; Sprachlich-literarisch-künstlerisches Aufgabenfeld: Englisch, Deutsch, Kunst, Spanisch, Französisch, Musik; Gesellschaftswissenschaftliches Aufgabenfeld: Geographie/Erdkunde, kath. Religion, ev. Religion, Geschichte, Philosophie/Ethik, Gesellschaftslehre/Politik; Sonstiges: Sport.

Die Studierenden waren zu einem Großteil aus dem 5. oder 6. Semester sowie dem 7. und 8. Semester, verteilten sich allerdings über fast alle Semester hinweg (siehe Anhang C6).

30 Studierende davon besuchten das klassische Seminarformat, 11 Studierende eine SL-Veranstaltung. 27 Studierende gaben an, sich in der Vergangenheit bereits ehrenamtlich engagiert zu haben, 13 verneinten das, einmal gab es keine Antwort. Die verschiedenen ehrenamtlichen Tätigkeiten ließen sich in verschiedene Kategorien clustern und waren divers gestreut (siehe Abbildung 15).

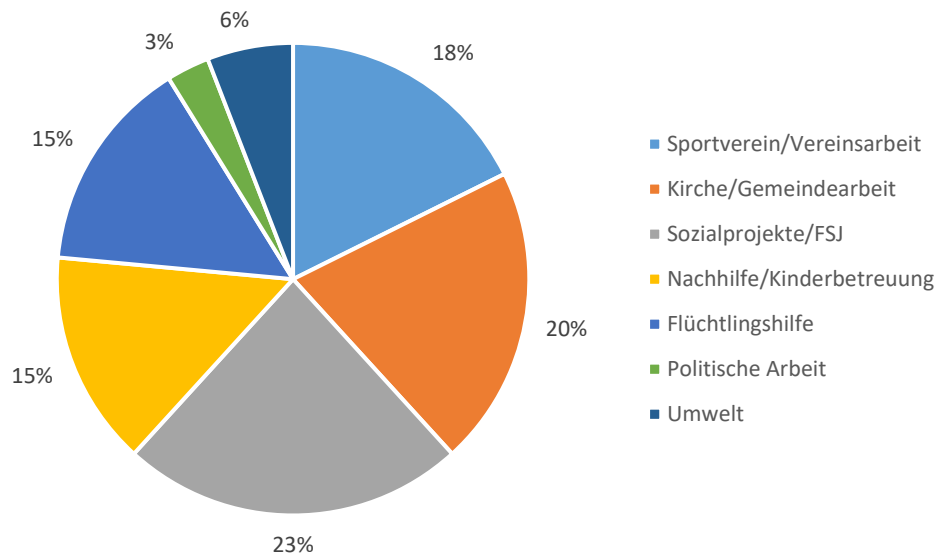


Abbildung 15. Ehrenamtliche Betätigungsfelder (geclustert) der Teilnehmenden der Pilotierung im SoSe17; FSJ=Freiwilliges Soziales Jahr.

Die Face-Reader Daten lagen an beiden Messzeitpunkten bei insgesamt 25 Personen vor. Der hohe Drop-Out kam hauptsächlich durch technische- oder Aufnahmefehler zustande. So gab es bei einigen ProbandInnen keine Videoaufnahme, obwohl die Kamera hätte aufzeichnen müssen ($n=6$), bei starkem Gegenlicht fokussierte die Kamera auf die Lichtquelle im Hintergrund, so dass das Gesicht für eine Analyse zu dunkel war ($n=6$) oder der Aufnahmeradius war ungünstig, so dass in dem erstellten Video Teile des Gesichts (bspw. Stirn oder Kinn) fehlten und die Software die Daten nicht auswerten konnte ($n=5$).

3.4.2 Testbatterie

Die Tests, die die ProbandInnen bearbeiteten, waren zum Teil die gleichen wie bei der vorher erfolgten Validierungsstudie. Neben dem IRI (Davis, 1980), wurde wieder der TEQ (Spreng et al., 2009) und der ESEK (Schönefeld & Roth, 2017) eingesetzt, letzterer allerdings in einer speziell für LehrerInnen adaptierten Empathieerwartungs-Abfrage, die auf den beruflichen Kontext abzielte. Zusätzlich wurden noch zwei Subskalen des *Emotionalen Kompetenz-*

Fragebogen - EKF von Rindermann (2009) in die Testbatterie mit einbezogen: Die Skala zum *Erkennen der Emotionen Anderer (EA)* und zum *Erkennen der eigenen Emotionen (EE)*. Die Skalen des EKF werden im Folgenden genauer beschrieben, alle anderen wurden bereits ausführlich im Kapitel 3.3.2 eingeführt, so dass an dieser Stelle auf erneute Erläuterungen verzichtet wird. Zusätzlich wurden auch wieder demografische Informationen erfragt. Neben Angaben zu Studium (Lehramtsstudiengang, Fächerwahl, Semester) und zur Person (Alter, Geschlecht) wurden auch ehrenamtliches Engagement, Besuch einer SL-Veranstaltung, Schätzung der geleisteten Praxisstunden in SL und Erfahrungen mit Praktika abgefragt, sowie die Rückmeldung, ob im vergangenen Semester über Empathie oder Perspektivenübernahme gesprochen wurde.

3.4.2.1 Emotionale Kompetenz Fragebogen - EKF

Der *emotionale Kompetenz Fragebogen*⁴ (Rindermann, 2009) ist ein Fragebogen zum Erkennen der eigenen und der Gefühle anderer sowie dem Umgang damit und floss als neues Instrument in die Testung ein. Das Instrument bietet sowohl die Möglichkeit der Eigen-, als auch der Fremdeinschätzung und gliedert sich in insgesamt vier Subdimensionen auf: 1. Das Erkennen eigener Emotionen, 2. Erkennen der Gefühle anderer, 3. Regulation eigener Gefühle und 4. der Gefühlsausdruck als emotionale Expressivität. Emotionale Kompetenz wird hier als multidimensionales Konstrukt verstanden, das eine Kombination aus Gefühl und Fähigkeit abbildet und sich somit von bisherigen Fragebögen im Modellbereich der emotionalen Intelligenz unterscheidet (Rindermann, 2009). Zugrunde gelegt wird die Fähigkeit sowohl bei sich als auch bei anderen Emotionen erkennen, verstehen, regulieren und sie ausdrücken zu können. Zudem erlaubt das Konzept auch, dass diese Fähigkeiten veränderbar sind, so dass

⁴ © Hogrefe Verlag GmbH & Co. KG, Göttingen

die Grundannahme der Trainierbarkeit emotionaler Fähigkeiten zu Grunde liegt und dadurch Veränderungsmessungen möglich werden. Die emotionale Kompetenz steht in Zusammenhang mit sozialer Kompetenz, wendet das Hauptaugenmerk allerdings im Gegensatz zu dieser eher auf intrapersonelle Vorgänge - d. h. in der Person selbst findet die Beschäftigung mit dem Verstehen und Erkennen der eigenen und fremden Gefühle, sowie dem Bereich der Regulation eigener Gefühle statt. Zudem wird in dem Instrument nicht zwischen verbalem und non-verbalem Emotionsausdruck unterschieden, wie es in anderen Verfahren zur Erfassung der emotionalen Kompetenz üblich ist. Die Items für das Instrument wurden aus verschiedenen Quellen entnommen (verschiedene Verfahren zur Messung emotionaler Kompetenz und Erlebens), so dass der erste Itempool aus 193 Items bestand und in verschiedenen Entwicklungsschritten auf 62 Items auf vier Dimensionen reduziert werden konnte. Der Entwicklungsprozess des Instrumentes gliederte sich in drei Schritte auf. Zuerst gab es eine Befragung mit Laien, die angeben sollten, welche Merkmale bei Personen zu erkennen sind, die sie als emotional intelligent bezeichnen würden. Die generierten Antworten flossen in die Itementwicklung ein. Im nächsten Schritt wurden zu den extrahierten Verhaltens- und Personenbeschreibungen der Laienbefragung passende, bereits bestehende Fragebögen und Instrumente gesucht und zusätzliche Items entwickelt, um einen ausreichend großen Itempool zu erhalten. Die gewonnenen Items zur Personenbeschreibung wurden einer Stichprobe von 233 Erwachsenen (Großteil: Nicht-Studierende) vorgelegt, die sich anhand einer 5-stufigen Likert Skala von *stimmt überhaupt nicht* (1) bis *stimmt vollständig* (5) selbst einschätzen sollten. Die danach durch Skalen- und Faktorenanalysen, sowie Korrelationen identifizierten Items wurden im nächsten Schritt zum Zwecke der Normierung 638 weiteren Erwachsenen (Frauenanteil von 61% Prozent) vorgelegt. Es wurden 62 Items extrahiert, die in die finale Testversion aufgeteilt in vier Dimensionen einfließen. In den

Normierungsstichproben konnte festgestellt werden, dass Frauen höhere Werte erreichten als Männer.

Für die vorliegende Pilotierungsstudie wurden die Selbsteinschätzungen der Skalen *Erkennen Eigener Emotionen (EE)* und *Erkennen der Emotionen Anderer (EA)* genutzt. In der Dimension EE wird geprüft, inwiefern die ProbandInnen ihre eigenen Gefühle im positiven, als auch im negativen Gefühlsspektrum erkennen, sowie verstehen können. In der Skala EA hingegen geht es darum einzuschätzen wie gut man in der Bewertung der Gefühle und Stimmungen anderer Menschen ist.

Die Skala EE besteht aus 15 Items (davon acht invertierte Items) und EA aus 17 Items (keine invertierten Items). Die internen Konsistenzen werden mit Cronbachs $\alpha=.88$ (EE) bzw. $\alpha=.91$ (EA) berichtet. Sehr ähnliche Werte zeigten sich auch bei den Studierenden dieser Testung.

3.4.3 Datenbereinigung

Zur Datenbereinigung wurden die Daten der beiden Testzeitpunkte in einen gemeinsamen SPSS-Datensatz integriert. Im finalen Datensatz verblieben nur die ProbandInnen, die an beiden Messzeitpunkten teilgenommen hatten ($n=41$). Die aufgezeichneten Videos der ProbandInnen waren etwa 20-30 Minuten lang und beinhalteten als Aufnahme das komplette Lösen des Testvignettenteils. Die Videos wurden anschließend dahingehend bearbeitet, dass sie pro Vignette zugeschnitten wurden und dann nur noch den Ausschnitt zeigten, wenn sich die ProbandInnen einen Videoclip ansahen. Als Marker dienten Screenshots, die den Anfang und das Ende für das Skript zum Schneiden kennzeichneten. Die so erhaltenen Videoclips wurden pro ProbandIn in Face-Reader (Noldus) in einem eigenen Projekt eingelesen, in dem die Aufnahmen aller Vignetten eingefügt wurden. Teilweise schauten die ProbandInnen ein Video mehrfach an, so dass von Face-Reader für diese Videos ein Mittelwert ausgegeben wurde, der in die weiteren Analysen einfluss. Jede Videosequenz wurde mit einem 2-

sekündigem, neutralen Teil des ungeschnittenen Komplett-Videos kalibriert, bei dem sie nach Sichtprüfung einen neutralen Gesichtsausdruck zeigten. In der Regel wurden zwei Sekunden innerhalb der ersten 30 Sekunden der Gesamtaufnahme dafür genutzt. Zu diesem Zeitpunkt lasen die ProbandInnen die Informationsseite, dass ab diesem Zeitpunkt die Aufnahme mitlaufen würde. Die Kalibrierung sorgte dafür, dass die Analysewerte von Face-Reader damit auf eine ProbandInnen-spezifische Baseline heruntergesetzt und geeicht wurden. Durch die helle Lampe und damit einhergehenden muskulären Anspannungen von Brauen und/oder Augen wurden ohne Kalibrierung bspw. häufig höhere Werte in Ärger oder Angst gemessen, was in diesen Fällen zu Verzerrungen der Ergebnisse geführt hätte und mit einer Kalibrierung vermieden werden konnte. Die Analyse-Ergebnisse wurden dann in den SPSS-Datensatz eingefügt.

Wie bereits berichtet, fehlten die Videoanalysen von siebzehn Personen, da es entweder Probleme mit der Aufnahme gab, so dass kein Video produziert wurde oder die Aufnahmen eine schlechte Qualität hatten, so dass sie von der Software nicht analysiert werden konnten.

Wie auch in der vorhergehenden Validierung wurde wieder der ESEK eingesetzt, diesmal allerdings in einer leicht modifizierten Form: *Essener Fragebogen zur Erfassung des subjektiven Empathiekonzepts für LehrerInnen – ESEK-L*. Die Instruktion der Skala zur Empathieerwartung wurden bei diesem Instrument für den Lehramtskontext angepasst, so dass nun die Erwartungen an das eigene empathische Verhalten gegenüber SchülerInnen abgefragt wurden. Die Items selbst waren die gleichen wie beim ESEK (vgl. Kapitel 3.3.2.1). In dieser Version gab es zudem zwei weitere Items, so dass die Skala der funktionalen Empathieerwartung nun aus zehn Items bestand und die dysfunktionale Empathieerwartung aus elf Items. Die neuen Items lauten „Schnell Ratschläge parat haben“, sowie „Auf aufgetauchte Probleme später noch einmal zurückkommen“. Fehlende Mittelwerte im *ESEK-L* wurden wieder durch den Mittelwert des jeweiligen Items ersetzt und dann die Skalen

funktional bzw. dysfunktional gebildet. Ebenfalls wurden die Testscores (Mittelwertberechnung) des TEQ und des EKF erstellt. Die Reliabilitäten entsprachen den erwarteten Normwerten.

3.4.4 Ergebnisse

In dem Versuch die interne Konsistenz anzuheben und höhere Reliabilitätswerte zu erhalten, wurde die Zusammensetzung der Videosequenzen anhand der deskriptiven Ergebnisse der vorherigen Messungen noch einmal angepasst. Für die Testversion zum Testzeitpunkt Prä kamen nun die Vignetten V6, V9, V10, V14, V21 und V30 zum Einsatz, zum Post-Messzeitpunkt die Vignetten V6, V8, V9, V21, V25 und V27. Die Missings der Antworten auf das Item „Intensität der gesehenen bzw. selbst erlebten Emotionen“ wurde durch 0 ersetzt, sofern bei einer der abgefragten Basisemotionen ein Wert >0 eingegeben wurde. Hier wurde wieder davon ausgegangen, dass es sich um einen Auslassungsfehler handelte, da konkrete Angaben gemacht wurden. Beim Ausfüllen der Antworten an den Tablets wurde festgestellt, dass die Antworten beim Item zum empathischen Lehrer-Verhalten teilweise nicht eingeloggt wurden, da es technische Schwierigkeiten beim Einordnen der Rankings gab. Wenn maximal ein Missing vorhanden war, wurde es im Datensatz durch die jeweilige noch fehlende Position ersetzt, wenn es mehrere waren, wurde -77 vergeben.

Alle Vignetten mussten bearbeitet worden sein, um die jeweiligen Testscores bilden zu können. Die durch SPSS-Befehle gebildeten Testwerte wurden händisch nochmal überprüft, um teilweise unentdeckte Missings bzw. Fehler zu korrigieren. Das Scoring selbst wurde angepasst, um die häufig zu leichten Schwierigkeiten anzupassen und den Effekt auf die Trennschärfen zu überprüfen, so dass nun für den Escore pro richtigem Antwortmuster (richtig-richtig und richtig-falsch) jeweils 1 Punkt vergeben wurde und somit insgesamt max. 6 Punkte erreichbar waren. Pro falsch gegebener Antwort (falsch-richtig und falsch-falsch)

wurde 1 Punkt abgezogen. Ebenfalls wurde das Scoring der Verhaltensantworten angepasst und pro richtig eingeschätzter Position 1 Punkt bzw. im Bereich von einer *SD* noch .75 Punkte vergeben. Die Verhaltensantwort von V25 wurde aus den Testwert-Berechnungen ausgeschlossen, da bei der Präsentation des Items fälschlicherweise die Antwortoptionen einer anderen Vignette angegeben wurden und somit eine Auswertung nicht möglich war. Die einzelnen Vignetten-Sequenzen wurden dann mit der Face Reader-Software ausgewertet, um sie mit den Selbstauskunftsdaten zum eigenen Gefühl und der Angabe der Intensität der gesehenen Emotion zu korrelieren. Die Face-Reader Aufnahmen diente dazu, noch einmal zusätzlich die jeweiligen Items anhand der sechs Basisemotionen Freude, Ärger, Angst, Traurigkeit, Überraschung und Ekel zu validieren. Der Auswertungsbereich der Face-Reader Analysen bewegt sich zwischen 0 und 1. Die Selbstauskunft zur eigenen Emotion wurde anhand einer 6-stufigen Likert Skala von *gar nicht* (0) *bis sehr stark* (5) ebenfalls anhand der sechs Basisemotionen erfasst. Dieses Item wurde erst zum Post-Zeitpunkt eingeführt und ersetzte die bis dato per Freitext erhobene Selbstauskunft, deren Auswertung und Implementation sinnvoll nicht möglich war.

Bevor weitere Rechenoperationen angewandt wurden, wurden alle gebildeten Werte mit Hilfe des Kolmogorov-Smirnov-Tests (KS-Test) auf Normalverteilung untersucht. Bei nicht-Erfüllen dieser Vorannahme würde ein non-parametrisches Verfahren zum Einsatz kommen. Da alle Werte normalverteilt zu sein schienen, wurde für die Korrelationsberechnungen Pearsons-Korrelation genutzt.

3.4.4.1 Itemanalyse

Für diese Pilotierung wurde neben der neuen Vignettentest-Zusammenstellung auch Das beschriebene neue Scoring des E- und VScores getestet, um es an der relevanten Zielgruppe

zu überprüfen. Zur besseren Einschätzung wurde auch das alte Scoring berechnet und den deskriptiven Ergebnissen gegenübergestellt.

Entgegen der Erwartung fielen die internen Konsistenzen des Tests geringer aus als bei den Validierungsversionen (siehe Tabelle 26).

Tabelle 26

Interne Konsistenzen des Vignettentests der Pilotierungsstichprobe mit neuem Scoring

Subskala	Eingeflossene Werte	Cronbachs α	
		Stichprobe SoSe17 Prä	Stichprobe SoSe17 Post
Escore	6/6	.24	.24
Iscore	6/6	.57	.57
Vscore	5/6	.51	.51

Anmerkungen. N=41; Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score; Vscore=Verhaltensscore.

Die Werte der Schwierigkeiten des neuen Scorings erstrecken sich zum Prä-Messzeitpunkt von 30.10 bis 88.62 (beim alten Scoring von 10.53 bis 92.68) und zum Post-Messzeitpunkt von 39.06 bis 86.18 (beim alten Scoring von 14.63 bis 87.80; siehe Abbildung 16 und Abbildung 17).

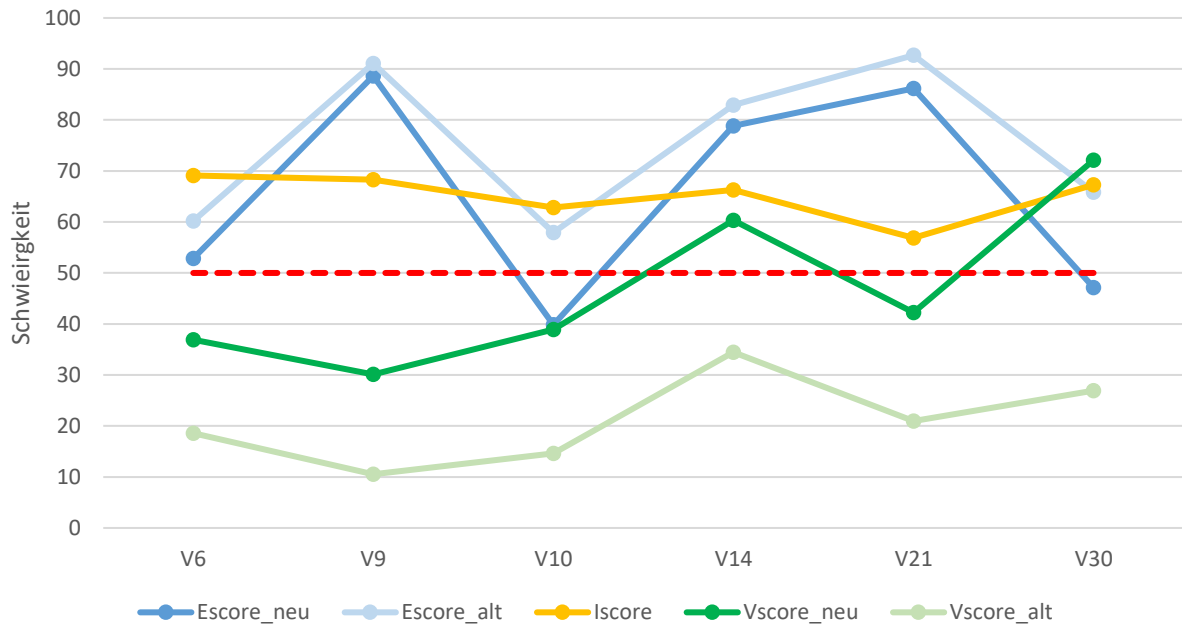


Abbildung 16. Itemschwierigkeiten mit altem und neuem E-, sowie V-Scoring zum Messzeitpunkt Prä; V=Vignette; Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score; Vscore=Verhaltensscore.

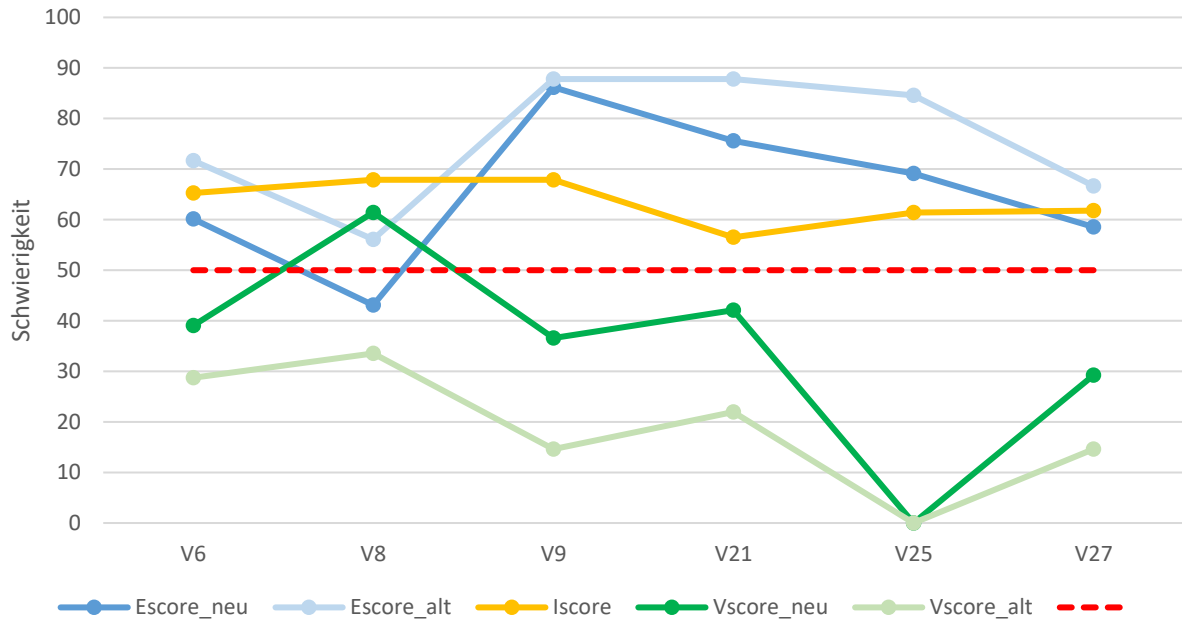


Abbildung 17. Itemschwierigkeiten mit altem und neuem E-, sowie VScoring zu Messzeitpunkt Post; der VScore zu Vignette 25 konnte nicht gerechnet werden und ist deswegen in dieser Grafik auf 0 gesetzt; V=Vignette; Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score; Vscore=Verhaltensscore.

Die Trennschärfen des neuen Scorings bewegen sich mit dem angepassten Scoring zum Messzeitpunkt Prä im Rahmen von .01 bis .44 (im alten Scoring zwischen -.02 und .38). Zum Messzeitpunkt Post verteilen sich die Werte von -.27 bis .30 (mit dem alten Scoring von -.09 bis .40; siehe Abbildung 18 und Abbildung 19).

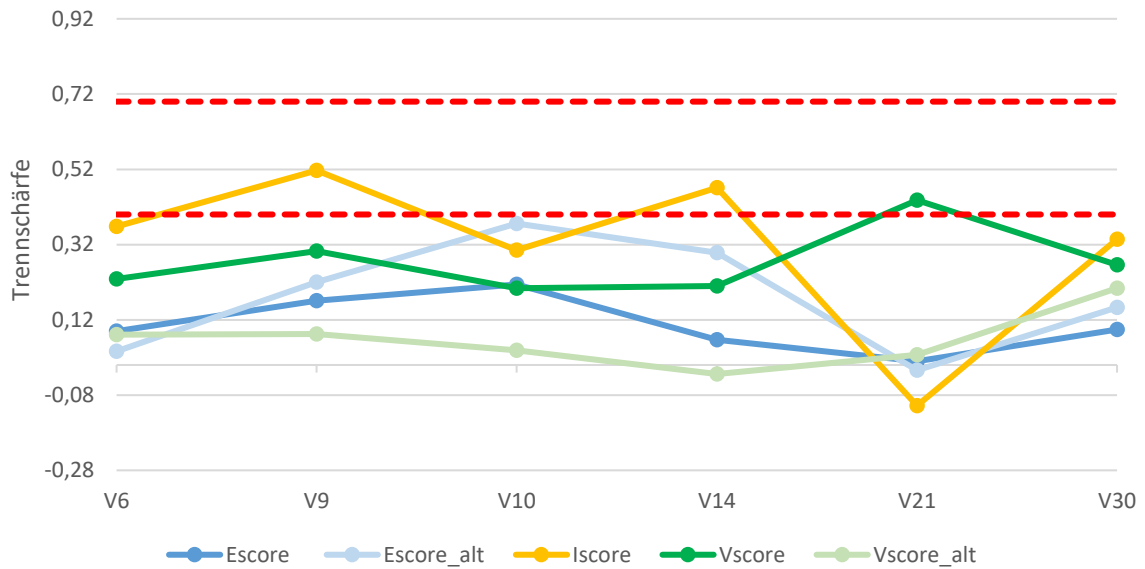


Abbildung 18. Trennschärfen des neuen und alten E-, sowie V-Scoring im Vergleich (Pilotierung SoSe17: Prä); V=Vignette; Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score; Vscore=Verhaltensscore.

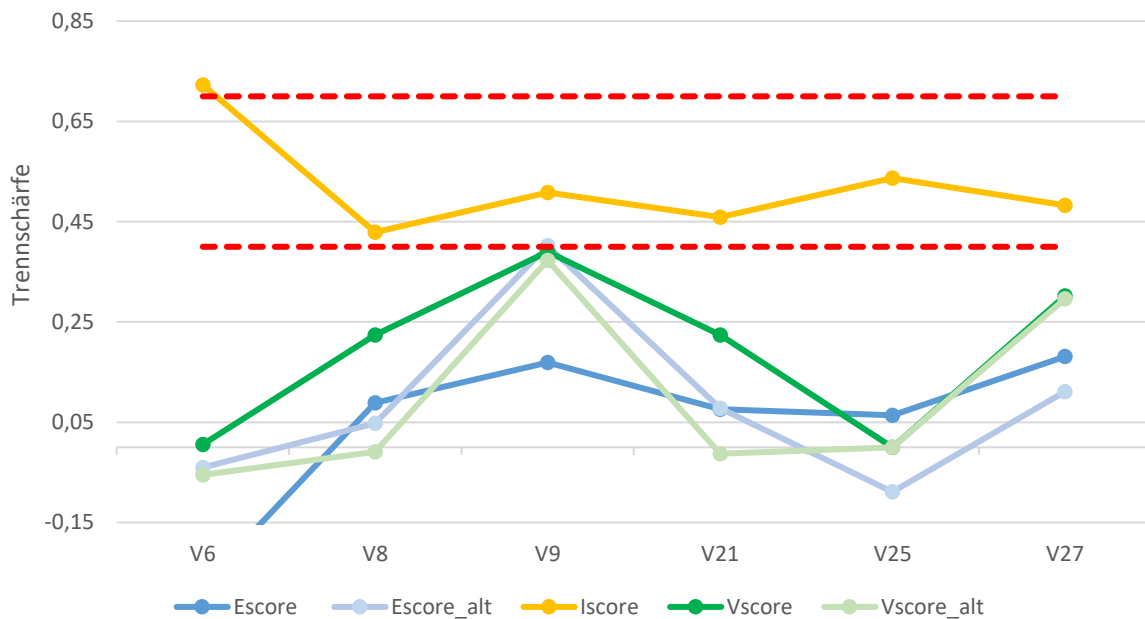


Abbildung 19. Trennschärpen des neuen und alten E-, sowie VScoring im Vergleich (Pilotierung SoSe17: Post); Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score; Vscore=Verhaltensscore.

Die Werte des Iscores liegen weitestgehend in einem mittleren Bereich (sowohl Schwierigkeit, als auch Trennschärfe betreffend).

3.4.4.2 Korrelationen mit anderen Testmaßen

Auch in dieser Pilotierung wurden die erhobenen Daten zum Vignettentest mit den zusätzlich erhobenen Testwerten korreliert, um die Zusammenhänge zu den interessierenden Konstrukten bei der Zielgruppe zu überprüfen. Hierfür wurden die Korrelationen nach Pearson berechnet. Die Ergebnisse sind in Tabelle 27 und Tabelle 28 für alle Werte dargestellt, bei denen signifikante Zusammenhänge festgestellt werden konnten. Die Werte bewegen sich zwischen -.54 bis .47.

Tabelle 27
 Korrelationen des Vignettentests Prä mit allen anderen Testwerten

	Escore			Iscore				Vscore			
	V6	V14	Testwert	V6	V9	V10	V14	Testwert	V6	V10	Testwert
ESEK funkt		.37*									
ESEK dysfunkt	.43**		.33*		-.38				-.34*	-.34*	-.37*
EKF_ EE				.35*		.46**	.39*	.47**			
EKF_ EA				.33*							

Anmerkungen. * $p < .05$, ** $p < .01$; $N=41$; Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score; Vscore=Verhaltensscore; ESEKfunkt=ESEK, Subskala funktional; ESEKdysfunkt=ESEK, Subskala dysfunktional; EKF_EE=EKF, Subskala *Erkennen der eigenen Emotionen*; EKF_EA=EKF, Subskala *Erkennen der Emotionen Anderer*; V=Vignette.

Tabelle 28
 Korrelationen des Vignettentests Post mit allen anderen Testwerten

	Escore			Iscore				Vscore		
	V6	V8	V27	Testwert	V6	V8	V9	V27	Testwert	V6
ESEK dysfunkt		-.54**		-.31*		-.38*			-.33*	-.34*
TEQ	.31*				.41**		.33*	.32*		
EKF_ EE							.44*		.37*	
EKF_ EA				-.42**			.46**			

Anmerkungen. * $p < .05$, ** $p < .01$; $N=41$; Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score; Vscore=Verhaltensscore; ESEKdysfunkt=ESEK, Subskala dysfunktional; EKF_EE=EKF, Subskala *Erkennen der eigenen Emotionen*; EKF_EA=EKF, Subskala *Erkennen der Emotionen Anderer*; V=Vignette.

3.4.4.3 Mitschwingen

Zusätzlich zu den Selbsteinschätzungen der ProbandInnen zu den eigenen Gefühlen beim Anschauen wurden auch Aufnahmen ihrer Gesichter gemacht und diese Aufnahmen mit einer Face Reader-Software (Noldus) analysiert. Die Ergebnisse der Face-Reader Analysen wurden sowohl mit der eingeschätzten Intensität in dem angesehenen Video, als auch mit der Selbstauskunft der eigenen Gefühle beim Anschauen dieses Videos korreliert. Da die Frage nach den eigenen Gefühlen erst zum Post-Zeitpunkt in eine Ratingskala konvertiert wurde, werden hier nur die Ergebnisse dieses Messzeitpunktes berichtet.

Besonders beim Gefühl *Freude* finden sich bei fünf Vignetten Zusammenhänge zwischen .57 bis .91 (sowohl beim selbsteingeschätzten Gefühl, als auch bei der Intensitätseinschätzung des Gefühls im Videoclip). Die Korrelationen der anderen Gefühlsqualitäten sind uneinheitlich. Es finden sich Zusammenhänge zwischen bei sich selbst eingeschätzter *Angst* und der Face-Reader-Analyse *Freude* mit .59 (V6), zwischen im Videoclip eingeschätzter *Traurigkeit* und der Face-Reader-Analyse *Überraschung* mit .63 (V8), zwischen im Videoclip eingeschätzter *Überraschung/Neugier* und der Face-Reader-Analyse von *Ekel* mit .56 (V8), bei sich selbst eingeschätztem *Ärger* und der Face-Reader-Analyse von *Traurigkeit* mit .42 (V8). Ebenfalls gab es Korrelationen zwischen dem im Videoclip eingeschätzten *Ärger* und der Face-Reader-Analyse von *Ekel* mit .49 (V9), zwischen im Videoclip eingeschätzter *Überraschung/Neugier* und der Face-Reader-Analyse *Freude* mit .66 (V25), sowie zwischen der bei sich selbst eingeschätzten *Überraschung/Neugier* und der Face-Reader-Analyse *Freude* mit .56 (V25). Des Weiteren korrelierten in V27 die selbst eingeschätzte *Freude* mit der Face-Reader-Analyse von *Ekel* mit .47 und der selbst eingeschätzte *Ekel* mit der Face-Reader-Analyse von *Ärger* mit .56. Negative Zusammenhänge waren zu finden bei der selbst eingeschätzten *Freude* mit der Face-Reader-Analyse von *Traurigkeit* in Höhe von -.48 (V8), bei dem selbst eingeschätzten *Ärger* mit der Face-Reader-Analyse von *Freude* mit -.48 (V9), ebenfalls bei

dem selbst eingeschätzten *Ärger* mit der Face-Reader-Analyse von *Freude* mit $-.43$ (V21),
sowie zwischen der im Videoclip eingeschätzten *Angst* und der Face-Reader-Analyse von
Ärger (V25; Übersicht siehe Anhang C7).

3.4.5 Auswertung der Pilotierung und Resümee der Testentwicklung

Die Gütekriterien der Vignettentestversionen haben sich in der Pilotierungstestung im Vergleich zur Validierungstestung leicht verändert. Die neue Zusammensetzung der Testversionen sollte einen Anstieg der Reliabilität bewirken, was in dieser Testung nicht nachgewiesen werden konnte. In der zuvor erfolgten Validierungstestung lagen die internen Konsistenzen in beiden Testversionen zwischen $\alpha=.53$ (Vscore, Test-Version 1) und $\alpha=.76$ (Iscore, Test-Version 2), in der nun durchgeführten Pilotierung wurden Werte zwischen $\alpha=.10$ (Escore, Test-Version Prä) und $\alpha=.77$ (Iscore, Test-Version) erreicht. Ein großer Unterschied, der einen Einfluss auf dieses Ergebnis gehabt haben könnte, lag bei beiden Testungen in der Zielgruppe. Während bei der Validierung verschiedene Studiengänge an der Untersuchung teilnahmen, waren es in der Pilotierung nur Lehramtsstudierende. Schaut man sich bei der Validierungsstudie nur die internen Konsistenzen der Lehramtsstudierenden an ($n=10$), bekommt man auch hier ähnliche Werte zwischen $\alpha=.27$ (Vscore, Test-Version 1) und $\alpha=.85$ (Escore, Test-Version 2) - ein Range, der sich entsprechend auch in der Pilotierung zeigte. Da die Zusammensetzung der Vignetten in der Validierung zumindest eine höhere Reliabilität aufwies, wurden die Vignetten-Konstellationen für die Hauptstudie auf die Validierungsabfolge zurückgesetzt (siehe Kapitel 3.2.5).

In der Pilotierung wurde für den E- und Iscore ein neues Scoring getestet. Es zeigte sich, dass das Scoring für bessere Werte bei Schwierigkeit und Trennschärfe sorgte und wird für die weiteren Testungen beibehalten. Die Schwierigkeitswerte lagen nun mit Werten zwischen 30.10 und 88.62 etwas näher am präferierten mittleren Bereich, in dem die größte Differenzierung zwischen Personen unterschiedlicher Merkmalsausprägung möglich wird (Moosbrugger & Kelava, 2007). Das Scoring des Iscore wurde beibehalten und zeigte auch diesmal wieder gute Werte im mittleren Bereich. Auch die Trennschärfe konnte leicht

verbessert werden, lag allerdings auch jetzt mit Werten zwischen $-.27$ bis $.44$ bei den meisten Vignetten deutlich unter dem angestrebten Bereich von $.40$ bis $.70$.

Die Kodierungen der verschiedenen Test-Scores werden so beibehalten und auch das Vorgehen selbst, die ExpertInnen-Antworten als Auswertungsgrundlage zu nutzen, ist ein Prozedere, das auch in anderen videobasierten Verfahren wie bspw. in der Observer-Studie⁵ angewandt wurde (Seidel & Prenzel, 2007; Seidel et al., 2010).

Die gefundenen Zusammenhänge der Vignettentest-Ergebnisse mit den eingesetzten Empathie- und Emotionsmessverfahren bewegten sich teilweise (Vignetten 6, 8, 9, 10, 14 und 27) im mittleren bis hohen Bereich, so dass davon ausgegangen werden kann, dass diese Vignetten Merkmale von Empathie und Emotionserkennung bei sich und anderen mit abbilden. Mit allen anderen Vignetten konnte in dieser Stichprobe kein signifikanter Zusammenhang festgestellt werden. Die Korrelationen mit dem Empathie-Messverfahren TEQ waren in dieser Stichprobe nur mit wenigen Vignettenwerten vorhanden. Eine mögliche Erklärung könnte sein, dass der TEQ Trait-Anteile von Empathie misst (Spreng et al., 2009), der Vignettentest allerdings versucht State-Anteile abzubilden. Somit beschreibt der TEQ stabile Empathieanteile, was die kleinen bis mittleren Zusammenhänge mit den Vignetten erklären würde (McCrae, Kurtz, Yamagata & Terracciano, 2011). Außerdem zeigten verschiedene Studien, dass kein Zusammenhang zwischen Selbstauskunftsfragebögen und Testungen der Performanz gezeigt werden konnte (Richter, 2009). Dies betraf bspw. Testungen zur empathischen Akkuratheit (Decety & Ickes, 2011), so dass denkbar ist, diesen Effekt auch hier vorliegen zu haben.

Die Auswertung der Face-Reader-Analysen zeigt, dass es zwischen der Selbsteinschätzung der selbst-empfundene Emotionen beim Anschauen der Videoclips, sowie der eingeschätzten

⁵ Entwicklung eines videobasierten Verfahrens zur Erfassung von professionellen Lehrerkompetenzen

Emotionsintensität der Personen in den Clips und den objektiveren Face-Reader-Analysen signifikante Zusammenhänge gibt. Das betrifft vor allem die Emotion *Freude*, die mittel bis hoch mit den anderen Daten korreliert. Das betrifft sowohl selbst getätigte Einschätzungen von Freude, als auch die Diskrimination und den erwartungskonformen, negativen Zusammenhang zwischen Gefühlen wie Ärger oder Traurigkeit. Bei den anderen Emotionen sind die Ergebnisse uneinheitlich. Vor allem *Überraschung/Neugier* korreliert einige Male mit Freude. Beim Blick in die Daten zeigt sich, dass in diesen Fällen kaum *Freude* gemessen werden konnte (niedriger Wert), ebenso wie bei der Angabe von *Überraschung/Neugier*. Dadurch erklärt sich der Zusammenhang dieser Ergebnisse.

Des Weiteren wurden bei der Pilotierung einige technische Schwierigkeiten offensichtlich, die für die nächsten Testungen behoben werden. Es stellte sich bspw. heraus, dass es technische Schwierigkeiten bei der Verhaltensfrage gab: wenn die Teilnehmenden die Antworten sortierten und auf die rechte Seite zogen, änderte sich bei manchen Tablets die Reihenfolge. Es konnte eruiert werden, dass sich das Problem beheben ließ, sobald die Antworten am Tablet über oben auf die andere Seite gezogen wurden. Das Vorgehen wurde in die Instruktion aufgenommen. Auch die Problematik der überbelichteten Aufnahmen wurde behoben, indem die Tische nur noch seitlich zu Fenstern gestellt wurden, um Gegenlicht zu vermeiden. Sofern möglich wurden im weiteren Verlauf auch alle Fenster verdunkelt und die Deckenbeleuchtung voll aufgedreht, um die Videoaufnahmen zu optimieren.

Kapitel 4: Praxiseinsatz des entwickelten Tests

4.1 Testung in SL-Veranstaltungen

Zur Untersuchung der Fragestellung, ob Empathie durch SL entwickelt werden kann, wurden ProbandInnen in insgesamt 17 Veranstaltungen rekrutiert. Die Veranstaltungen fanden im oder ab dem Wintersemester 2017/18 an der Goethe Universität Frankfurt statt. Zehn Veranstaltungen davon waren SL-Veranstaltungen (siehe Anhang D1), fünf waren konventionelle Seminare im Lehramtsstudium. In drei SL-Veranstaltungen war es nur möglich eine schriftliche Bewerbung der Studie durch die Dozierenden an die Studierenden weiterleiten zu lassen. In weiteren 14 Veranstaltungen wurde in den ersten beiden Semesterwochen die Studie in den Veranstaltungen selbst beworben und InteressentInnen zur Teilnahme animiert. Bei der Anwerbung bekamen die Studierenden einen kurzen Überblick über die Fragestellung und die Testung, sowie Informationen über die Belohnung von 30€ bei der Teilnahme an allen drei Testungsterminen.

4.1.1 Stichprobe

Zum ersten Messzeitpunkt nahmen 110 Studierende teil, beim zweiten Messzeitpunkt 72 Studierende und beim dritten Messzeitpunkt nochmals 72 Studierende (allerdings nicht die exakt gleichen Personen wie zu Messzeitpunkt zwei). Das entspricht einem Dropout von 35.5%. Final flossen die Daten von 68 Studierenden in die Analysen mit ein, deren Daten zu allen drei Messzeitpunkten vorhanden waren. 30 Personen davon besuchten SL-Seminare und wurden der Experimentalgruppe (EG) zugeordnet, 38 Personen besuchten kein SL-Seminar und wurden der Kontrollgruppe (KG) zugeteilt. Als SL-Besuch wurde die Teilnahme an einer einsemestrigen oder zweisemestrigen SL-Veranstaltung gewertet. Das Alter der ProbandInnen in der EG lag durchschnittlich bei 24.82 Jahren ($SD=4.90$ Jahre) und in der KG bei durchschnittlich 24.37 Jahren ($SD=4.70$ Jahre). Unter den Teilnehmenden waren in der EG 19 Frauen und 11 Männer und in der KG 29 Frauen und 9 Männer. Unter den ProbandInnen

waren alle Lehramts-Formen vertreten, die an der Goethe-Universität angeboten werden (siehe Abbildung 20).

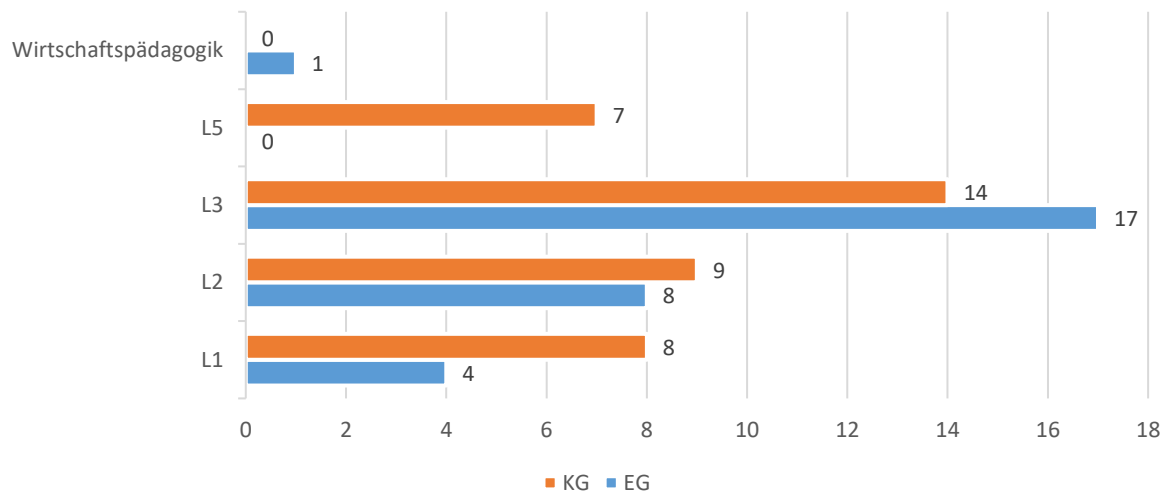


Abbildung 20. Verteilung der studierten Lehrformen der ProbandInnen; L1=Lehramt an Grundschulen, L2=Lehramt an Real- und Hauptschulen, L3=Lehramt an Gymnasien, L5=Lehramt an Förderschulen; KG=Studierende aus Non-SL-Veranstaltungen, EG=Studierende in SL-Veranstaltungen.

Auch die Fachsemesterzahl war heterogen und Semester eins bis dreizehn Semester vertreten (siehe Anhang D2). Die ProbandInnen studierten meistens zwei oder mehr Fächer. Eine Übersicht über die kategorisierte Fächereinteilung ist in Abbildung 21 dargestellt.

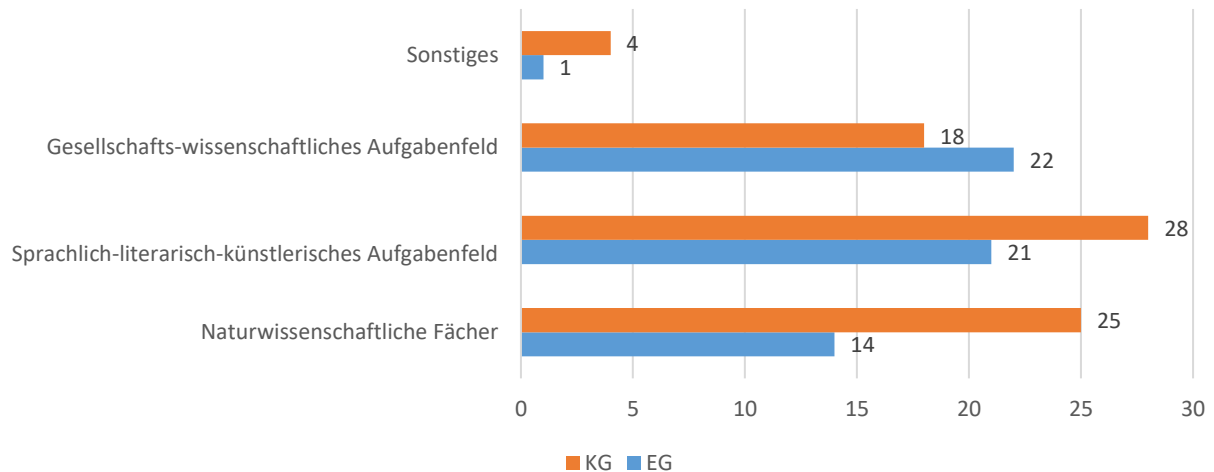


Abbildung 21. Studierte Fächer der ProbandInnen; Naturwissenschaftliche Fächer: Mathe, Physik, Informatik, Chemie Biologie; Sprachlich-literarisch-künstlerisches Aufgabenfeld :Englisch, Deutsch, Kunst, Französisch, Musik, Spanisch, Italienisch; Gesellschafts-wissenschaftliches Aufgabenfeld: Geographie/Erkunde, Kath. Religion, Geschichte, Philosophie, Islamische Religion, ev. Religion; Sonstiges: Sachunterricht, Sport, Emotionale soziale Entwicklung/EH, BWL, Lernhilfe/LH, Praktisch bildbar.

Zwischen EG und KG gab es in fast allen demografischen Variablen keine signifikanten Unterschiede, lediglich bezüglich der studierten Lehrform gab es den Unterschied, dass L5-Lehramts-Studierende nur in der Kontrollgruppe teilnahmen. Zudem wurde das Thema *Perspektivenübernahme* über das WiSe hinweg signifikant häufiger in der Experimentalgruppe angesprochen (Selbstauskunft der Studierenden; Mann-Whitney-U-Test; siehe Anhang D3). Die Ausgangswerte des TEQ und die Ergebnisse des Persönlichkeitsfragebogens HEXACO-PI-R waren in beiden Gruppen annähernd gleich verteilt, so dass zu Beginn keine auffallenden Gruppenunterschiede festzustellen waren.

4.1.2 Durchführung

Als Design der Studie wurde ein längsschnittliches Kontrollgruppen-Design mit drei Messzeitpunkten gewählt (siehe Abbildung 22).

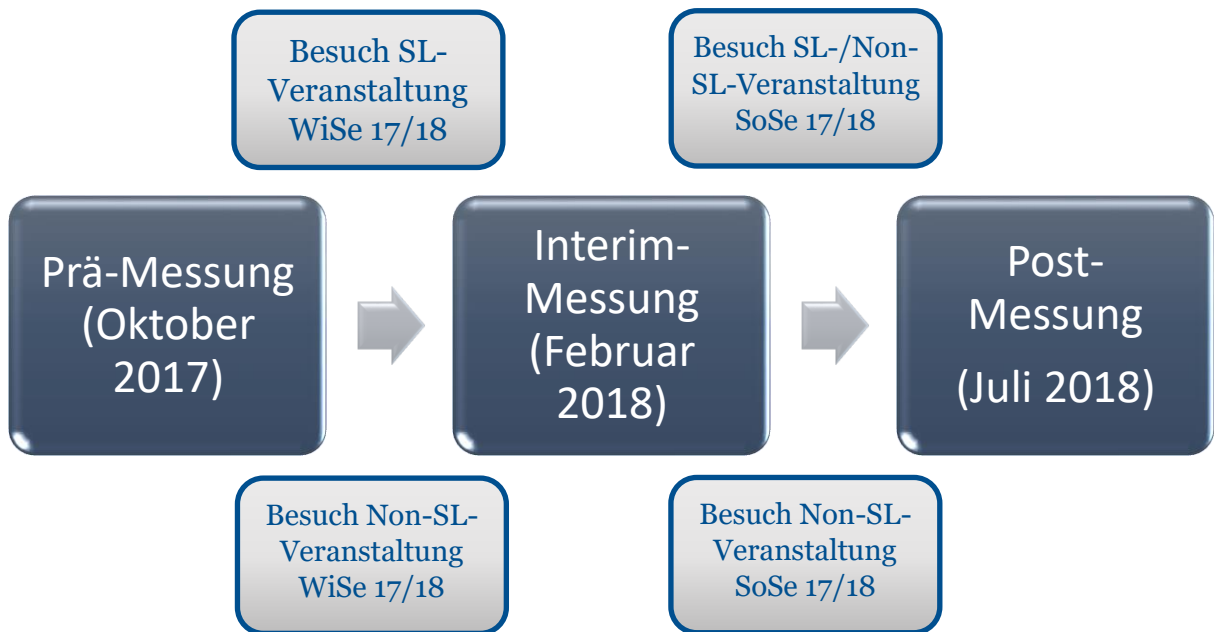


Abbildung 22. Erhebungsdesign des Praxiseinsatzes.

Die Studierenden wurden vor Beginn der Testung ausführlich instruiert und bekamen die Informationen über die Videoaufnahme ihres Gesichts und den Hinweis auf Freiwilligkeit ihrer Teilnahme beim ersten Testungstermin (siehe Anhang D4). Ebenfalls wurde ihnen erläutert, was mit ihren Daten geschehen und dass Filmaufnahmen sowie Testdaten getrennt auf einem geschützten abteilungsinternen Server gespeichert würden. Alle Teilnehmenden waren mit den Bedingungen einverstanden und sich der Freiwilligkeit ihrer Teilnahme bewusst. Zudem wurden Sie darauf hingewiesen während des Vignettentest-Teiles das Tablet im Ständer zu belassen, während der Videoaufnahmen keinen Kaugummi zu kauen, nichts zu essen oder zu trinken und darauf zu achten die Hand nicht ins Gesicht zu nehmen (zum Aufstützen o.ä.). Die Teilnehmenden wurden über die Incentives informiert (30€ am Ende der dritten Testungssitzung) und es gab am Ende eine Unterschriftenliste, die von jedem Studierenden nach erfolgter Teilnahme abgezeichnet wurde. Hierbei blieb es den Studierenden freigestellt, ob sie den Test ernsthaft bearbeiteten oder die Seiten ohne Bearbeitung nur bis zum Ende durchklickten, was möglich war. Die Testungen wurden wieder in einem

Gruppensetting durchgeführt, so dass zeitgleich an jedem Testungstermin 34 Personen getestet werden konnten. Die Tische wurden seitlich zum Fenster aufgestellt, so dass starkes Licht im Rücken der ProbandInnen ausgeschlossen werden konnte und wenn möglich die Fenster verdunkelt. Für Messzeitpunkt zwei und drei wurden die ProbandInnen per Email angeschrieben und gebeten sich für einen von mehr als 20 möglichen Testterminen in den letzten beiden Semesterwochen einzutragen bzw. eine Rückmeldung zu geben, um einen individuellen, passenderen Termin zu vereinbaren. Die Testung selbst bestand wieder aus einer Testbatterie mit verschiedenen Verfahren und sechs Vignetten aus dem Vignettentest. Empathie wurde in der Testbatterie durch verschiedene Variablen operationalisiert. Da es für Lehrkräfte nicht nur darauf ankommt den empathischen Prozess zu durchlaufen, sondern auch empathisch zu agieren, wurden auch Verhaltenskomponenten mit erhoben, die auf professionelles Wissen und empathisches Verhalten abzielen (vgl. Abbildung 23 und Abbildung 24).

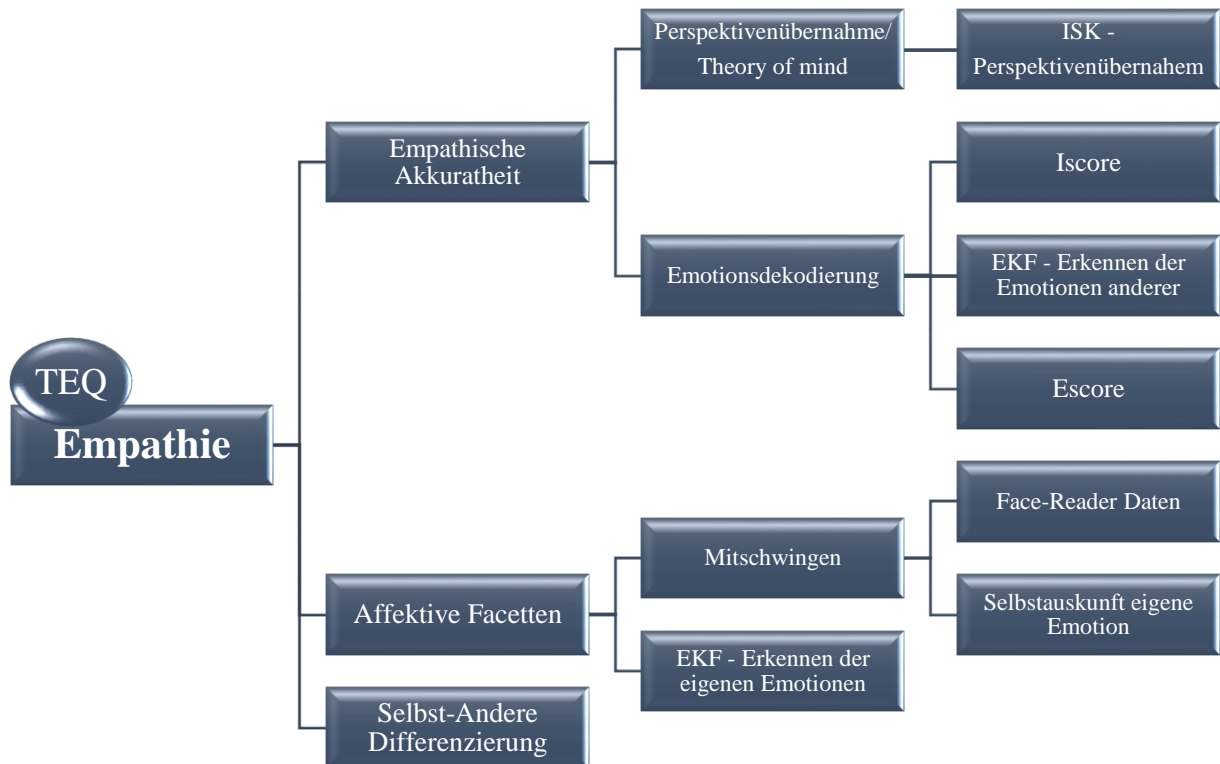


Abbildung 23. Operationalisierung von Empathie bei angehenden Lehrkräften beim Praxiseinsatz des Vignettentests; die AVs sind in den Kästen ganz rechts aufgelistet.

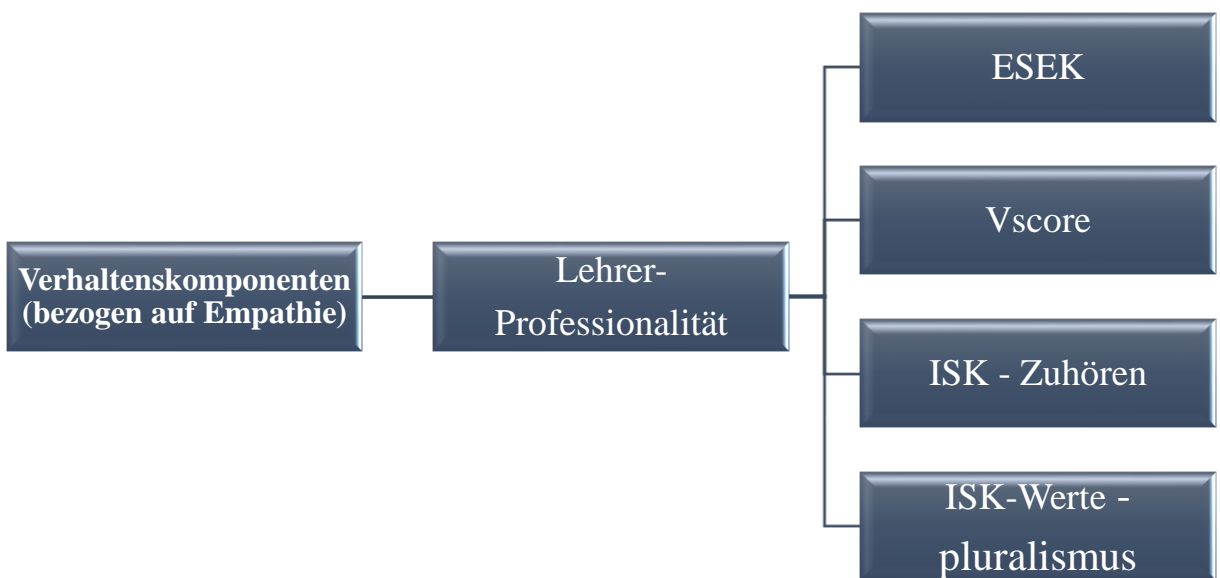


Abbildung 24. Operationalisierung empathischen Handelns bei angehenden Lehrkräften beim Praxiseinsatz des Vignettentests; die AVs sind in den Kästen ganz rechts aufgelistet.

Zum Erfassen empathischen Handelns wurden der *ESEK-L* (Skala funktional $\alpha=.61$, $n=68$; Skala dysfunktional $\alpha=.74$, $n=68$) und Sekundärfaktoren der Subdimension *Soziale Orientierung* des *Inventar Sozialer Kompetenzen* (Kanning, 2009) vorgelegt. Zur Messung von Empathie wurden zur allgemeinen Einschätzung der *TEQ* ($\alpha=.67$, $n=66$) und zum Erfassen der einzelnen Facetten die Subdimensionen *Erkennen der Emotionen anderer* ($\alpha=.92$, $n=64$), sowie *Erkennen der eigenen Emotionen* ($\alpha=.92$, $n=57$) des *EKF* zur Bearbeitung gegeben. Zusätzlich wurde als Persönlichkeitsfragebogen noch der *HEXACO-PI-R* (Vries, 2013) vorgelegt. Die meisten der benutzten Verfahren wurden bereits an früherer Stelle ausführlich vorgestellt (siehe Kapitel 3.4), die neu hinzu gekommenen Verfahren werden im Folgenden erläutert.

4.1.2.1 HEXACO-PI-R-24

Der *HEXACO-PI-R-24* (Vries, 2013) ist ein Selbstauskunftsfragebogen zum Erfassen verschiedener Facetten der Persönlichkeit bestehend aus 24 Items. Die Items werden auf einer Skala von *stimme überhaupt nicht zu* (1) bis *stimme voll zu* (6) eingeschätzt. Er basiert auf dem *HEXACO-Modell* mit sechs Persönlichkeitseigenschaften, das sich in den letzten Jahren mehr und mehr etabliert hat und das fünf-Faktoren Modell der Persönlichkeit erweitert, auf dem beispielsweise auch der *BFI* beruht (Moshagen, Hilbig & Zettler, 2014; Rammstedt & Danner, 2016; siehe Kapitel 3.3.2.5). Die computerisierte Technik machte es möglich den lexikalischen Ansatz des Fünf-Faktoren-Modells nicht nur sprach- und kulturspezifisch, sondern über diese Begrenzung hinaus über eine Vielzahl verschiedener Sprachen hinweg auszuwerten. Dadurch ergab sich ein passenderes Modell mit sechs Faktoren (Moshagen et al., 2014). Die sich herauskristallisierenden Faktoren konnten als *Ehrlichkeit-Bescheidenheit*, *Emotionalität*, *Extraversion*, *Verträglichkeit vs. Ärger*, *Gewissenhaftigkeit* und *Offenheit für neue Erfahrungen* identifiziert werden. Sie sind teilweise deckungsgleich mit denen des Fünf-

Faktoren-Modells, unterscheiden sich allerdings auch in manchen Fällen. So gibt es Unterschiede in der Untergliederung der Faktoren *Emotionalität* (bzw. *Neurotizismus*) und *Verträglichkeit*, bei denen verschiedene Items jeweils dem anderen Faktor zugeordnet werden (Moshagen et al., 2014). Der zusätzliche Faktor *Ehrlichkeit-Bescheidenheit* zeigte mit verschiedenen anderen Bereichen wie Delinquenz, sozialen und politischen Einstellungen, sowie Kooperation hohe Zusammenhänge und sorgte für mehr Aufklärung als das vorherige Modell, so dass er aufgenommen wurde (vgl. Vries, 2013). Die Lang-Version des HEXACO-Persönlichkeitsfragebogens HEXACO-192 (Lee & Ashton, 2004) besteht aus 192 Items mit jeweils 32 Items pro Eigenschaft. Eine kürzere Version bildet der HEXACO-96 mit jeweils 16 Items auf jedem Faktor, sowie der HEXACO-60, bei dem es 10 Items pro Eigenschaft gibt (Ashton & Lee, 2009). Die psychometrischen Eigenschaften der Selbstauskunftsfragebögen liegen alle in einem annehmbaren Bereich und die internen Konsistenzen werden selbst in der HEXACO-60 Version zwischen $\alpha=.77$ und $\alpha=.80$ berichtet (deutsche Version: zwischen $\alpha=.74$ bis $\alpha=.83$), was für Persönlichkeitstests in einem zufriedenstellenden Bereich liegt (Moosbrugger & Kelava, 2007; Moshagen et al., 2014; Schmidt-Atzert & Amelang, 2012). Die hier verwendete Version des HEXACO-24 ist mit 24 Items noch kürzer und innerhalb weniger Minuten bearbeitbar, so dass er aufgrund ökonomischer Gesichtspunkte ausgewählt wurde, um die Testungsdauer unter 60 Minuten halten zu können. Die berichteten internen Konsistenzen der niederländischen Version sind zwar mit Werten zwischen $\alpha=.55$ und $\alpha=.69$ niedriger als die empfohlenen Richtwerte für Persönlichkeitsfragebögen, haben allerdings recht hohe Retest-Reliabilitätswerte und hohe Konvergenzwerte mit der Langform und Kurzform des HEXACO-PI-R (Vries, 2013). Vries bezieht sich auf McCrae et al. (2011) und argumentiert, dass hohe interne Konsistenzen nicht zwangsläufig etwas mit hoher Validität zu tun haben müssten, so dass Stabilität des Merkmals und hohe Zusammenhänge mit ähnlichen Konstrukten mehr Bedeutung zukäme, so dass auch die Kurzversion mit 24 Items vor allem

in umfangreichen Testbatterien eine vertretbare Alternative darstelle. In der hier genutzten Version liegen die Reliabilitäten deutlich niedriger als die berichteten zwischen $\alpha=.30$ (*Verträglichkeit vs. Ärger*) und $\alpha=.55$ (*Extraversion*), so dass aufgrund der niedrigen Reliabilität und dem damit einhergehenden größeren Messfehler die Daten von weiteren Auswertungen ausgeschlossen werden.

4.1.2.2 Inventar Sozialer Kompetenzen – ISK

Als weiteres Verfahren kam die Subskala *Soziale Orientierung* des *Inventars sozialer Kompetenzen-ISK*⁶ (Kanning, 2009) zum Einsatz. Soziale Kompetenz wird bei diesem Verfahren als ein multidimensionales Merkmal angesehen und es wird zwischen sozialen Kompetenzen und sozial kompetentem Verhalten unterschieden. Ersteres bezeichnet die vorhandenen Anlagen oder das Potential in der Person und letzteres den Ausdruck dieses Potentials im tatsächlichen Handeln. Das Instrument ist ein Selbstauskunftsverfahren mit 108 Items bestehend aus vier Primärfaktoren (Soziale Orientierung, Offensivität, Selbststeuerung, Reflexivität) und insgesamt 17 Sekundärfaktoren (vier bis fünf Faktoren pro Primärfaktor). Die Likert-Skala reicht von *trifft gar nicht zu* (1) bis *trifft sehr zu* (4). Da in der vorliegenden Studie nur einige Sekundärfaktoren des Primärfaktors *Soziale Orientierung (SO)* zum Einsatz kamen, werden nur diese ausführlicher beschrieben. Die Skala SO besteht aus insgesamt 32 Items mit den fünf Subdimensionen *Prosozialität (PS)*, *Perspektivenübernahme (PÜ)*, *Wertepluralismus (WP)*, *Kompromissbereitschaft (KO)* und *Zuhören (ZU)*. Eingeflossen in die Studie sind die Subdimensionen PÜ, WP und ZU. Die Primärskala im Ganzen gibt Auskunft darüber wie offen und wohlwollend sich eine Person anderen gegenüber verhält. Perspektivenübernahme gibt den Grad der Fähigkeit sich in die Lebenswelt eines anderen

⁶ © Hogrefe Verlag GmbH & Co. KG, Göttingen

hineinzusetzen an. Sie besteht aus sechs Items, zwei Items in invertierter Form. Wertpluralismus erfasst die Ausprägung an Offenheit andere Personen so zu akzeptieren wie sie sind (bezogen auf Werte und Einstellungen). Hier liegen ebenfalls sieben Items vor, zwei in invertierter Form. Zuhören misst das Ausmaß, in dem man anderen konzentriert zuhören kann. Es gibt sechs Items, die alle invertiert vorliegen.

Diese Subdimensionen wurden bereits in anderer Forschung zu Empathie eingesetzt und erwiesen sich als änderungssensitiv (Altmann, 2013), so dass sie als Instrument auch hier zum Einsatz kommen. Die berichteten Reliabilitätswerte der Sekundärfaktoren finden sich in Tabelle 29, die Werte in dieser Studie lagen in einem ähnlichen Wertebereich.

Tabelle 29

Interne Konsistenzen des ISK, Subskala Soziale Orientierung

	Cronbachs Alpha			
	SO	PÜ	WP	ZU
Normierungsstichprobe im Rahmen der Testentwicklung (n=4208)	.75	.80	.69	.80

Anmerkungen. SO=Subskala des ISK *Soziale Orientierung*; PÜ= Subskala des ISK *Perspektivenübernahme*; WP= Subskala des ISK *Wertpluralismus*; ZU= Subskala des ISK *Zuhören*.

Fehlende Daten werden als Missings gekennzeichnet und die jeweiligen Skalen konnten nur gebildet werden, wenn maximal ein Wert fehlte.

4.1.2.3 Aufbau Vignettentest

Nach den Ergebnissen der Pilotierungstestung im Sommersemester (siehe Kapitel 3.4) wurde die Zusammensetzung der Vignetten für jeden Messzeitpunkt noch einmal geändert. Die Vignetten wurden weiterhin in einer randomisierten Reihenfolge gezeigt und auch die Items

jeder Vignette blieben so erhalten wie in den Vorgängerversionen. Ziel war es die internen Konsistenzen der jeweiligen Testversionen zu erhöhen (vgl. Tabelle 30).

Tabelle 30

Übersicht Aufbau Vignettentests im Praxiseinsatz und interne Konsistenzwerte

		Prä	Interim	Post
Zusammensetzung der Testversionen		Vignette 2	Vignette 6	Vignette 2
		Vignette 6	Vignette 8	Vignette 6
		Vignette 9	Vignette 10	Vignette 8
		Vignette 21	Vignette 14	Vignette 21
		Vignette 25	Vignette 27	Vignette 25
		Vignette 30	Vignette 30	Vignette 30
Cronbachs Alpha	Escore	.29	.26	.25
	Iscore	.60	.69	.80
	Vscore	-.76	-.26	.10
	Eigenes Gefühl	.80	.90	.89

Anmerkungen. Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score; Vscore=Verhaltensscore.

Auch wenn die Prä- und Post-Testversion bis auf eine Vignette identisch sind, sind deskriptiv-statistisch sowohl Test-Mittelwerte, als auch -Standardabweichungen in den Testversionen leicht unterschiedlich (vgl. Tabelle 31).

Tabelle 31
Deskriptiv-statistische Auswertung der Testwerte in der Hauptstudie

	<i>N</i>	<i>MW</i>	<i>SD</i>
Testwert Escore Prä	68	25.49	3.53
Testwert Escore Inter	68	22.75	3.13
Testwert Escore Post	68	24.55	3.49
Testwert Iscore Prä	68	18.85	3.46
Testwert Iscore Inter	68	21.34	3.75
Testwert Iscore Post	68	18.79	4.80
Testwert Vscore Prä	67	9.31	1.64
Testwert Vscore Inter	68	10.15	1.88
Testwert Vscore Post	67	11.11	2.20

Anmerkungen. Escore=Emotionsdekodierungs-Score; Iscore=Intensitäts-Score; Vscore=Verhaltensscore.

Da die internen Konsistenzen nach erneutem Anpassen der Vignettenzusammenstellung und dem Scoring im Escore und Vscore – und auch nach Prüfung des Auslassens einzelner Vignettenwerte - immer noch unzureichende Werte lieferten, wird von einer Auswertung dieser Scores an dieser Stelle abgesehen. Die Items sollten für eine zukünftige Verwendung noch einmal revidiert und bearbeitet werden. Die Werte des *I*scores und der *Einschätzung des eigenen Gefühls* fließen weiterhin in die folgenden Analysen mit ein.

Zur Prüfung der Veränderungssensitivität müssen parallele Testformen vorliegen. Zur Feststellung der Parallelität wird hier das Mindestkriterium angelegt, das besagt, dass gleiche wahre Werte und Standardabweichungen vorliegen müssen, um von einer parallelen Testform sprechen zu können (Moosbrugger & Kelava, 2007). Da der Mittelwert einen erwartungstreuen Schätzer des wahren Wertes darstellt, wird dieser als Parameter zu Grunde gelegt. Die Mittelwerte der Iscore-Testwerte der Prä- und Post-Testversion weisen keine signifikanten Unterschiede auf (T-Test, $\alpha=0.05$, $p=.921$), so dass von Testparallelität ausgegangen wird. Zwischen den Testversionen Prä und Inter, sowie Inter und Post gibt es signifikante Unterschiede (T-Test, $\alpha=0.05$, $p=.000$), so dass nicht von Parallelität ausgegangen

werden kann. Entsprechend können Vergleiche über Veränderung nur zwischen der Prä- und Post-Testversion gezogen werden.

Die Kovariaten wurden auf Relevanz überprüft. Hierfür wurde eine Korrelationsanalyse nach Kendalls-Tau-B gerechnet, da keine Normalverteilung der Daten vorlag und nicht zwischen allen AVs und erhobenen Kontrollvariablen Zusammenhänge gegeben waren (siehe Anhang D5). Von den signifikanten neun Variablen Geschlecht, Alter, geclusterte Fachwahl, Fachsemester, ob im SoSe18 ein studienbezogenes Praktikum absolviert wurde, ob in der Vergangenheit ein Ehrenamt geleistet wurde und ob die ProbandInnen Kinder haben, wurde die Relevanz für die jeweiligen Gruppen überprüft. Bei der Frage nach dem Ehrenamt gab es keine signifikanten Unterschiede zwischen den Gruppen (T-Test, 2-seitig, $\alpha=.05$: T-Wert: -1.24, $p=.221$), so dass die Variable nicht in die Analyse einfluss. Alle anderen wurden als Kovariaten in die Analysen mit aufgenommen. Die Kovariate zum Praktikum wurde aufgrund der heterogenen Länge der angegebenen Praktika und der dadurch schlechten Bewertbarkeit von weiteren Analysen ausgeschlossen.

4.1.3 Datenaufbereitung

Die Daten wurden auch in dieser Studie wieder in einen gemeinsamen Datensatz in SPSS integriert, wodurch nur diejenigen ProbandInnen verblieben, die an allen drei Messzeitpunkten teilnehmen konnten ($N=68$). Es fielen die Daten von sechs Personen heraus, die lediglich einen Code eingegeben hatten. Der Datensatz wurde zusätzlich um Dubletten bereinigt bzw. die Daten derselben Personen zusammengefügt, die aufgrund technischer Problematiken von ProbandInnen durch die Nutzung unterschiedlicher Tablets entstanden. Invertierte Items wurden umkodiert und Missings wurden wieder durch den Mittelwert ersetzt (ESEK) oder mit -99 gekennzeichnet. Dadurch konnte in der Prä- und Inter-Version der ISK-Testwert jeweils einer Person nicht gebildet werden.

Die aufgezeichneten Videos von 20-30 Minuten wurden so geschnitten, dass jeweils die Sequenzen übrigblieben, die die Teilnehmenden während des Anschauens der Videoclips zeigten. Diese Videoausschnitte wurden erneut pro Person in einem Face-Reader-Projekt angelegt und nach der individuellen Kalibrierung jedes Videos mit der Software ausgewertet. Die generierten Ergebnisdaten wurden in den SPSS-Datensatz integriert. Einige Analysedaten fehlten aus unsystematischen Gründen (siehe Anhang D6). Bei einigen Tablets wurde aus nicht eruierbaren Gründen kein Video aufgezeichnet, teilweise ließen sich die Videos nicht analysieren, da der Kopfausschnitt der Probandinnen ungünstig ausfiel und Teile des Kopfes abgeschnitten wurden oder es wurden nicht alle Screenshot-Marker aufgezeichnet, so dass keine Einordnung erfolgen konnte, an welchen Stellen Videosequenzen angeschaut wurden. In der Testversion Prä fehlten von zwei bis drei Personen die Face-Reader-Auswertungen (je nach Vignette), in der Testversion Inter fehlten Daten von zwei bis vier Personen (je nach Vignette) und in der Testversion Post fehlten die Werte von fünf bis sieben Personen (je nach Vignette). Zusätzlich wurden die demografischen Angaben händisch überarbeitet (bspw. bei der Semesterangabe entfernen von Interpunktion, so dass die Variable als numerisch erkannt werden konnte u.ä.). Außerdem wurde die Zugehörigkeit zu EG oder KG durch eine Gruppierungsvariable mit 0 (KG) und 1 (EG) kodiert.

4.1.4 Ergebnisse

Zur Berechnung der Unterschiede zwischen den Gruppen und den Verläufen über die Zeit kamen Varianzanalysen zum Einsatz. Hierbei wurden die Mittelwerte der Messzeitpunkte bzw. der Gruppen miteinander verglichen und überprüft, ob sich Gruppen signifikant voneinander unterscheiden (EG von KG oder Messzeitpunkte von der Baseline). Zur querschnittlichen Untersuchung kam die MANOVA zum Einsatz und zur längsschnittlichen Analyse ANOVAs mit Messwiederholung. Beide Analysen gehören zu den allgemeinen linearen Modellen und geben Auskunft darüber wieviel Varianz durch die geprüften Variablen aufgeklärt werden kann. Grundvoraussetzungen zur Anwendung beider Verfahren sind in etwa gleich große Gruppen, die Gruppenwerte sollten normalverteilt sein und es sollte Varianzhomogenität gegeben sein (Field, 2013). Da beide Verfahren allerdings recht robust sind, kann die Verletzung von Normalverteilung und gegebener Varianzhomogenität ausgeglichen werden solange die Gruppen annähernd gleich groß sind, was bei dieser Studie der Fall ist (Field, 2013). Alle Analysen wurden aufgrund der forschungsleitenden Annahmen einseitig und mit einem Signifikanzniveau von $\alpha=0.05$ geprüft.

4.1.4.1 Prüfen der Vorannahmen und längsschnittliche Analysen

Für das Prüfen der Normalverteilung wurde erneut der Shapiro-Wilk Test genutzt. Es ist ein robustes Verfahren, das prüft, ob sich eine Verteilung von der Normalverteilung unterscheidet. Sollte dies der Fall sein, wird das Ergebnis bei einem $\alpha=.05$ signifikant. Über die drei Messzeitpunkte hinweg ist das bei acht Variablen der Fall (siehe Anhang D7). Da die Stichproben allerdings ≥ 30 sind, können die Analysen trotzdem durchgeführt werden, da die Verletzung der Vorannahmen dadurch vernachlässigbar ist (Blanca, Alarcón, Arnau, Bono & Bendayan, 2017).

Zum Abbilden einer Veränderung wurden in den einzelnen Gruppen ANOVAs mit Messwiederholung durchgeführt. Da hier die Interkorrelationen der AV's miteinander eine untergeordnete Rolle spielen, wurde auf das Durchführen einer MANOVA mit Messwiederholung verzichtet. Neben den bereits genannten Vorannahmen sollte auch noch Sphärizität erfüllt sein. Sollte dies nicht der Fall sein, kann die Greenhouse-Geisser-Korrektur angewandt werden. Angefangen mit der EG werden die Ergebnisse im Folgenden dargestellt. Es werden alle Haupteffekte berichtet und Effekte mit Kovariaten sofern Effekte mit einer Fehlerwahrscheinlichkeit von $\alpha \leq 0.05$ auftraten. Eine deskriptivstatistische Übersicht der Testwerte findet sich in Tabelle 32.

Tabelle 32

Deskriptivstatistische Übersicht der Testwerte über die Messzeitpunkte in der EG (Praxiseinsatz)

Variable	MZP	MW	SD
ESEK funktional	Prä	42.57	3.60
	Inter	41.53	3.34
	Post	42.13	3.73
ESEK dysfunktional	Prä	37.70	4.72
	Inter	36.57	4.86
	Post	35.47	5.91
TEQ	Prä	3.16	.34
	Inter	3.11	.32
	Post	3.04	.38
Erkennen eigener Gefühle – EKF	Prä	3.59	.65
	Inter	3.65	.76
	Post	3.76	.70
Erkennen der Gefühle Anderer – EKF	Prä	4.00	.47
	Inter	3.87	.40
	Post	3.95	.51
ISK – Zuhören	Prä	17.33	4.21
	Inter	17.13	3.80
	Post	17.83	3.92
ISK – Wertepluralismus	Prä	21.17	3.61
	Inter	21.70	3.31
	Post	21.83	2.70
ISK -Perspektivenübernahme	Prä	20.07	2.90
	Inter	20.27	3.33
	Post	19.73	2.79
Iscore	Prä	18.83	3.68
	Post	19.20	3.94

Anmerkungen. Iscore=Intensitäts-Score; MZP=Messzeitpunkt.

Aufgrund der zahlreichen Einzelanalysen wurde die α -Fehlerkorrektur nach Bonferroni angewandt, wodurch sich das α -Level auf .006 reduzierte, und kein Effekt mehr signifikant wurde. Die einzelnen Werte der ANOVA Analysen sind in Tabelle 33 zu sehen.

Tabelle 33

Haupt- und relevante Nebeneffekte in der EG (Praxiseinsatz)

	Sphärizität	<i>F</i>	<i>p</i>	η^2
ESEK funktional	angenommen	(2, 54) =.88	.420	.03
ESEK dysfunktional	angenommen	(2, 56) =3.20	.048*	.10
TEQ	Greenhouse-Geisser-Korrektur	(1.63, 42.41) =.04	.936	.00
Erkennen eigener Gefühle – EKF	angenommen	(2, 42) =.16	.849	.01
Erkennen der Gefühle Anderer – EKF	angenommen	(2, 52) =2.63	.082	.09
ISK – Zuhören	Greenhouse-Geisser-Korrektur	(1.38, 38.65) =.68	.459	.02
ISK – Wertepluralismus	angenommen	(2, 52) =2.56	.087	.09
ISK – Wertepluralismus* Naturwissensch. Cluster	angenommen	(2, 52) =6.57	.003**	.20
ISK – Perspektivenüberahme	Greenhouse-Geisser-Korrektur	(1.43, 34.19) =.14	.802	.01
ISK – Perspektivenüberahme * Naturwissensch. Cluster	Greenhouse-Geisser-Korrektur	(1.43, 34.19) =6.28	.010*	.21
Gesamt Iscore	angenommen	(1, 26) =.26	.639	.01

Anmerkungen. * $p < .05$, ** $p < .01$; Naturwissensch. Cluster=ProbandIn studiert eines der folgenden Fächer: Mathe, Physik, Informatik/EDV, Chemie, Biologie; p = Signifikanz; η^2 =Effektstärke; Iscore=Intensitäts-Score.

In der Subskala *Wertepluralismus-ISK* in Abhängigkeit des *naturwissenschaftlichen Fachclusters* zeigt sich bei allen, die kein naturwissenschaftliches Fach studieren, zum zweiten Messzeitpunkt ein Werteeinbruch (vgl. Abbildung 25).

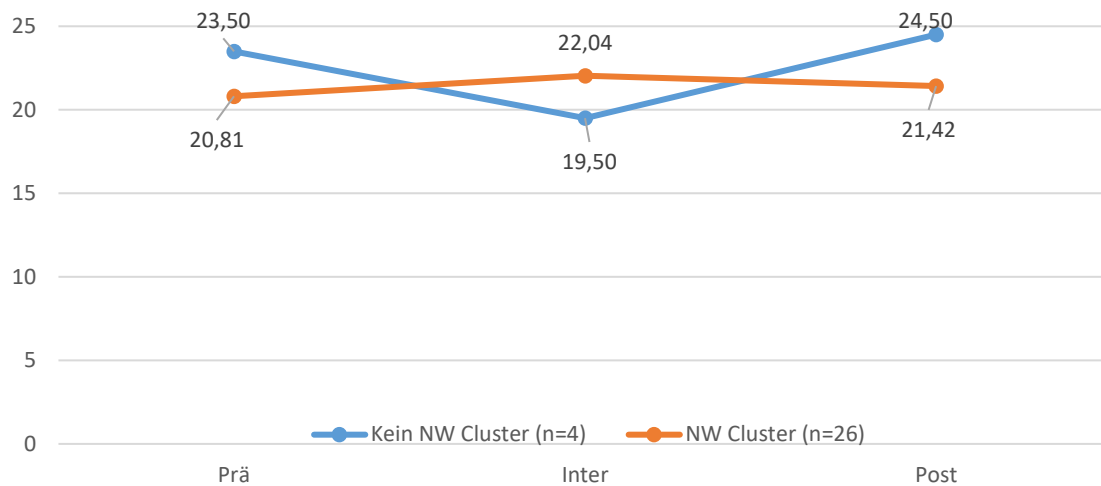


Abbildung 25. Längsschnittliche Verläufe der Subskala Wertepluralismus des ISK innerhalb der EG und der Subgruppe, die ein naturwissenschaftliches Fach studieren (NW Cluster) versus denen, die kein naturwissenschaftliches Fach studieren (Kein NW-Cluster).

In der Subskala *Perspektivenübernahme-ISK* zeigt sich ein ähnlicher Verlauf mit dem *naturwissenschaftlichen Cluster* wie in der Subskala *Wertepluralismus* (vgl. Abbildung 26).

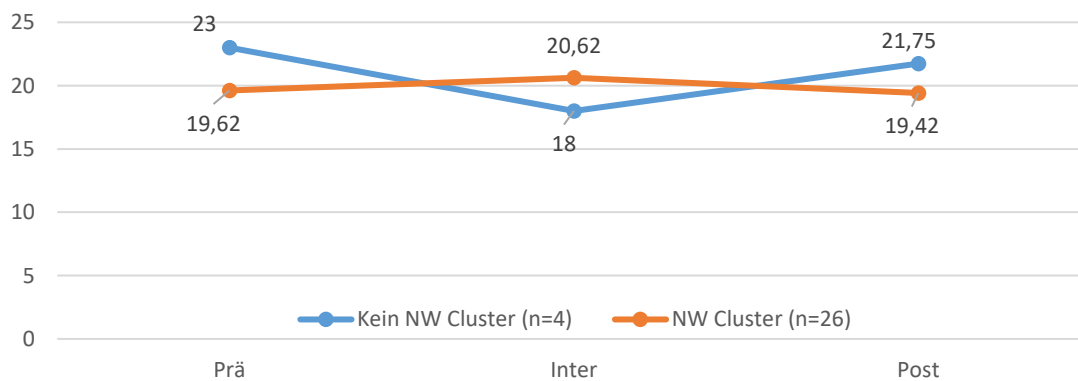


Abbildung 26. Längsschnittliche Verläufe der Subskala Perspektivenübernahme des ISK innerhalb der EG und der Subgruppe, die ein naturwissenschaftliches Fach studieren (NW Cluster) versus denen, die kein naturwissenschaftliches Fach studieren (Kein NW-Cluster).

Ähnliche Ergebnisse wie in der EG sind auch in der KG zu finden. Die deskriptivstatistischen Werte sind als Überblick in Tabelle 34 abgetragen.

Tabelle 34

Deskriptivstatistische Übersicht der Testwerte über die Messzeitpunkte in der KG (Praxiseinsatz)

Variable	MZP	MW	SD
ESEK funktional	Prä	43.18	2.92
	Inter	41.89	3.80
	Post	42.58	3.55
ESEK dysfunktional	Prä	36.08	5.04
	Inter	34.50	4.91
	Post	34.97	5.39
TEQ	Prä	3.17	.39
	Inter	3.12	.40
	Post	3.11	.39
Erkennen eigener Gefühle – EKF	Prä	3.36	.68
	Inter	3.51	.63
	Post	3.56	.66
Erkennen der Gefühle Anderer – EKF	Prä	4.13	.47
	Inter	4.06	.52
	Post	4.10	.59
ISK – Zuhören	Prä	18.58	2.60
	Inter	19.05	3.19
	Post	18.95	3.50
ISK – Wertepluralismus	Prä	21.22	3.51
	Inter	22.22	3.63
	Post	22.11	3.06
ISK -Perspektivenübernahme	Prä	20.00	2.45
	Inter	19.86	2.99
	Post	20.29	2.55
Iscore	Prä	18.86	3.32
	Post	18.47	5.41

Anmerkungen. Iscore=Intensitäts-Score; MZP=Messzeitpunkt.

Auch hier wurden ursprünglich signifikante Ergebnisse nach der Bonferroni-Korrektur ($\alpha=.006$) nicht mehr signifikant (siehe Tabelle 35).

Tabelle 35

Haupt- und relevante Nebeneffekte in der KG (Praxiseinsatz)

	Sphärizität	<i>F</i>	<i>p</i>	η^2
ESEK funktional	angenommen	(2, 72) =4.20	.019*	.11
ESEK funktional * Alter	angenommen	(2, 72) =3.17	.048*	.08
ESEK dysfunktional	angenommen	(2, 72) =2.00	.143	.05
TEQ	angenommen	(2, 68) =.09	.913	.00
Erkennen eigener Gefühle – EKF	angenommen	(2, 60) =.23	.799	.01
Erkennen der Gefühle Anderer – EKF	Greenhouse-Geisser-Korrektur	(1.66, 57.98) =2.90	.072	.08
Erkennen der Gefühle Anderer – EKF * Alter	Greenhouse-Geisser-Korrektur	(1.66, 57.98) =5.13	.013*	.13
ISK – Zuhören	Greenhouse-Geisser-Korrektur	(1.55, 54.13) =.43	.603	.01
ISK – Wertepluralismus	Greenhouse-Geisser-Korrektur	(1.46, 48.24) =1.92	.167	.06
ISK – Perspektivenüberahme	Greenhouse-Geisser-Korrektur	(1.27, 40.76) =1.96	.166	.06
Gesamt Iscore	angenommen	(1, 35) =.00	.979	.00

Anmerkungen. * $p < .05$; p = Signifikanz; η^2 = Effektstärke; Iscore = Intensitäts-Score.

Deskriptivstatistisch lässt sich das Alter mit einem Median von 23 Jahren in zwei Gruppen aufteilen. Schaut man sich den Verlauf im *ESEK funktional* getrennt nach Altersgruppen an sind die Verläufe ähnlich mit einem leichten Abfall zum Inter-Messzeitpunkt und erneutem Anstieg zum Post-Messzeitpunkt (siehe Abbildung 27).

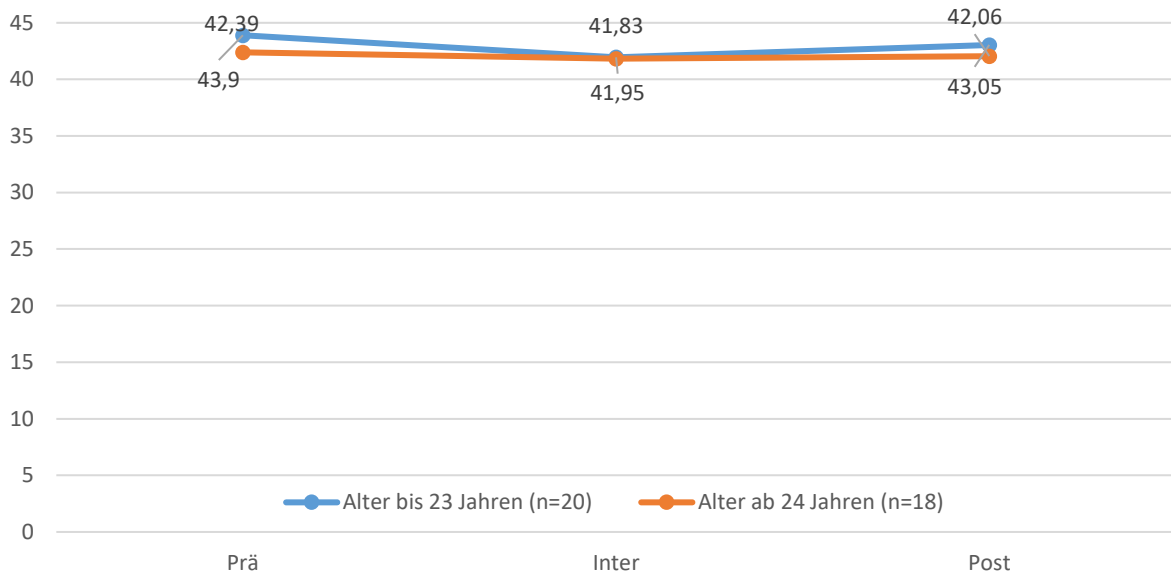


Abbildung 27. Längsschnittliche Verläufe der Subskala funktional des ESEK innerhalb der KG und der Subgruppe Alter (mit Median-Split zur Einteilung in zwei Untergruppen).

Sieht man sich den deskriptivstatistischen Verlauf des *EKF – Erkennen der Emotionen* anderer an, zeigt sich zum Post-Zeitpunkt ein Anstieg des Mittelwertes in der Subgruppe der bis 23-jährigen (siehe Abbildung 28).

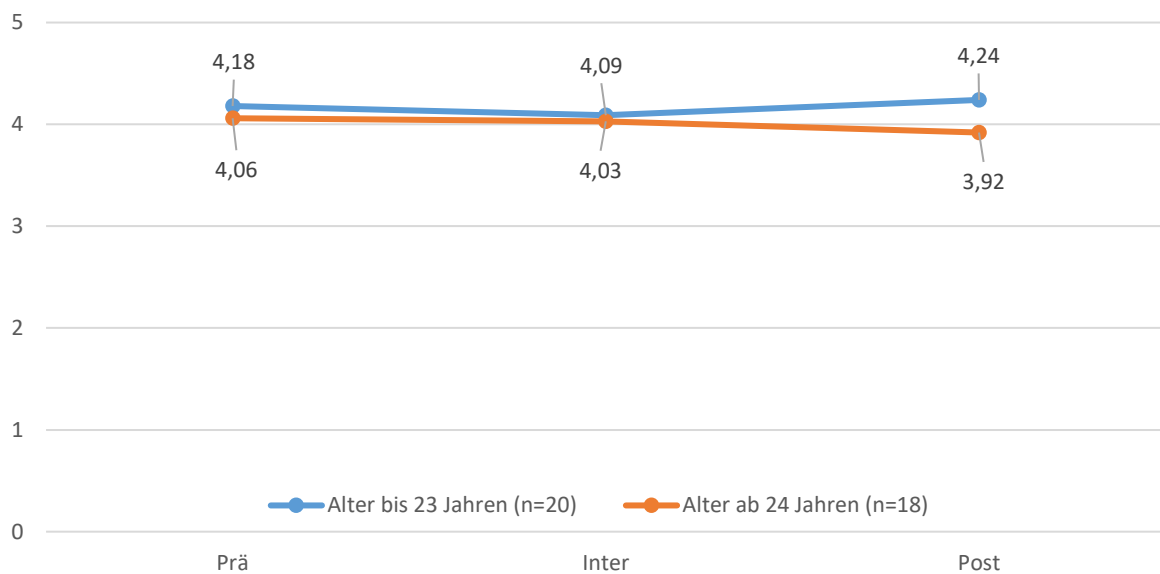


Abbildung 28. Längsschnittliche Verläufe der Subskala Erkennen der Emotionen anderer - EKF in der KG und der Subgruppe Alter (mit Median-Split zur Einteilung in zwei Untergruppen).

4.1.4.2 Querschnittliche Analysen

Beim querschnittlichen Vergleich der Gruppen wurden die Werte der drei Messzeitpunkte zwischen den Gruppen mit multivariater Varianzanalyse (MANOVA) miteinander verglichen. Hierbei wurden alle Maße zusammen ausgewertet, die eingesetzt wurden, um Empathie zu repräsentieren (empathische Akkuratheit sowie empathisches Handeln, siehe Abbildung 23 und Abbildung 24). Die Stichproben sind unabhängig, repräsentativ für die Gesamtpopulation und in etwa gleich groß, so dass die MANOVA zur Anwendung kommen kann. Die Pillai-Bartlett-Spur ist auch bei kleinen Stichproben und wenn Vorannahmen verletzt sind ein robustes Maß, so dass es hier für die Auswertung herangezogen wurde (Field, 2013). Zum Messzeitpunkt Prä und Post sind die Kovarianzen in den Gruppen gleich und der Box-Test negativ. Außer beim ISK-ZU sind auch die Ergebnisse des Levene-Tests nicht signifikant.

Der Haupteffekt *Gruppe* der AVs zu *Empathie* zum Messzeitpunkt Prä wird nicht signifikant ($F(4,57)=1.61, p=.184, \eta^2=.10$), wie ebenfalls der Effekt mit dem *Naturwissenschaftlichen Fachcluster* ($F(4,57)=.76, p=.539, \eta^2=.05$), dem *Sprachlich-literarischen Fachcluster* ($F(4,57)=.80, p=.531, \eta^2=.05$), dem *Gesellschaftswissenschaftlichen Fachcluster* ($F(4,57)=.65, p=.627, \eta^2=.04$) sowie dem *Sonstigen Fachcluster* ($F(4,57)=2.32, p=.068, \eta^2=.14$). Der Effekt von *Alter* auf die AVs wird signifikant ($F(4,57) = 3.30, p=.017, \eta^2=.19$). Auf Ebene der einzelnen Variablen gibt es einen Effekt von *Alter* auf den *I-score* ($F(1, 60)=5.16, p=.027, \eta^2=.08$) sowie auf die Subskala *Erkennen eigener Gefühle – EKF* ($F(1,60)=4.58, p=.036, \eta^2=.07$).

Der Haupteffekt *Gruppe* der AVs *empathisches Handeln* wird ebenfalls nicht signifikant ($F(4, 56)=1.09, p=.373, \eta^2=.07$). Ebenfalls nicht signifikant wird der Effekt von *Alter* ($F(4,56)=.13, p=.972, \eta^2=.01$), dem *Naturwissenschaftlichen Fachcluster* ($F(4,56)=.62, p=.650, \eta^2=.04$), dem *Sprachlich-literarischen Fachcluster* ($F(4,56)=2.03, p=.102, \eta^2=.13$), dem *Gesellschaftswissenschaftlichen Fachcluster* ($F(4,56)=.76, p=.557, \eta^2=.05$) sowie dem *Sonstigen Fachcluster* ($F(4,56)=.84, p=.505, \eta^2=.06$) auf die entsprechenden AVs.

Der Haupteffekt *Gruppe* der AVs zu *Empathie* zum Messzeitpunkt Post wird erneut nicht signifikant ($F(4,42)=1.92, p=.125, \eta^2=.15$), ebenfalls wie der Interaktionseffekt von *Alter* auf die AVs ($F(64,180)=1.27, p=.114, \eta^2=.31$) sowie *Naturwissenschaftlichen Fachcluster* ($F(4,42)=.75, p=.561, \eta^2=.07$), dem *Sprachlich-literarischen Fachcluster* ($F(4,42)=.93, p=.457, \eta^2=.08$), dem *Gesellschaftswissenschaftlichen Fachcluster* ($F(4,42)=.75, p=.567, \eta^2=.07$) sowie dem *Sonstigen Fachcluster* ($F(4,42)=.90, p=.475, \eta^2=.08$).

Der Haupteffekt *Gruppe* der AVs *empathisches Handeln* wird nicht signifikant ($F(4,42)=.64, p=.639, \eta^2=.06$), ebenfalls wie der Interaktionseffekt *Alter* ($F(64,180)=1.12, p=.285, \eta^2=.28$), des *Naturwissenschaftlichen Fachcluster* ($F(4,42)=1.29, p=.290, \eta^2=.11$), dem *Sprachlich-literarischen Fachcluster* ($F(4,42)=.69, p=.601, \eta^2=.06$), dem *Gesellschaftswissenschaftlichen Fachcluster* ($F(4,42)=1.96, p=.120, \eta^2=.16$) sowie dem *Sonstigen Fachcluster* ($F(4,42)=.15, p=.963, \eta^2=.01$).

4.1.4.3 Mitschwingen

Die gemachten Videoaufnahmen wurden im Rahmen der Testentwicklung durchgeführt, um längsschnittliche Verläufe und Veränderungen abzubilden, so dass bei der Auswertung nicht zwischen den beiden Gruppen EG/KG differenziert wird. Gemeinsam mit der Selbsteinschätzung des eigenen Gefühls beim Anschauen der Videosequenz sowie der eingeschätzten Intensität der im Videoclip eingeschätzten Emotion dienen sie dazu, das empathische Mitschwingen abzubilden. Zur Auswertung wurde die non-parametrische Korrelationsanalyse Kendall-Tau-B genutzt (2-seitig, $\alpha=.05$). Die Ergebnisse sind im Folgenden pro Vignette für die Testversionen Prä und Post dargestellt und es werden nur signifikante Zusammenhänge berichtet.

Bei Vignette 2 zeigen sich in der Prä- und auch in der Post-Version kleine bis mittlere Zusammenhänge zwischen der selbst eingeschätzten Emotion und der eingeschätzten Intensität der im Videoclip gesehenen Emotion. Dieser Effekt ist bei allen Basisemotionen kongruent zu

finden, allerdings gibt es auch weitere Effekte mit nicht-kongruenten Emotionen (vgl. Tabelle 36 und Tabelle 37).

Tabelle 36

Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 2, Prä-Version

	EE Freude	EE Ärger	EE Angst	EE Traurigkeit	EE Überraschung/ Neugier	EE Ekel
EI Freude	.29**					
EI Angst			.33**	.24*		
EI Traurigkeit			.27*	.26*		
EI Überraschung/ Neugier			.27*		.44**	
EI Ekel		.32**	.30**	.29**		.49**

Anmerkungen. * $p < .05$, ** $p < .01$; $N=68$; EE=Intensität der selbst-eingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Emotion.

Tabelle 37

Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 2, Post-Version

	EE Freude	EE Ärger	EE Angst	EE Traurigkeit	EE Überraschung/ Neugier	EE Ekel
EI Freude	.50**				.27**	
EI Ärger		.28**	.25*	.24*		.30**
EI Angst	.24*		.55**	.24*		.23*
EI Traurigkeit			.59**	.29**		.36**
EI Überraschung/ Neugier	.40**				.41**	
EI Ekel			.37**	.28*		.56**

Anmerkungen. * $p < .05$, ** $p < .01$; $N=68$; EE=Intensität der selbst-eingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Emotion.

Beim Vergleich der Face-Reader-Daten mit der selbst-eingeschätzten Emotion, als auch der eingeschätzten Intensität der Emotionen im Videoclip der Vignette 2 zeigen sich kleine bis mittlere Effekte, allerdings vor allem mit nicht-kongruenten Emotionen (vgl. Tabelle 38 und Tabelle 39).

Tabelle 38

Korrelationen der Face-Reader-Daten mit der selbst-ingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 2, Prä-Version

	Face-Reader Ärger	Face-Reader Traurigkeit	Face-Reader Überraschung
EI Überraschung/ Neugier			.21*
EE Traurigkeit	.23*		.23*
EE Überraschung/ Neugier		.19*	

Anmerkungen. * $p < .05$, ** $p < .01$; $N=66$; EE=Intensität der selbst-ingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Emotion.

Tabelle 39

Korrelationen der Face-Reader-Daten mit der im Videoclip eingeschätzten Intensität der Vignette 2, Post-Version

	Face-Reader Freude
EI Angst	-.21*

Anmerkungen. * $p < .05$, ** $p < .01$; $N=61$; EI=Intensität der im Videoclip eingeschätzten Emotion.

In Vignette 6 zeigen sich in der Prä- und auch in der Post-Version kleine bis große Zusammenhänge zwischen der selbst eingeschätzten Emotion und der eingeschätzten Intensität der im Videoclip gesehenen Emotion. Dieser Effekt ist ebenfalls bei allen Basisemotionen kongruent zu finden, allerdings gibt es auch weitere Effekte mit nicht-kongruenten Emotionen (vgl. Tabelle 40 und Tabelle 41).

Tabelle 40

Korrelationen der selbst-ingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 6, Prä-Version

	EE Freude	EE Ärger	EE Angst	EE Traurigkeit	EE Überraschung/ Neugier	EE Ekel
EI Freude	.55**				.26*	
EI Ärger		.47**				.29*
EI Angst			.26*	.22*		
EI Traurigkeit		.27**	.22*	.39**		
EI Überraschung/ Neugier					.51**	
EI Ekel		.27*				.82**

Anmerkungen. * $p < .05$, ** $p < .01$; N=68; EE=Intensität der selbst eingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Emotion.

Tabelle 41

Korrelationen der selbst-ingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 6, Post-Version

	EE Freude	EE Ärger	EE Angst	EE Traurigkeit	EE Überraschung/ Neugier	EE Ekel
EI Freude	.56**		.24*		.30**	
EI Ärger	.34**	.30**				.40**
EI Angst		.35**	.22*	.20*		
EI Traurigkeit		.41**	.24*	.46**		
EI Überraschung/ Neugier	.27*				.65**	
EI Ekel	.48**	.33**	.33**		.23*	.71**

Anmerkungen. * $p < .05$, ** $p < .01$; $N=68$; EE=Intensität der selbst-ingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Emotion.

Die Face-Reader-Daten betreffend gibt es bei Vignette 6 einen kongruenten, kleinen Zusammenhang zwischen der Intensität des eingeschätzten Ärgers im Videoclip und der Face-Reader-Analyse. Einen negativen, kleinen Zusammenhang gibt es zwischen der selbst-ingeschätzten Freude und den Face-Reader-Daten zu Angst in der Post-Version. Alle anderen gefundenen Effekte sind kleine Zusammenhänge mit nicht-kongruenten Gefühlen und sogar ein

negativer Zusammenhang zwischen selbst-eingeschätztem Ekel und der Face-Reader-Analyse zu Ekel in der Post-Version (vgl. Tabelle 42 und Tabelle 43).

Tabelle 42

Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 6, Prä-Version

	Face-Reader Freude	Face- Reader Ärger	Face- Reader Angst	Face- Reader Überraschung	Face- Reader Ekel
EI Ärger		.21*			
EI Traurigkeit					.19*
EE Freude	.21*				
EE Traurigkeit	.21*		.27**	.27**	.23*
EE Überraschung/ Neugier	.19*		.23*		

Anmerkungen. * $p < .05$, ** $p < .01$; $N=65$; EE=Intensität der selbst-eingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Emotion.

Tabelle 43

Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 6, Post-Version

	Face-Reader Freude	Face-Reader Angst	Face-Reader Ekel
EI Überraschung/ Neugier	.22*		
EE Freude		-.23*	-.26*
EE Ekel			-.21*

Anmerkungen. * $p < .05$; $N=62$; EE=Intensität der selbst-eingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Emotion.

In Vignette 21 finden sich in der Prä- und auch in der Post-Version ebenfalls kleine bis große Zusammenhänge zwischen der selbst eingeschätzten Emotion und der eingeschätzten Intensität der im Videoclip gesehenen Emotion. Dieser Effekt ist bei nahezu allen Basisemotionen kongruent zu finden (außer bei Ärger in der Prä-Version), allerdings gibt es auch weitere Effekte mit nicht-kongruenten Emotionen (vgl. Tabelle 44 und Tabelle 45).

Tabelle 44

Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 21, Prä-Version

	EE Freude	EE Angst	EE Traurigkeit	EE Überraschung/ Neugier	EE Ekel
EI Freude	.53**				
EI Angst	-.21*	.34**	.31**		.23*
EI Traurigkeit			.37**		.23*
EI Überraschung/ Neugier				.28**	
EI Ekel					.34**

Anmerkungen. * $p < .05$, ** $p < .01$; $N=68$; EE=Intensität der selbst-eingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Intensität.

Tabelle 45

Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 21, Post-Version

	EE Freude	EE Ärger	EE Angst	EE Traurigkeit	EE Überraschung/ Neugier	EE Ekel
EI Freude	.63**	-.34**		-.22*	.20*	-.27**
EI Ärger		.20*				
EI Angst		.45**	.32**	.28**		.32**
EI Traurigkeit		.39**	.33**	.57**		.38**
EI Überraschung/ Neugier	.24*				.54**	
EI Ekel			.27*	.32**		.55**

Anmerkungen. * $p < .05$, ** $p < .01$; $N=68$; EE=Intensität der selbst-eingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Intensität.

Die Zusammenhänge zwischen den Face-Reader-Daten und der selbst-eingeschätzten Emotion sowie der eingeschätzten Intensität der Emotion im Videoclip der Vignette 21 liegen im kleinen bis mittleren Bereich und sind bis auf *Überraschung* in der Post-Version meist bei nicht-kongruenten Emotionen zu finden (vgl. Tabelle 46 und Tabelle 47).

Tabelle 46

Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 21, Prä-Version

	Face-Reader Angst	Face-Reader Überraschung
EI Ärger	.20*	
EI Traurigkeit		.31**
EE Traurigkeit		.24*

Anmerkungen. * $p < .05$, ** $p < .01$; $N=65$; EE=Intensität der selbst-eingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Emotion.

Tabelle 47

Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 21, Post-Version

	Face-Reader Ärger	Face-Reader Traurigkeit	Face-Reader Überraschung
EI Freude	-.22*		.20*
EI Ärger		.28**	
EE Freude	-.32**		
EE Überraschung/ Neugier	-.19*		.24*

Anmerkungen. * $p < .05$, ** $p < .01$; $N=62$; EE=Intensität der selbst-eingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Emotion.

Bei Vignette 25 zeigen sich in der Prä- und auch in der Post-Version ähnliche Zusammenhänge wie bei den vorherigen Vignetten auch kleine bis mittlere Zusammenhänge zwischen der selbst eingeschätzten Emotion und der eingeschätzten Intensität der im Videoclip gesehenen Emotion.

Dieser Effekt ist bei allen Basisemotionen kongruent zu finden, allerdings gibt es auch hier weitere Effekte mit nicht-kongruenten Emotionen (vgl. Tabelle 48 und Tabelle 49).

Tabelle 48

Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 25, Prä-Version

	EE Freude	EE Ärger	EE Angst	EE Traurigkeit	EE Überraschung/ Neugier	EE Ekel
EI Freude	.49**					
EI Ärger		.43**				
EI Angst		.36**	.38**	.36**		
EI Traurigkeit		.28**		.35**		
EI Überraschung/ Neugier					.37**	
EI Ekel	.35**					.61**

Anmerkungen. * $p < .05$, ** $p < .01$; $N=68$; EE=Intensität der selbst-eingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Emotion.

Tabelle 49

Korrelationen der selbst-ingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 25, Post-Version

	EE Freude	EE Ärger	EE Angst	EE Traurigkeit	EE Überraschung/ Neugier	EE Ekel
EI Freude	.26*					
EI Ärger		.39**		.30**		.29**
EI Angst		.38**	.47**	.47**		
EI Traurigkeit		.32**	.31**	.56**	.22*	.32**
EI Überraschung/ Neugier					.40**	
EI Ekel			.23*	.27*		.70**

Anmerkungen. * $p < .05$, ** $p < .01$; $N=68$; EE=Intensität der selbst-ingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Emotion.

Die Korrelationen der Vignette 25 zwischen den Face-Reader-Daten, sowie der selbst-ingeschätzten Emotion und der eingeschätzten Emotionsintensität des Videoclips ergeben kleine Zusammenhänge bei nicht-kongruenten Emotionen (vgl. Tabelle 50 und Tabelle 51).

Tabelle 50

Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 25, Prä-Version

	Face-Reader Freude	Face-Reader Ekel
EI Traurigkeit	.25**	
EE Traurigkeit		.29**

Anmerkungen. ** $p < .01$; $N=63$; EE=Intensität der selbst-eingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Emotion.

Tabelle 51

Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion der Vignette 25, Post-Version

	Face-Reader Ärger	Face-Reader Überraschung
EE Angst	-.19*	-.23*
EE Traurigkeit	-.19*	
EE Ekel	-.21*	

Anmerkungen. * $p < .05$; $N=63$; EE=Intensität der selbst-eingeschätzten Emotion.

In Vignette 30 zeigen sich in der Prä- und auch in der Post-Version erneut ähnliche Zusammenhänge wie bei den vorherigen Vignetten auch. Kleine bis große Zusammenhänge zwischen der selbst eingeschätzten Emotion und der eingeschätzten Intensität der im Videoclip gesehenen Emotionen. Dieser Effekt ist bei allen Basisemotionen kongruent zu finden, allerdings gibt es auch hier (außer bei *Überraschung/Neugier*) weitere Effekte mit nicht-kongruenten Emotionen (vgl. Tabelle 52 und Tabelle 53).

Tabelle 52

Korrelationen der selbst-ingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 30, Prä-Version

	EE Freude	EE Ärger	EE Angst	EE Traurigkeit	EE Überraschung/ Neugier	EE Ekel
EI Freude	.36**					
EI Ärger		.29**	.27*			.27*
EI Angst			.45**	.40**		.33**
EI Traurigkeit						.27*
EI Überraschung/ Neugier					.24*	
EI Ekel				.30*		.42**

Anmerkungen. * $p < .05$, ** $p < .01$; $N=66$; EE=Intensität der selbst-ingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Emotion.

Tabelle 53

Korrelationen der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 30, Post-Version

	EE Freude	EE Ärger	EE Angst	EE Traurigkeit	EE Überraschung/ Neugier	EE Ekel
EI Freude	.42**					
EI Ärger		.56**	.34**	.25*	.20*	.25*
EI Angst		.31**	.42**	.32**		.33**
EI Traurigkeit		.24*	.37**	.50**		.55**
EI Überraschung/ Neugier					.48**	
EI Ekel		.37**		.42**	.26*	.46**

Anmerkungen. * $p < .05$, ** $p < .01$; $N=68$; EE=Intensität der selbst-eingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Emotion.

Zwischen den Face-Reader-Daten, sowie der selbst-eingeschätzten Emotion und der eingeschätzten Emotionsintensität des Videoclips in Vignette 30 ergeben sich kleine Zusammenhänge bei kongruenten, sowie nicht-kongruenten Emotionen (vgl. Tabelle 54 und Tabelle 55). Einen negativen Zusammenhang gibt es in der Post-Version zwischen den Face-Reader-Daten von *Ekel* und der selbst eingeschätzten Intensität von *Ekel*.

Tabelle 54

Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion und der im Videoclip eingeschätzten Intensität der Vignette 30, Prä-Version

	Face-Reader Ärger	Face-Reader Angst	Face-Reader Traurigkeit	Face-Reader Überraschung
EI Angst		.27**		
EI Traurigkeit		.30**		.27**
EE Ärger				
EE Angst	.23*		-.23*	
EE Ekel	.22*			.24*

Anmerkungen. * $p < .05$, ** $p < .01$; $N=63$; EE=Intensität der selbst-eingeschätzten Emotion; EI=Intensität der im Videoclip eingeschätzten Emotion.

Tabelle 55

Korrelationen der Face-Reader-Daten mit der selbst-eingeschätzten Emotion der Vignette 30, Post-Version

	Face-Reader Freude	Face-Reader Ekel
EE Freude	.27**	
EE Ekel		-.27*

Anmerkungen. * $p < .05$, ** $p < .01$; $N=63$; EE=Intensität der selbst-eingeschätzten Emotion.

4.1.5 Diskussion

Beim Praxiseinsatz sollte der entwickelte Vignettentest geprüft werden und ob das Lehr-Lernformat SL dazu geeignet ist, Empathie bei Lehramtsstudierenden zu fördern. Im längsschnittlichen Kontrollgruppen-Design wurden dazu Studierende konventioneller Lehrveranstaltungen (klassisches Seminar) mit Studierenden verglichen, die an SL-Veranstaltungen teilnahmen. Es wurden sowohl die längsschnittlichen Verläufe, als auch querschnittliche Vergleiche zum Prä- und Post-Zeitpunkt analysiert. Um die Effektgröße einordnen zu können, wird vorgeschlagen η^2 so zu interpretieren wie Pearson's-Produkt-Moment-Korrelation (Cohen, 2001). In der Regel bedeutet das .10 als einen kleinen, .30 als einen mittleren und .50 als einen großen Effekt zu interpretieren.

4.1.5.1 Längs- und Querschnittliche Analysen

In den längsschnittlichen Analysen konnten nach Anwendung der Bonferroni-Korrektur ($\alpha=.006$) keine signifikanten Veränderungen in den interessierenden Variablen festgestellt werden. Weder bei *ESEK*, noch beim *TEQ*, noch bei den Skalen des *EKF* und *ISK* oder des *I-scores* gab es in der EG oder der KG signifikante Effekte. Lediglich in der Subskala *Wertpluralismus-ISK* gibt es einen kleinen, signifikanten Effekt in Verbindung mit dem *naturwissenschaftlichen Fachcluster*. Bei allen Studierenden, die kein naturwissenschaftliches Fach studierten, gab es zum Inter-Messzeitpunkt einen kleinen Mittelwert-Einbruch, der sich allerdings zum Post-Zeitpunkt über den Prä-Wert steigerte. Auffallend ist hierbei erst einmal, dass es nur vier Studierende gibt, die in der EG kein naturwissenschaftliches Fach studierten, so dass die Substichprobe äußerst klein war. Es ist anzunehmen, dass es sich hierbei um eine Verzerrung aufgrund der unterschiedlichen Gruppengröße handelt. Zur weiteren Überprüfung könnte man gezielt gleich große Subgruppen erheben, um einen Effekt mit größerer Sicherheit nachweisen zu können. Eine alternative Erklärung könnte natürlich auch sein, dass gerade die Studierenden mit naturwissenschaftlichen Fächern mehr von SL profitierten, so dass sich

dadurch die steigenden Mittelwerte erklären ließen. Durch die SL-Veranstaltungen wurden sie mit sozialen Herausforderungen und mit diversen Werten und Einstellungen konfrontiert (Eyler & Giles, 1999), was erst einmal zu einer Überforderung führte, so dass die selbst eingeschätzte Offenheit für Diversität niedriger hatte ausfallen können.

Das Fehlen weiterer signifikanter Werte lässt sich zum einen dadurch erklären, dass durch das Anwenden der Alphafehler-Korrektur das Alpha-Niveau auf sehr niedrige Werte heruntersetzt wurde. Sieht man sich alle Werte an, die vor der Fehlerkorrektur bei einem Alpha-Niveau von .05 signifikant geworden wären, sind ein paar Effekte zu nennen.

In der EG gibt es einen Unterschied über die Zeit in der dysfunktional-Skala des *ESEK* mit einem kleinen Effekt. Schaut man in die Daten zeigt sich hier eine Abnahme der Punktzahl, was als Zunahme eines Verständnisses für professionelles empathisches Verhalten gesehen werden kann. Die weniger funktionalen Annahmen über Empathie scheinen sich verändert und in relevantem Maße abgenommen zu haben. Dieser Effekt zeigt sich statistisch in der KG nicht, obwohl es auch hier deskriptivstatistisch eine Abnahme der durchschnittlichen Testwerte gibt. Ein Grund für diesen Befund kann auch sein, dass die SL-Seminare durch ihre Praxisnähe alle unterrichts-ähnliche Teile aufweisen, in denen sich die Studierenden als Lehrkräfte erproben können. Dies stärkt natürlich die Kompetenzen und Fähigkeiten im pädagogisch-psychologischen Bereich auch hinsichtlich eines besseren Verständnisses der Lehrendenrolle (Goodwin, 1996; Voss et al., 2011). Der *ESEK* fragt Lehrendenverhalten ab und speziell danach was empathisches Verhalten in diesem Kontext der eigenen Einschätzung nach bedeutet (Schönefeld & Roth, 2017). Dieses Wissen scheint sich durch die Teilnahme an SL-Veranstaltungen positiv und in höherem Ausmaß zu verändern als in der KG. Es könnte auch gewinnbringend sein SL-Veranstaltungen dahin gehend mit Veranstaltungen rund um die Praxisphase im Lehramtsstudium zu vergleichen (z.B. Schulpraktische Studien oder Praxissemester). Hierbei ist anzunehmen, dass sich ähnliche Ergebnisse zeigen könnten.

Trotzdem muss erwähnt werden, dass die statistische Relevanz aufgrund der erniedrigten Alpha-Niveaus nur als Hinweis für weitergehende Forschung dienen kann.

Eine weitere Veränderung über die Zeit konnte in der Subskala *Perspektivenübernahme* des *ISK* in Interaktion mit dem Studium eines Faches des *Naturwissenschaftlichen Fächerclusters* gefunden werden. Hierbei zeigte sich (vor Anwendung der Fehlerkorrektur des Alpha-Niveaus) ein signifikanter, kleiner Effekt in der EG. Deskriptivstatistisch reduzierte sich der Test-Mittelwert bei allen vier Personen, die kein naturwissenschaftliches Fach studierten zum Inter-Messzeitpunkt, um zum Post-Messzeitpunkt wieder zu steigen. Der Verlauf bei der Gruppe mit den Studierenden mit *naturwissenschaftlichem Fach* startete und endete im Vergleich mit niedrigeren Werten und erhöhte sich leicht zum Inter-Messzeitpunkt. Auch bei diesem Testwert kann erst einmal von einem Stichprobenartefakt aufgrund der kleinen Substichprobe ausgegangen werden. Auch in diesem Fall sollte zukünftige Forschung auf ausgewogenere Gruppengrößen achten, um aussagekräftige Ergebnisse zu erhalten und belastbare Schlussfolgerungen ziehen zu können. Zudem ist auch hier denkbar, dass sich die Subgruppe der vier Studierenden ohne naturwissenschaftliches Fach bei der Ausübung in der SL-Veranstaltung zum Inter-Messzeitpunkt in einer überfordernden Situation befand und die eigenen Fähigkeiten zur *Perspektivenübernahme* dem geschuldet niedriger bewerteten. Die Selbsteinschätzung zum Prä-Messzeitpunkt fiel höher aus, so dass in Konfrontation mit der Zielstichprobe der Eindruck entstanden sein könnte, die Erlebenswelt der anderen nicht gut erfassen zu können, so dass die *Perspektivenübernahme* nicht als erfolgreich erlebt wurde (Bischof-Köhler, 2009; Decety & Ickes, 2011). Durch die vormalige Einschätzung die Erlebenswelten anderer gut nachvollziehen zu können, könnten die Studierenden die Erfahrung gemacht haben, ihre Fähigkeiten zu überschätzen und ihr Können und Wissen für ausgeprägter gehalten haben als sich in der Praxistätigkeit dann zeigte (Bird & Viding, 2014).

Ein weiterer kleiner, signifikanter Effekt zeigte sich in der KG auf der *funktionalen Skala des ESEK*. Deskriptivstatistisch fiel der Mittelwert zum Inter-Messzeitpunkt und stieg wieder ein

wenig zum Post-Zeitpunkt an. Es ist denkbar, dass sich das Verständnis von funktionalen Anteilen der Empathie über das Semester hinweg verschlechterte. Auch hier sind die Ergebnisse allerdings nur einschränkend bewertbar, da auch dieser Effekt nach Bonferroni-Korrektur nicht mehr signifikant wurde. Einen ähnlichen Verlauf gibt es bei dem kleinen Effekt des *Alters* auf die funktionale Skala des *ESEK*. Auch hier könnten ähnliche Gründe eine Rolle spielen.

In der Subskala *Erkennen der Gefühle Anderer* des *EKF* zeigt sich in Abhängigkeit des Alters über die Zeit ein kleiner, signifikanter Effekt in der KG. Der *EKF* ist ein Selbstauskunftsverfahren, mit dem die ProbandInnen ihre Fähigkeit einschätzen, im allgemeinen Gefühle bei anderen erkennen zu können (Rindermann, 2009). Es werden also keine konkreten Gesichtsausdrücke benannt, sondern nur die selbst erlebte Fähigkeit dazu. Es ist aus der Forschung bekannt, dass die Genauigkeit bestimmte Emotionen zu erkennen im Alter abnimmt (für eine Übersicht vgl. Richter, 2009). Dies scheint vor allem Traurigkeit, Ärger und Angst zu betreffen. Allerdings ist diesen Studien allen gemein, dass die Fähigkeiten mit objektiven Verfahren erhoben wurden, so dass eine Übertragbarkeit auf die Befunde in dieser Studie nur eingeschränkt möglich ist. Zu ähnlichen Ergebnissen kommt auch Rindermann selbst (2009) bei der Entwicklung seines Fragebogens. Ältere Personen scheinen in der Regel zurückhaltender zu werden, was den emotionalen Ausdruck angeht und geübter zu sein was den Umgang mit eigenen Gefühlen betrifft. Allerdings betrifft dies vornehmlich Altersgruppen über 50 Jahren, die hier nicht vorliegen. Die Selbsteinschätzung, bei anderen treffend zu erkennen was sie fühlen, scheint bei ProbandInnen, die älter als 23 Jahre sind (Median=23 Jahre), zumindest in der KG über die Messzeitpunkte abzunehmen. Es ist denkbar, dass dieser Effekt auch mit dem Bearbeiten des Vignettentests zusammenhängen könnte. Die Skalen des *EKF* wurden in der Testbatterie nach den Vignetten präsentiert, so dass zu vermuten ist, dass sich eine Unsicherheit bzgl. der Bewertung der Emotionen in den Videos auch im Antwortverhalten des *EKF* widerspiegelt. In der KG gibt es (wie in der EG auch) zwar keine signifikanten

Unterschiede zwischen den Messzeitpunkten beim Vignettentest, allerdings lässt sich deskriptivstatistisch sehen, dass in der KG bezogen auf das Alter eine Reduktion des Mittelwertes auftritt während in der EG bezogen auf das Alter der Mittelwert steigt. Daraus ließe sich ableiten, dass die Reduktion bei der EG nicht zu beobachten ist, da durch die praktische Erfahrung der SL-Veranstaltung eine größere Zuversicht herrscht, Gefühle bei anderen erkennen zu können.

Als letzter Wert im Vignettentest verblieb lediglich die Einschätzung der Intensität der gesehenen Emotionen. Sowohl in der EG, als auch der KG gab es längsschnittlich betrachtet keine signifikanten Veränderungen. Schaut man deskriptivstatistisch auf die Daten fällt allerdings auf, dass es in der KG fallende und in der EG steigende Mittelwerte von Prä zu Post gibt. Es lässt sich spekulieren, ob diese Ergebnisse rein zufällig sind (was die statistische Analyse bekräftigt) oder zumindest eine Tendenz aufzeigen, dass SL-Studierende besser im Einschätzen von Emotionen sein könnten. Nach Überarbeitung des Vignettentests lässt sich in Kombination mit den anderen Werten, die hier aufgrund der niedrigen Gütemaße außen vor gelassen wurden, sicher mehr sagen.

Bei der Analyse der querschnittlichen Auswertungen zwischen den Gruppen zeigt sich zu allen drei Messzeitpunkten kein signifikanter Unterschied. Unabhängig davon ob die Studierenden eine SL-Veranstaltung oder keine besucht haben, ist die Entwicklung in beiden Gruppen über die Empathie relevanten Maße so ähnlich, dass keine Unterschiede nachgewiesen werden können. Zum Messzeitpunkt Prä gibt es allerdings einen Effekt zwischen dem *Alter* und den abhängigen Variablen, der zum Post-Zeitpunkt allerdings verschwindet. Mit Blick auf die Einzelvariablen zeigt sich der Effekt auf der Skala *Erkennen der eigenen Gefühle – EKF* und dem *IScore*. Hierbei haben alle Studierenden >23 Jahre im Schnitt höhere Werte als die jüngeren Studierenden. Dieser Effekt verschwindet allerdings wieder, so dass davon ausgegangen werden kann, dass sich der Unterschied über die *Zeit auswächst*. Bei den Unterschieden im *IScore* ist denkbar, dass hier ein Lerneffekt stattfindet und über alle

Altersgruppen durch die mehrfache Präsentation der Vignetten eine bessere Einschätzung erfolgen konnte.

Zwischen allen Variablen des empathischen Handelns gibt es zwischen den Gruppen zu allen Messzeitpunkten keine signifikanten Unterschiede – weder Haupteffekte, noch Interaktionseffekte zwischen dem Variablencluster der abhängigen Variablen und den getesteten Kovariaten. Es lässt sich annehmen, dass die besuchte Veranstaltungsform keinen Einfluss auf diese Variablen nimmt und keine seminarspezifische Entwicklung stattfindet.

4.1.5.2 Mitschwingen

Um objektivere Aussagen über die erlebten Emotionen beim Anschauen der Videosequenzen zu bekommen, wurden Aufnahmen der Gesichter der ProbandInnen gemacht. Dadurch konnten die mimischen Veränderungen mit Hilfe einer Face-Reader Software diskreten Emotionen zugeordnet und ein Mittelwert für jede Basisemotion ermittelt werden. Zusätzlich wurde per Selbstauskunft die *Intensität der selbst erlebten Emotion (EE)* erhoben und in Relation gesetzt zu *der Intensität der Emotionen, die in der Videosequenz bei der entsprechenden Person gesehen und bewertet wurden (EI)*. Diese Maße sollen als Einschätzung für das Mitschwingen herangezogen werden. Generell ist zu sagen, dass die ermittelten Werte über alle Vignetten und Messzeitpunkte recht niedrig ausfielen, was sicherlich damit zusammenhängt, dass die in den Videos gezeigten emotionalen Situationen wenig ausgeprägt waren. Zudem muss davon ausgegangen werden, dass sich Gewöhnungseffekte einstellten, sobald Videos mehrfach gezeigt wurden und damit das Interesse an der gezeigten Person sowie Situation abflachte, was sich in der emotionalen Reaktion der ProbandInnen zeigte.

Betrachtet man die Vignetten zum Prä-Messzeitpunkt lässt sich erkennen, dass es bei den meisten Vignetten zwischen der *EE* und *EI* kleine bis große emotionskongruente Zusammenhänge gibt. Diese Effekte werden zum Post-Zeitpunkt kleiner und die Anzahl der nicht-emotions-kongruenten Zusammenhänge nimmt zu. Das lässt den Schluss zu, dass

zumindest zum Prä-Testzeitpunkt selbst eingeschätzt ein emotionales Mitschwingen stattgefunden zu haben scheint. Ob es sich dabei allerdings nur um einen Verzerrungseffekt im Antwortverhalten handelt oder die tatsächlich stattfindende emotionale Schwingung widerspiegelt, lässt sich durch die Auswertung der Face-Reader-Analysen genauer sagen. Nimmt man die Face-Reader Daten in die Bewertung mit hinein zeigt sich, dass die *EE* in den meisten Fällen nicht bestätigt werden kann. Übereinstimmungen zeigen sich in der Prä-Version in der Vignette 6 mit Freude und der *EI* Ärger, die laut Face-Reader-Analyse in der Gruppe auch geteilt wird. Ansonsten zeigen sich weitestgehend eher inkongruente Zusammenhänge in der Emotionseinschätzung/-bewertung, so dass ein Mitschwingen nur bedingt gefunden werden konnte.

Vor Auswertung der aufgenommenen Videos wurden die Gesichtsaufnahmen für den jeweiligen Studierenden kalibriert, da bekannt ist, dass Face-Reader in manchen Fällen bspw. Konzentration als Ärger interpretiert und deswegen eine Kalibrierung anhand einer neutralen Sequenz hilfreich ist (Zaman & Shrimpton-Smith, 2006). Ebenfalls trug die helle Beleuchtung zu mimischen Veränderungen bei, so dass auch diese Rahmenbedingung eine Kalibrierung notwendig machte. Das kann allerdings auch dazu geführt haben, dass eigentlich durch Emotionen bedingte mimische Bewegungen herausgerechnet oder in sehr viel niedrigerem Ausmaß erkannt wurden. In einer Studie von Lewinski et al. (2014) konnte gezeigt werden, dass die Korrektheit von Face-Reader im Erkennen von Emotionen bei .67 liegt (menschliche Kodierer bestehen das Training des Facial Acting Coding Systems, auf dem Face-Reader aufbaut, mit einer Analyseleistung von .70). Damit sind die Ergebnisse recht genau, allerdings nicht fehlerfrei. Erschwerend kommt beim Erkennen der Emotionen in den Videovignetten hinzu, dass die Gesichter der SchülerInnen aufgrund der Aufnahmequalität teilweise nur schwer oder gar nicht zu erkennen waren. Es liegt die Vermutung nahe, dass die Teilnehmenden kognitiv zu schlussfolgern versuchten welche Emotionen zu sehen waren, dadurch allerdings die mentale Repräsentation entweder unvollständig war oder zumindest nicht in eine passende,

adäquate Emotion umgesetzt wurde. Für zukünftige Forschung sollte das Stimulusmaterial emotional gehaltvoller sein und die Möglichkeit gegeben sein, objektiv die Emotionen im Stimulusmaterial selbst auswerten zu können (bspw. auch über Face-Reader, was in dieser Studie wegen schlechter Aufnahmequalität und Gruppensituationen nicht umsetzbar war), so dass ein Mitschwingen daran festgemacht werden kann.

Im Allgemeinen scheint es sehr sinnvoll objektive Bewertungsmöglichkeiten in Empathiemessungen zu implementieren, um vor allem auch Facetten des Mitschwingens valide erfassen zu können. Die Umsetzung scheint allerdings an gewisse Kriterien gebunden zu sein, die bereits mit dem Stimulusmaterial beginnen und auch die Aufnahmequalität betreffen. Face-Reader ist ein ökonomisches Werkzeug, das zusätzlich zu anderen Instrumenten wertvolle Informationen liefern kann.

Kapitel 5: Fazit der Untersuchungen

5.1 Fazit zur Untersuchung von Empathie in SL Veranstaltungen

Die Befundlage zu den Ergebnissen der Testungen in SL-Veranstaltungen und der Frage danach, ob derartige Veranstaltungen dazu geeignet sind, Empathie bei Lehramtsstudierenden besser zu entwickeln als konventionelle Seminare, ist ernüchternd. In der hier vorliegenden Untersuchung konnte ein solcher Vorteil nicht gefunden werden. Diese Ergebnisse können allerdings auf verschiedene Gründe zurückgeführt werden. Zum einen ist es denkbar, dass die eingeschränkte Auswahl an SL-Seminaren und die damit nicht einher gegangene inhaltliche Qualitätsprüfung einen Einfluss gehabt haben könnte. Die Veranstaltungen wurden aufgrund ihrer Verfügbarkeit aus dem Lehrveranstaltungskatalog der Goethe-Universität ausgewählt und aufgrund dieser Kursbeschreibungen als geeignet befunden. Es fehlte eine objektive, zusätzliche Befragung der Dozierenden, um eine qualitative Einordnung anhand der Schwerpunkteinordnung von Service und Learning (Sigmon, 1997) sowie der K-12 Kriterien vornehmen zu können (Billig & Weah, 2008). Es bleibt zudem unklar wie viele Praxisstunden die Studierenden tatsächlich in ihrem Kurs hatten und wie hoch der Stellenwert jedes Bausteins (*Service* und *Learning*) tatsächlich war. Schon Eyler und Giles (1999) sehen in der Heterogenität des Veranstaltungsaufbaus eine Möglichkeit genau zu erforschen, welche Aspekte vermutlich zu welchem Outcome führen. Dementsprechend scheint es lohnend sich die Bestandteile der SL-Veranstaltungen genauer anzuschauen sowie zu bewerten und auf Basis dessen Auswertungen durchzuführen. Mit dieser Einordnung hätte die Gruppe der Studierenden in SL-Veranstaltungen noch einmal aufgeteilt oder gewichtet werden können. Unter Umständen hätte sich das auch in den Ergebnissen widerspiegelt. Vermutlich wäre es in diesem Fall ein methodisch gewinnbringenderes Vorgehen gewesen auch außerhalb der Goethe-Universität zu rekrutieren, um einen größeren Pool an Veranstaltungen und final eine größere Stichprobe haben zu können. Aus forschungsökonomischen Gründen wurde in dieser explorativen Studie auf ein solches Vorgehen verzichtet. Zukünftige Forschung sollte diesen Punkt allerdings

berücksichtigen, auch wenn das aufgrund universitäts-unterschiedlichen Lehramtsausbildungen andere Problematiken mit sich bringen kann.

Des Weiteren ist es möglich, dass durch die Rekrutierung in den jeweiligen Veranstaltungen methodische Artefakte vorliegen, da es insgesamt zwei Dozierende gab, die mehrere ihrer Veranstaltungen zur Rekrutierung zur Verfügung stellten. Ein Dozierender ließ in drei seiner Seminare die Kontroll-Gruppen testen und eine weitere Dozentin bot die Möglichkeit in drei ihrer Seminare Teilnehmende zu gewinnen, von denen zwei konventionell und eines ein SL-Seminar war. Ursprünglich sollte es zwei SL-Seminare geben, eines musste jedoch umgewandelt werden, da der zivilgesellschaftliche Partner kurzfristig absprang. Aufgrund dieser Tatsache könnte eine Verzerrung in den Daten vorliegen, da die teilnehmenden Studierenden eigentlich eine SL-Veranstaltung besuchen wollten. Diesen Problemen könnte man in zukünftiger Forschung auch hier mit einer gezielten Befragung der Dozierenden entgegenwirken, um genauere Informationen über Seminarinhalte zu erhalten und dadurch Unterschiede in den verschiedenen Gruppen aufklären zu können.

Ein weiterer Aspekt, der bei der Auswertung der Ergebnisse wenig ins Gewicht fiel, war die Länge der Veranstaltungen. Es gab einige SL-Veranstaltungen, die über zwei Semester angelegt waren und einige, bei denen das nicht der Fall war. Die Auswertung ließ eine genauere Spezifikation nur schlecht zu. Die Studierenden wurden zwar danach befragt in welchem Seminar sie teilnahmen und wie viele Stunden Praxistätigkeit bereits erfolgt waren, aber nicht alle beantworteten diese Fragen, so dass eine Auswertung und der Einbezug der Informationen schwierig waren. Somit wurde sich bei der Auswertung an den Messzeitpunkten orientiert und nicht an der bereits geleisteten Praxistätigkeit in den Veranstaltungen. Die Länge der Veranstaltung spielt insofern eine Rolle, da bei Empathie-Trainings festgestellt werden konnte, dass vor allem mittellange Trainings im Umfang von 10-20 Stunden die besten Erfolge zeigten (vgl. Altmann, 2015). Da das Semester in der Regel 12 bis 15 Wochen dauert, aber die Service-Tätigkeit in SL-Seminaren meist erst mit Verzögerung beginnt, kann es durchaus sein, dass

Studierende weniger Praxis-Stunden hatten und eine optimale Trainingszeit gar nicht erreicht werden konnte, was auch eine fehlende Empathieentwicklung erklären könnte.

Erschwerend kommt auch hinzu, dass die SL-Seminare, in denen rekrutiert wurde, keinen expliziten Schwerpunkt auf das Thema Empathie hatten. In den bisherigen Studien, die signifikante Veränderungen zeigen konnten, wurde das Thema jedoch auch inhaltlich behandelt (Lundy, 2007; Wilson, 2011). Auch hier war der Versuch diese Information über die Selbstauskunft der Studierenden einzuholen mäßig erfolgreich. Die entsprechenden Items waren sehr allgemein formuliert und ließen, da keine Definition gegeben wurde, Spielraum zur Interpretation, was eigentlich mit Empathie oder dem Baustein Perspektivenübernahme gemeint sein könnte. Auch hier füllten nicht alle Studierenden die Frage aus und es wäre wiederum günstig gewesen diese Information mit den Dozierenden der Veranstaltungen abzugleichen, um diese Informationen angemessen berücksichtigen zu können. Ebenfalls zeigte sich bei einigen Studierenden, dass sie sich in Lehrveranstaltungen mit dem Thema beschäftigten, die keine SL-Veranstaltung waren. Aber was genau und in welchem Umfang konnte aufgrund der oben genannten Problematik nicht eruiert werden. Diese inhaltlichen Informationen hätten auch zu einer besseren Einschätzung führen können, ob es besondere Prompts zu dem Thema gab (sowohl in SL als auch in Non-SL-Veranstaltungen). Aus der Forschung ist bekannt, dass Personen mit geringem Vorwissen an ein Thema herangeführt werden müssen, um Wissen aufbauen zu können (Hmelo-Silver et al., 2007; Wild & Möller, 2009a). Es ist denkbar, dass das auch für die Entwicklung des Konstruktes Empathie eine relevante Rolle spielt. Ob das in den beforschten Veranstaltungen der Fall war, bleibt fraglich, da keine kognitiven Zuwächse oder Effekte erfasst wurden.

Ein weiterer methodischer Kritikpunkt ist wie gewissenhaft die ProbandInnen die Testbatterie bearbeitet haben. Gerade bei der Inter- und Post-Testung wurden die Bearbeitungsdauern kürzer und die gesehenen Videosequenzen des Vignettentests weniger häufig angeschaut als bei der Prä-Testung. Teilweise waren die Videoclips auch schon bekannt und vor allem V21 (die

Sequenz mit dem Brief *Desiree*) blieb vielen im Gedächtnis. Es ist denkbar, dass das einen Einfluss auf die Ergebnisse gehabt hat.

5.2 Fazit zur Entwicklung einer Testbatterie zur Erfassung von Empathie

Empathie ist ein mehrdimensionales Konstrukt, das aus kognitiven und affektiven Anteilen besteht, was die Erfassung komplex gestaltet. Die bislang genutzten Verfahren bringen in den meisten Fällen die Einschränkung mit, dass sie lediglich affektive oder kognitive Facetten berücksichtigen und in den meisten Fällen ein reines Selbstauskunftsverfahren sind. Im Vergleich zu objektiven oder objektiveren Messverfahren birgt diese Messmethode Verzerrungstendenzen und ist sehr stark von der Introspektionsfähigkeit der ProbandInnen abhängig (Schmidt-Atzert & Amelang, 2012). Im Bereich der Empathieerfassung kommen deswegen immer häufiger Methoden zum Einsatz, die mit zu bewertenden Videosequenzen arbeiten, um dadurch eine tatsächliche empathische Episode auszulösen (Decety & Ickes, 2011; Richter, 2009). Zusätzlich werden objektive Marker wie physiologisches Feedback und/oder mimische Aufnahmen erhoben, um eine Bewertung der erlebten Emotionen vornehmen zu können. Die Arbeit mit Videoaufnahmen ist auch in der Lehrerbildung ein gängiges Verfahren, um Wissen in komplexen Situationen anwenden und vertiefen zu können (Palmer et al., 2005; Sabers et al., 1991; Seidel et al., 2010; Stürmer et al., 2013). Dementsprechend erwies sich die Entwicklung eines videobasierten Vignettentests als ein ökonomisches und vielversprechendes Instrument, um als zusätzliche Informationsquelle zur Einschätzung von Empathie bei Lehramtsstudierenden zum Einsatz zu kommen. Zur Entwicklung gab es verschiedene Schritte, die zur Umsetzung herangezogen wurden. Hierbei wurden mit Hilfe von ExpertInnen inhaltsbasiert geeignete Videosequenzen ausgewählt, mit Fragen angereichert und anhand der gegebenen Antworten ein Auswertungsmanual erstellt. Nach den ersten Probedurchläufen wurden zusätzlich Gesichtsaufnahmen der ProbandInnen angefertigt, die mit der Emotionsanalyse-Software Face-Reader ausgewertet wurden. Neben dem videobasierten Vignettentest kamen zusätzlich noch Selbstauskunftsverfahren zum Einsatz, die neben

affektiven Empathiefacetten auch noch emotionale, sowie soziale Kompetenzen erfassen. Ebenfalls wurde Wissen über empathisches Verhalten im beruflichen Kontext erhoben.

Allgemein gesprochen hat die Testbatterie all die Bereiche abgedeckt, die abgedeckt werden sollten, allerdings gibt es vor allem beim Videovignettentest einige Einschränkungen, die im Verlauf sichtbar geworden sind und noch zusätzlicher Überarbeitung bedürfen. Um die Situationen in den Videosequenzen so real wie möglich abzubilden und trotzdem ökonomisch vorgehen zu können, wurden bereits existierende Videoaufnahmen anderer Schulforschungsprojekte nach Situationen durchsucht, die einen emotional eingefärbten Inhalt hatten. Aus anfangs 29 Videosequenzen wurden final 10 Videosituationen gewählt, die zum Einsatz kamen. Die in diesen Clips gesehenen Emotionen und Situationen wurden durch ExpertInnen eingeschätzt, was eine inhaltliche Auswertung ermöglichte. In diesem Vorgehen lässt sich allerdings auch eine der Schwachstellen festmachen. Empathie ist ein kommunikativer Prozess und die empathische Wahrnehmung bzw. Perspektivenübernahme der Erlebenswelt der anderen Person wird besser, wenn es von der betroffenen Person eine Rückmeldung zu dieser Einschätzung gibt (Barrett-Lennard, 1981; Kunyk & Olson, 2001). Dadurch kann die empathische Wahrnehmung noch einmal korrigiert und angepasst werden und wird somit genauer. Dieses Vorgehen findet man auch immer wieder in anderen Settings, bei denen dyadische Gesprächssituationen gefilmt und die involvierten Personen zu ihren Gedanken und Gefühlen in dieser Situation befragt werden (Decety & Ickes, 2011). Das hat den Vorteil, dass sofort eine Auswertungsmöglichkeit vorliegt, um eine Perspektivenübernahme vornehmen zu lassen, die auf Korrektheit geprüft werden kann. Bei diesem Vignettentest gab es keine Möglichkeit mehr, eine Rückmeldung von den SchülerInnen zu bekommen, da das Stimulusmaterial schon vorlag und teilweise vor über zehn Jahren aufgenommen wurde. Für Weiterentwicklungen bietet es sich an, eigene Aufnahmen anzufertigen, die dann bzgl. der durchlebten Gedanken und Gefühle direkt von den SchülerInnen kommentiert werden könnten. Da die Fähigkeit für bewusste Emotionen und eine

gewisse Reflexionsfähigkeit gegeben sein sollte, die sich erst zwischen sechs bis zwölf Jahren ausdifferenziert (vgl. Kullik & Petermann, 2012), wäre es sinnvoll ältere SchülerInnen zwischen zehn bis achtzehn Jahren für die Aufzeichnungen zu gewinnen. Die Situationen sollten real und nicht durch SchauspielerInnen gestellt sein, da echte emotionale Episoden Empathie zuverlässiger auslösen (vgl. Richter, 2009). In den hier genutzten Videosequenzen handelt es sich um authentische Emotionen, die allerdings dadurch bedingt auch recht schwach ausgeprägt sind. Diese Einschätzung wurde durch alle Entwicklungsschritte hindurch immer wieder gegeben, so dass am Ende die Vignetten mit den intensivsten emotionalen Anteilen im Test verblieben, die allerdings trotzdem recht schwache emotionale Situationen zeigten. Es erscheint sinnvoll, emotional gehaltvollere Situationen zu wählen und diese unter Umständen extra filmen und herstellen zu lassen. Hierbei sollte das Augenmerk darauf liegen, dass die Situationen als authentisch und echt erlebt werden, damit die affektiven Anteile von Empathie beim Zusehenden auch aktiviert werden können.

Ein weiteres wichtiges Merkmal, das in dieser Arbeit nur indirekt mit bedacht wurde, ist Interesse. Interesse scheint eine nicht unbedeutende Rolle im Zusammenhang mit Empathie zu spielen, da die intrapersonellen Prozesse der Aktivierung mentaler Modelle und dadurch ausgelöster adäquater Emotionen nur stattfinden, wenn die Aufmerksamkeit bei der auslösenden Person liegt und eine Motivation zum Verständnis der Erlebenswelt des Gegenübers gegeben ist (Cheng et al., 2007; Leventhal & Scherer, 1987; Rameson & Lieberman, 2009; Richter, 2009). Die persönliche Relevanz ist hierbei zentral. Im vorliegenden Fall wird diese Variable nur indirekt bei der Auswertung der Videoaufnahmen des Gesichts der ProbandInnen durch Face-Reader einbezogen. Bei der Auswertung und Angabe der Intensität der Emotionen wird die Position des Kopfes mit einberechnet. Blick abwenden oder Ändern der Blickrichtung durch Kopfdrehungen hat einen Einfluss auf die Ergebnisse der Analysen. Da es allerdings auch denkbar ist, dass ProbandInnen beim direkten Anschauen eigentlich kein Interesse an der Person oder der Situation haben, wäre es bei einer Weiterentwicklung sinnvoll

das Interesse direkt oder indirekt zu erfragen. Das erscheint über eine Selbstauskunft denkbar oder über entsprechend formulierte Fragen, die die entwickelten mentalen Modelle der Perspektivenübernahme erfragen. Bei vorheriger Befragung der SchülerInnen in den Videos über ihre erlebten Emotionen und Gedanken in den aufgenommenen Situationen könnte man dann die Antworten der ProbandInnen sehr viel valider bewerten und müsste nicht auf die Fremdeinschätzungen von ExpertInnen zurückgreifen, die möglicherweise nicht ganz stimmen. Folglich könnte Interesse etwas klarer mit einbezogen und berücksichtigt werden. Dieses Vorgehen hätte darüber hinaus den Vorteil, dass man durch die explizit gemachte Perspektivenübernahme und objektive Erfassung von emotionalen Reaktionen eine differenzierte Aussage tätigen kann, ob es sich bei dem erlebten Phänomen tatsächlich um Empathie oder aber eher um selbst zentrierte Reaktionen wie Gefühlsansteckung handelt (Rameson & Lieberman, 2009).

Bei Empathie handelt es sich um ein Konstrukt, das sowohl State-, als auch Trait-Aspekte beinhaltet. In der Forschung zeigt sich, dass State-Anteile den Bereich ausmachen, der durch Trainings entwickelt und verändert werden kann (Butters, 2010). Einige der eingesetzten Verfahren und Instrumente erwiesen sich bereits in früheren Testungen zu Empathie als änderungssensitiv und sind somit geeignet um State-Aspekte abzubilden (Altmann, 2015), weswegen sie auch in dieser Studie zum Einsatz kamen. Allerdings wurden zusätzlich noch Verfahren wie der *TEQ* (Spreng et al., 2009) genutzt, die auf Trait-Anteile von Empathie abzielen. Eine höhere Ausrichtung an und bessere Erfassung von State-Anteilen hätte bei der längsschnittlichen Messung erreicht werden können, wenn in der *TEQ*-Instruktion (bzw. ab Messzeitpunkt *Inter* in allen Testinstruktionen) eine zeitliche Begrenzung mit eingeflossen wäre. So hätte die Aufforderung nicht mehr darin bestanden, seine generellen Fähigkeiten einzuschätzen, sondern die der letzten beiden Wochen, des letzten Monats o.ä. (Butters, 2010; Schmidt-Atzert & Amelang, 2012). Es ist denkbar, dass sich dadurch andere Ergebnisse gezeigt hätten.

Die auch im Praxiseinsatz festgestellten niedrigen internen Konsistenzwerte im Vignettentest könnten auf zwei Ursachen zurückzuführen sein: Zum einen auf die Multidimensionalität der Situationen in den Videoclips, als auch auf methodische Problematiken in der Testentwicklung. Dadurch dass in den Videosequenzen komplexe Unterrichts- bzw. Klassensituationen gezeigt werden, umfassen diese Interaktionen natürlicherweise mehr an interpersonellen Geschehnissen als nur empathische bzw. affektive Episoden. Allein dadurch entsteht eine Verzerrung der Daten (Davis, 1980). Vergleicht man die Reliabilitätswerte der Testungen von Studierenden verschiedener Studienrichtungen mit den Ergebnissen der Erhebungen mit Lehramtsstudierenden zeigen sich bei den Lehramtsstudierenden niedrigere Werte. Es ist denkbar, dass die Videosequenzen im Sinne einer professionellen Wahrnehmung für Lehramtsstudierende mehr verwertbare Informationen enthalten als für andere Studierende. Lehramtsstudierende nehmen die Informationen in den Sequenzen sehr viel differenzierter wahr und müssen dementsprechend mehr Informationen verarbeiten (Goodwin, 1996; Seidel et al., 2010). Das bedeutet, dass sich die Bereiche *noticing* und *knowledge-based reasoning* bei Studierenden unterschiedlicher Fächer jeweils auf andere Bereiche konzentrieren und es allen Studierenden außer Lehramtsstudierenden leichter fallen könnte, die nur für Empathie und die Items relevanten Informationen herauszufiltern und zu verarbeiten. Bei Lehramtsstudierenden dürfte diese Verarbeitung etwas komplexer ausfallen und vermutlich auch einen Einfluss auf die Bewertung der Situation haben. Dementsprechend kann es für zukünftige Forschung eine Überlegung sein, ebenfalls Vignetten mit Kontexten außerhalb des Unterrichts zu nutzen (gleiche Altersgruppe wie die der SchülerInnen, aber Setting außerhalb der Schule).

Die niedrigeren Werte des E- und Vscore betreffend könnte dies auf die Konstruktion der Items zurückzuführen sein. Bei der Auswertung der Emotionsqualitätsantwort und der Verhaltensantworten hätte man optimalerweise noch weitere Auswertende hinzugezogen. Die Komprimierung der Antworten erfolgte über die Autorin, wurde allerdings über nächste Entwicklungsschritte mit weiteren ExpertInnen zumindest evaluiert. Günstig wäre gewesen

tatsächlich eine Delphi-Befragung im klassischen Sinne durchzuführen und die extrahierten Antwortformate nochmals den beteiligten ExpertInnen vorzulegen (Mayring, 2015). Das hätte zwar die Anonymisierung der Befragung erschwert und unter Umständen negative Auswirkungen auf die Partizipation gehabt, trotzdem hätte man validere Ergebnisse erhalten können. Ebenfalls hätte man die Antworten des Items zum empathischen Verhalten nochmal kritisch prüfen sollen. Die Rückmeldungen einiger Studierenden zum letzten Messzeitpunkt zeigten, dass die Antworten teilweise die Lösung von Item eins enthielten, was trotz sorgfältiger Prüfungen im Vorfeld nicht aufgefallen war. Des Weiteren wäre es wünschenswert gewesen weitere Validierungen mit größeren Stichproben durchzuführen und auch die Zusammenhänge zwischen anderen performanzbasierten Emotionserkennungsverfahren und dem Vignettentest weiter zu prüfen. Durch die hohe Anzahl an Vignetten wurden mehrere Testversionen gebildet, um eine ausreichende Menge an Informationen zu erhalten. Das sorgte trotz einer ausreichenden Partizipation für recht kleine Stichprobengrößen. Da nun nur noch eine Anzahl von zehn Vignetten verblieben ist, ließen sich diese mit größeren Stichproben validieren.

Wie sich in der Durchführung zeigte, erweist sich das Erfassen der Facette *Mitschwingen* als herausfordernd. Obwohl mit Hilfe der Software Face-Reader eine ökonomische, objektive Auswertungsmöglichkeit der Emotionen der Teilnehmenden beim Anschauen des Stimulusmaterials zur Verfügung stand, sind die Ergebnisse nicht eindeutig und nur eingeschränkt interpretierbar. Es gibt vor allem in der Prä-Version emotions-kongruente Zusammenhänge zwischen Selbstauskunft und Face-Reader Analysen, allerdings nicht über alle Emotionen und Vignetten hinweg. Zum Post-Messzeitpunkt werden die Effekte noch weniger. Die Gründe hierfür lassen sich sicherlich wie bereits erwähnt auf der methodischen Ebene im Stimulusmaterial und auch in den angefertigten Videoaufnahmen der Teilnehmenden finden. Für zukünftige Forschung sollte dem Rechnung getragen und das Stimulusmaterial für die videobasierten Vignettentests kontextspezifischer ausgewählt bzw. hergestellt werden, um eindeutige Aussagen zur empathischen Akkuratheit bereits im Vorfeld generieren zu können.

Alles in allem lässt sich allerdings sagen, dass die Erhebung von Empathie im Rahmen einer vielfältigen Testbatterie einen guten Zugang zur Erfassung dieses multidimensionalen Konstruktes darstellt. Das Umsetzen objektiver Verfahren zum Erfassen der Performanz bspw. in Form eines Videovignettentests erscheint nach wie vor als ein geeignetes Vorgehen. Hierbei sollte der Entwicklungsphase viel Zeit eingeräumt werden, um einen stabilen und validen Test zu konzipieren, der sich situationsübergreifend für verschiedenste Gruppen von Erwachsenen einsetzen lässt. In der Konstruktionsphase sollten bereits Analysen mit Face-Reader umgesetzt werden. Die Erfahrung mit dieser Studie zeigt, dass selbst mit dieser Erhebungsmethode Gruppensettings umsetzbar sind, so dass ökonomisch machbar eine ausreichende Stichprobengröße erreicht werden kann. Im Vergleich zu reinen Selbstaussagenverfahren zur Messung von Empathie ist der Aufwand höher, erscheint allerdings lohnenswert, um Empathie umfassend zu erheben.

Literaturverzeichnis

- ABL - Akademie für Bildungsforschung und Lehrerbildung. (2018a). *Goethe-Universität — Akademie für Bildungsforschung und Lehrerbildung*. Zugriff am 04.12.2018. Verfügbar unter http://www.uni-frankfurt.de/62115214/Akademie_f%C3%BCr_Bildungsforschung_und_Lehrerbildung
- ABL - Akademie für Bildungsforschung und Lehrerbildung. (2018b). Prüfungsordnung für den bildungswissenschaftlichen Studienanteil. Zugriff am 04.12.2018. Verfügbar unter http://www.uni-frankfurt.de/65107035/Fachspezifische_Anh%C3%A4nge_und_Modulscheine_L1
- ABL - Akademie für Bildungsforschung und Lehrerbildung. (2018c). *Service Learning in der ABL*. Zugriff am 04.12.2018. Verfügbar unter http://www.uni-frankfurt.de/70268208/Service_Learning
- Adolphs, R., Tranel, D. & Damasio, A. R. (2003). Dissociable neural systems for recognizing emotions. *Brain and Cognition*, 52(1), 61–69. [https://doi.org/10.1016/S0278-2626\(03\)00009-5](https://doi.org/10.1016/S0278-2626(03)00009-5)
- Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J. & Sanford, R. N. (1950). *The authoritarian personality*. New York: Harper.
- Albiero, P., Matricardi, G., Speltri, D. & Toso, D. (2009). The assessment of empathy in adolescence: A contribution to the Italian validation of the "Basic Empathy Scale". *Journal of Adolescence*, 32(2), 393–408. <https://doi.org/10.1016/j.adolescence.2008.01.001>
- Alexander, K. L. (2001). *Prosocial behaviors of adolescents in work and family life: Empathy and conflict resolution strategies with parents and peers*. Unveröffentlichte Dissertation. Ohio State University.
- Allport, G. W., Vernon, P. E. & Lindzey, G. (1951). *Manual: Study of values*. Boston: Houghton Mifflin.
- Altenschmidt, K. (Hrsg.). (2009). *Raus aus dem Elfenbeinturm? Entwicklungen in Service Learning und bürgerschaftlichem Engagement an deutschen Hochschulen*. Weinheim u.a.: Beltz.
- Altmann, T. (2013). *Funktionale Empathie: Entwicklung und Evaluation eines Empathietrainings auf Basis eines integrativen Prozessmodells zur Vermeidung empathisch kurzschlüssigen Handelns*. Unveröffentlichte Dissertation. Universität Duisburg-Essen. Zugriff am 19.04.2016. Verfügbar unter https://duepublico.uni-duisburg-essen.de/servlets/DerivateServlet/Derivate-34497/DISS.Text_UB_online_2013-12-12.pdf
- Altmann, T. (2015). *Empathie in sozialen und Pflegeberufen. Entwicklung und Evaluation eines Trainingsprogramms*. Wiesbaden: Imprint: Springer.
- Altmann, T. & Roth, M. (2013). The evolution of empathy: From single components to process models. In C. Mohiyeddini, M. Eysenck & S. Bauer (Hrsg.), *Psychology of emotions* (S. 171–187). New York: Nova Science Publishers.
- Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention. Mit 58 Tabellen; [mit 3 neuen Anwendungsfeldern]* (Springer-Lehrbuch, 4., vollst. überarb. und erw. Aufl.). Heidelberg: Springer.
- Asendorpf, J. B. (2007). *Psychologie der Persönlichkeit*. Heidelberg: Springer.
- Ashton, M. C. & Lee, K. (2009). The HEXACO-60: a short measure of the major dimensions of personality. *Journal of Personality Assessment*, 91, 340–345. <https://doi.org/10.1080/00223890902935878>

- Aspy, D. N. (1969). The effect of teacher-offered conditions of empathy, positive regard, and congruence upon student achievement. *Florida Journal of Educational Research*, *11*(1), 39–48.
- Aznar, A. & Tenenbaum, H. R. (2015). Gender and age differences in parent-child emotion talk. *The British Journal of Developmental Psychology*, *33*, 148–155. <https://doi.org/10.1111/bjdp.12069>
- Backhaus-Maul, H. & Roth, C. (2013). *Service Learning an Hochschulen in Deutschland*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-00124-7>
- Baker, S. B., Daniels, T. G. & Greeley, A. T. (1990). Systematic Training of Graduate-Level Counselors: Narrative and Meta-Analytic Reviews of Three Major Programs. *The Counseling Psychologist*, *18*, 355–421.
- Baldwin, S. C., Buchanan, A. M. & Rudisill, M. E. (2007). What Teacher Candidates Learned About Diversity, Social Justice, and Themselves From Service-Learning Experiences. *Journal of Teacher Education*, *58*, 315–327. <https://doi.org/10.1177/0022487107305259>
- Balu & Du. (2018). *Das MentorInnenprogramm Balu und Du*. Zugriff am 26.11.2018. Verfügbar unter <https://www.balu-und-du.de/home/>
- Bandura, A. & Walters, R. H. (1963). *Social learning and personality development*. New York: Holt Rinehart and Winston.
- Baron-Cohen, S. (1995). *Mindblindness. An essay on autism and theory of mind*. Cambridge, Mass: MIT Press.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry*, *42*, 241–251. <https://doi.org/10.1017/S0021963001006643>
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17. <https://doi.org/10.1023/A:1005653411471>
- Barrett-Lennard, G. T. (1981). The empathy cycle: Refinement of a Nuclear concept. *Journal of Counseling Psychology*, *28*, 91–100.
- Barron, F. (1954). *Charades as an assessment device* (Air Force Personnel, Hrsg.) (Technical Memorandum Project No. 7730). Alabama: Officer Education Research Laboratory, Maxwell Air Force Base.
- Bartens, W. (2017). *Empathie. Die Macht des Mitgefühls: weshalb einfühlsame Menschen gesund und glücklich sind*. München: Knauer.
- Bass, B. M. (1955). Authoritarianism or acquiescence? *Journal of Abnormal and Social Psychology*, *51*, 616–623.
- Batchelder, T. H. & Root, S. (1994). Effects of an undergraduate program to integrate academic learning and service: cognitive, prosocial cognitive, and identity outcomes. *Journal of adolescence*, *17*, 341–355.
- Batson, C. D. (1987). Prosocial Motivation: Is it ever Truly Altruistic? In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 20 (Advances in Experimental Social Psychology, vol. 20, pp. 65–122). San Diego: Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60412-8](https://doi.org/10.1016/S0065-2601(08)60412-8)
- Batson, C. D. (1997). Self-other merging and the empathy-altruism hypothesis: Reply to Neuberg et al. (1997). *Journal of Personality and Social Psychology*, *73*, 517–522. <https://doi.org/10.1037/0022-3514.73.3.517>

- Batson, C. D. (2011). These Things Called Empathy: Eight Related but Distinct Phenomena. In J. Decety & W. Ickes (Hrsg.), *The social neuroscience of empathy* (S. 3–15). Cambridge, Mass. u.a.: MIT Press.
- Batson, C. D., Batson, J. G., Slingsby, J. K., Harrell, K. L., Peekna, H. M. & Todd, R. M. (1991). Empathic joy and the empathy-altruism hypothesis. *Journal of Personality and Social Psychology*, *61*, 413–426. <https://doi.org/10.1037/0022-3514.61.3.413>
- Batson, C. D., Fultz, J. & Schoenrade, P. A. (1987). Distress and Empathy: Two Qualitatively Distinct Vicarious Emotions with Different Motivational Consequences. *Journal of Personality*, *55*, 19–39. <https://doi.org/10.1111/j.1467-6494.1987.tb00426.x>
- Batson, C. D., Sager, K., Garst, E., Kang, M., Rubchinsky, K. & Dawson, K. (1997). Is Empathy-Induced Helping Due to Self-Other Merging? *Journal of Personality and Social Psychology*, *73*, 495–509.
- Bellmann, J. (2014). *Nicht-intendierte Effekte Neuer Steuerung im Schulsystem - Theoretische Konzeptualisierung und Instrumententwicklung zur empirischen Erfassung (NeFo)*, BMBF-Bundesministerium für Bildung und Forschung. Zugriff am 05.05.2019. Verfügbar unter <https://www.empirische-bildungsforschung-bmbf.de/de/418.php>
- Berking, M. (2015). *Training emotionaler Kompetenzen* (Psychotherapie, 3., vollst. überarb. Auflage). Berlin: Springer. <https://doi.org/10.1007/978-3-642-54017-2>
- Bernhardt, D. (2008). *Service Learning. Aus Sicht der Studierenden*. Saarbrücken: VDM Verl. Müller.
- Beven, J. P., O'Brien-Malone, A. & Hall, G. (2004). Using the Interpersonal Reactivity Index to Assess Empathy in Violent Offenders. *International Journal of Forensic Psychology*, *1*, 33–41.
- Bierhoff, H.-W. (2005). Prosoziales Verhalten. In H. Weber, T. Rammsayer & J. Bengel (Hrsg.), *Handbuch der Persönlichkeitspsychologie und differentiellen Psychologie* (Handbuch der Psychologie, / hrsg. von J. Bengel ... ; Bd. 2, S. 457–466). Göttingen: Hogrefe.
- Biggs, J. B. & Tang, C. (2011). *Teaching for quality learning at university. What the student does*. Berkshire: McGraw-Hill.
- Billig, S. H. & Weah, W. (2008). K-12 Service Learning Standards for Quality Practice. In *Growing to Greatness: The State of Service Learning*, (S. 8–15). Saint Paul, Minnesota. National Youth Leadership Council.
- Bird, G. & Viding, E. (2014). The self to other model of empathy: providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience and Biobehavioral Reviews*, *47*, 520–532. <https://doi.org/10.1016/j.neubiorev.2014.09.021>
- Bischof-Köhler, D. (2009). Empathie, Mitgefühl und Grausamkeit. Und wie sie zusammenhängen. *Psychotherapie*, *14*(1), 52–57.
- Blanca, M. J., Alarcón, R., Arnau, J., Bono, R. & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, *29*, 552–557. <https://doi.org/10.7334/psicothema2016.383>
- Block, J. (1961). *The Q-sort method in personality assessment and psychiatric research*. Springfield: Charles C Thomas Publisher. <https://doi.org/10.1037/13141-000>
- Boeckler, A., Tusche, A. & Singer, T. (2016). The Structure of Human Prosociality. Differentiating Altruistically Motivated, Norm Motivated, Strategically Motivated, and Self-Reported Prosocial Behavior. *Social Psychological and Personality Science*, *7*, 530–541. <https://doi.org/10.1177/1948550616639650>

- Bonn, J., Helbrich, h., Lohan, M., Peindl, S. & Schabert, J. (2008). *Evaluation eines Fragebogens zur Erfassung von Empathie gegenüber Tieren*. Universität Regensburg.
- Bowlby, J. (1969). *Attachment and Loss: Attachment* (Bd. 1). New York: Basic Books.
- Brown, J. S., Collins, A. & Duguid, P. (1989). Situated Cognition and the Culture of Learning. *Educational Researcher*, 18, 32–42.
- Brown, M. A. (2011). Learning From Service: The Effect of Helping on Helpers' Social Dominance Orientation. *Journal of Applied Social Psychology*, 41, 850–871. <https://doi.org/10.1111/j.1559-1816.2011.00738.x>
- Buccino, G., Binkofski, F., Fink, G. R., Fadiga, I., Fogassi, L., Gallese, V. et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, 13, 400–404.
- Burgess, P. W., Alderman, N., Evans, J. J., Wilson, B. A. & Emslie, H. (1996). The Dysexecutive Questionnaire. In P. W. Burgess, H. Emslie & J. J. Evans (Hrsg.), *Behavioral assessment of the dysexecutive syndrome*. Bury St. Edmunds, England: Thames Valley Test.
- Butters, R. P. (2010). *Meta-analysis of empathy training programs for client populations*. Unveröffentlichte Dissertation. University of Utah.
- Campus Compact. (2018, November 23). *Campus Compact*. Verfügbar unter <https://compact.org/>
- Cappella, J. N. & Planalp, S. (1981). Talk and Silence Sequences in Informal Conversations III: Interspeaker Influence. *Human Communication Research*, 7, 117–132.
- Cattell, R. B. & Cattell, A. K. S. (1959). *The Culture Fair Intelligence Test. A Measure of "g"*. Champaign, Illinois, USA: Institute for Personality and Ability Testing.
- Celio, C. I., Durlak, J. & Dymnicki, A. (2011). A Meta-analysis of the Impact of Service-Learning on Students. *Journal of Experiential Education*, 34, 164–181. <https://doi.org/10.5193/JEE34.2.164>
- Chartrand, T. L. & Bargh, J. A. (1999). The Chameleon Effect: The PErception-Behavior Link and Social Interaction. *Journal of Personality and Social Psychology*, 76, 893–910.
- Cheng, Y., Meltzoff, A. N. & Decety, J. (2007). Motivation modulates the activity of the human mirror-neuron system. *Cerebral Cortex*, 17, 1979–1986. <https://doi.org/10.1093/cercor/bhl107>
- Cialdini, R. B., Brown, S. L., Lewis, B. P., Carol, L. & Neuberg, S. L. (1997). Reinterpreting the empathy-altruism relationship: When one into one equals oneness. *Journal of Personality and Social Psychology*, 73, 481–494. <https://doi.org/10.1037/0022-3514.73.3.481>
- Cierpka, M. (2004). Das Fördern der Empathie bei Kindern mit FAUSTLOS. *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie (GIO)*, 35, 37–50. <https://doi.org/10.1007/s11612-004-0004-y>
- Cliffordson, C. (2002). The hierarchical structure of empathy: Dimensional organization and relations to social functioning. *Scandinavian Journal of Psychology*, 43, 49–59.
- Cochin, S., Barthelemy, C., Lejeune, B., Roux, S. & Martineau, J. (1998). Perception of motion and qEEG activity in human adults. *Electroencephalography and Clinical Neurophysiology*, 107, 287–295. [https://doi.org/10.1016/S0013-4694\(98\)00071-6](https://doi.org/10.1016/S0013-4694(98)00071-6)
- The Cognition and Technology Group at Vanderbilt. (1990). Anchored Instruction and Its Relationship to Situated Cognition. *Educational Researcher*, 19(6), 2–10. Zugriff am 26.11.2018.
- Cohen, B. H. (2001). *Explaining psychological statistics* (2.ed.). New York: Wiley. Retrieved from <http://www.loc.gov/catdir/bios/wiley042/00036655.html>

- Combs, A. W., Soper, D. W., Gooding, C. T., Benton, J. A., Dickman, J. F. & Usher, R. H. (1969). *Florida studies in the helping professions*. Gainesville, Florida: University of Florida Press.
- Conway, J. M., Amel, E. L. & Gerwien, D. P. (2009). Teaching and Learning in the Social Context: A Meta-Analysis of Service Learning's Effects on Academic, Personal, Social, and Citizenship Outcomes. *Teaching of Psychology*, 36, 233–245. <https://doi.org/10.1080/00986280903172969>
- Cooper, B. (2011). *Empathy in education. Engagement, values and achievement*. London: Continuum. Retrieved from <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10505738>
- Corcoran, K. J. (1982). An exploratory investigation into self-other differentiation Empirical evidence for a monistic perspective on empathy. *Psychotherapy: Theory, Research and Practice*, 19, 63–68.
- Costanzo, M. & Archer, B. (1994). *The Interpersonal Perception Task 15 (IPT– 15): A guide for researchers and teachers*. Berkeley, CA: University of California Center for Media and Independent Learning.
- Cross, D. G. & Sharpley, C. F. (1982). Measurement of Empathy with the Hogan Empathy Scale. *Psychological Reports*, 50, 62. <https://doi.org/10.2466/pr0.1982.50.1.62>
- Crowne, D. P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349–354. <https://doi.org/10.1037/h0047358>
- Cunico, L., Sartori, R., Marognolli, O. & Meneghini, A. M. (2012). Developing empathy in nursing students: A cohort longitudinal study. *Journal of Clinical Nursing*, 21, 2016–2025.
- Cusi, A. M., Macqueen, G. M., Spreng, R. N. & McKinnon, M. C. (2011). Altered empathic responding in major depressive disorder: relation to symptom severity, illness burden, and psychosocial outcome. *Psychiatry Research*, 188, 231–236. <https://doi.org/10.1016/j.psychres.2011.04.013>
- Danner, D. D., Snowdon, D. A. & Friesen, W. V. (2001). Positive emotions in early life and longevity: Findings from the nun study. *Journal of Personality and Social Psychology*, 80, 804–813. <https://doi.org/10.1037//0022-3514.80.5.804>
- Davis, M. H. (1980). *A Multidimensional Approach to Individual Differences in Empathy*. Austin: The University of Texas.
- Davis, M. H. (1983). Measuring individual Differences in Empathy: Evidence for a Multidimensional Approach. *Journal of Personality and Social Psychology*, 44, 113–126.
- Davis, M. H., Hull, J. G., Young, R. D. & Warren, G. G. (1987). Emotional reactions to dramatic film stimuli: The influence of cognitive and emotional empathy. *Journal of Personality and Social Psychology*, 52, 126–133. <https://doi.org/10.1037/0022-3514.52.1.126>
- Decety, J. & Ickes, W. (Hrsg.). (2011). *The social neuroscience of empathy*. Cambridge, Mass. u.a.: MIT Press.
- Decety, J. & Moriguchi, Y. (2007). The empathic brain and its dysfunction in psychiatric populations: implications for intervention across different clinical conditions. *BioPsychoSocial Medicine*, 1, 22. <https://doi.org/10.1186/1751-0759-1-22>
- Dewey, J. (1916). *Democracy and education : an introduction to the philosophy of education*. New York: Macmillan.
- Didaktische Werkstatt. (2018). *Goethe-Universität — Praxisprojekte*. Zugriff am 04.12.2018. Verfügbar unter <https://www.uni-frankfurt.de/49228126/Praxisprojekte>

- Dimberg, U. (1990). Facial electromyography and emotional reactions. *Psychophysiology*, 27, 481–494. Zugriff am 24.10.2018.
- Dimberg, U., Thunberg, M. & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science*, 11, 86–89. <https://doi.org/10.1111/1467-9280.00221>
- Dixon, W. R. & Morse, W. C. (1961). The Prediction of Teaching Performance: Empathic Potential. *Journal of Teacher Education*, 12, 322–329. Zugriff am 07.12.2018.
- Drummond, M. J. (1993). *Assessing children's learning*. London: Fulton.
- Dubs, R. (1995). *Lehrerverhalten. Ein Beitrag zur Interaktion von Lehrenden und Lernenden im Unterricht*. Zürich: Verlag:SKV. Zugriff am 07.12.2018.
- Duden. *Synonyme. Online-Wörterbuch*. Zugriff am September 2016. Verfügbar unter <https://www.duden.de/hilfe/synonyme>
- Dudenredaktion (Hrsg.). (2017). *Duden, die deutsche Rechtschreibung. Das umfassende Standardwerk auf der Grundlage der neuen amtlichen Regeln ; rund 135000 Stichwörter mit über 500000 Beispielen, Bedeutungserklärungen und Angaben zu Worttrennung, Aussprache, Grammatik, Stilebenen und Etymologie* (Der Duden in zwölf Bänden, Bd. 1, 27., völlig neu bearb. und erw. Aufl.). Berlin: Dudenverlag. Zugriff am 21.08.2018. Verfügbar unter <https://www.duden.de/suchen/dudenonline/Empathie>
- Dymond, R. F. (1949). A scale for the measurement of empathic ability. *Journal of Consulting Psychology*, 13, 127–133. <https://doi.org/10.1037/h0061728>
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.
- EHEA. (1999). *The bologna Declaration*. THE EUROPEAN HIGHER EDUCATION AREA. Zugriff am 10.07.2017. Verfügbar unter <https://www.ehea.info/pid34363/ministerial-declarations-and-communications.html>
- EHEA. (2003). *“Realising the European Higher Education Area” - Communiqué of the Conference of Ministers responsible for Higher Education in Berlin*. Bologna Prozess. Zugriff am 10.07.2017. Verfügbar unter <https://www.ehea.info/pid34363/ministerial-declarations-and-communications.html>
- EHEA. (2007). *London Communiqué -Towards the European Higher Education Area: responding to challenges in a globalised world*. Bologna Process. Zugriff am 10.07.2017. Verfügbar unter <https://www.ehea.info/pid34363/ministerial-declarations-and-communications.html>
- EHEA. (2012). *Making the Most of Our Potential: Consolidating the European Higher Education Area - Bucharest Communiqué*. EHEA. Zugriff am 14.07.2017. Verfügbar unter <https://www.ehea.info/pid34363/ministerial-declarations-and-communications.html>
- Eisenberg, N. (1986). *Altruistic Emotion, cognition, and behavior*. Hillsdale, NJ: Erlbaum.
- Eisenberg, N., Cumberland, A. & Spinrad, T. L. (1998). Parental Socialization of Emotion. *Psychological Inquiry*, 9, 241–273. https://doi.org/10.1207/s15327965pli0904_1
- Eisenberg, N., Guthrie, I. K., Murphy, B. C., Shepard, S. A., Cumberland, A. & Carlo, G. (1999). Consistency and Development of Prosocial Dispositions: A Longitudinal Study. *Child Development*, 70, 1360–1372. <https://doi.org/10.1111/1467-8624.00100>
- Eisenberg, N. & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin*, 94, 100–131. <https://doi.org/10.1037/0033-2909.94.1.100>
- Eisenberg, N. & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin*, 101, 91–119. <https://doi.org/10.1037//0033-2909.101.1.91>

- Eisenberg, N., Shea, C. L., Carlo, G. & Knight, G. P. (1991). Empathy related responding and cognition: A "Chicken and the Egg" Dilemma. In W. M. Kurtines, J. Gewirtz & J. L. Lamb (Hrsg.), *Handbook of Moral Behavior and Development*. Hillsdale, NJ: Erlbaum.
- Eisenberg, N. & Strayer, J. (1987). Critical issues in the study of empathy. In N. Eisenberg & J. Strayer (Hrsg.), *Cambridge studies in social and emotional development. Empathy and its development* (S. 3–13). New York: Cambridge university press.
- Eisenberger, N. I. (2012). The pain of social disconnection: examining the shared neural underpinnings of physical and social pain. *Nature Reviews. Neuroscience*, 13, 421–434. <https://doi.org/10.1038/nrn3231>
- Eisenberger, N. I. & Lieberman, M. D. (2004). Why rejection hurts: a common neural alarm system for physical and social pain. *Trends in Cognitive Sciences*, 8, 294–300. <https://doi.org/10.1016/j.tics.2004.05.010>
- Eisenberger, N. I., Lieberman, M. D. & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science*, 302, 290–292. Zugriff am 24.10.2018.
- Ekman, P., Levenson, R. & Friesen, W. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, 221(4616), 1208–1210. <https://doi.org/10.1126/science.6612338>
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6, 169–200. <https://doi.org/10.1080/02699939208411068>
- Erickson, J. & Anderson, J. B. (Eds.). (2012). *Learning With the Community. Concepts and Models for Service Learning in Teacher Education* (Service Learning in the Disciplines Series). Herndon: Stylus Publishing. Retrieved from <http://gbv.ebib.com/patron/FullRecord.aspx?p=4438581>
- Erpenbeck, J., Scharnhorst, A., Ebeling, W., Martens, D., Nachtigall, C., North, K. et al. (2006). *Metakompetenzen und Kompetenzentwicklung* (Arbeitsgemeinschaft Betriebliche Weiterbildungsforschung e. V./Projekt Qualifikations-Entwicklungs-Management, Hrsg.) (Teil I Nr. 95). Berlin. Zugriff am 26.11.2018.
- Eyler, J. & Giles, D. (1999). *Where's the learning in service-learning?* (The Jossey-Bass higher and adult education series, 1st ed). San Francisco: Jossey-Bass.
- Eysenck, H. J. (1959). *The Maudsley Personality Inventory*. London: University of London Press.
- Ferrari, J. R. & Worrall, L. (2000). Assessments by Community Agencies: How "the Other Side" Sees Service-Learning. *Michigan Journal of Community Service Learning*, 7, 35–40. Zugriff am 30.11.2018.
- Feshbach, N. D. & Roe, K. (1968). Empathy in Six- and Seven-Year-Olds. *Child Development*, 39, 133–145. Zugriff am 12.11.2018.
- Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics. And sex and drugs and rock 'n' roll* (4th ed.). Los Angeles [etc.]: Sage.
- Floyd, K., Pauley, P. M. & Hesse, C. (2010). State and Trait Affectionate Communication Buffer Adults' Stress Reactions. *Communication Monographs*, 77, 618–636. <https://doi.org/10.1080/03637751.2010.498792>
- Frith, C. D. & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50, 531–534. <https://doi.org/10.1016/j.neuron.2006.05.001>
- Gerdes, K. E., Lietz, C. A. & Segal, E. A. (2011). Measuring Empathy in the 21st Century: Development of an Empathy Index Rooted in Social Cognitive Neuroscience and Social Justice. *Social Work Research*, 35, 83–93. <https://doi.org/10.1093/swr/35.2.83>
- Gerholz, K.-H., Liszt, V. & Klingsieck, K. B. (2015). Didaktische Gestaltung von Service Learning – Ergebnisse einer Mixed Methods-Studie aus der Domäne der

- Wirtschaftswissenschaften. *bwp-Berufs- und Wirtschaftspädagogik-Online*, (28), 1–23. Zugriff am 12.08.2015. Verfügbar unter http://www.bwpat.de/ausgabe28/gerholz_etal_bwpat28.pdf
- Gignac, G. E. & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78. <https://doi.org/10.1016/j.paid.2016.06.069>
- Goodwin, C. (1996). Professional Vision. *American Anthropologist*, 96, 606–633. Zugriff am 20.08.2015.
- Gough, H. G. (1956). California Psychological Inventory. Palo Alto, CA, England: Consulting Psychologists Press.
- Govekar, M. A. & Rishi, M. (2007). Service Learning: Bringing Real-World Education Into the B-School Classroom. *Journal of Education for Business*, 83, 3–10. <https://doi.org/10.3200/JOEB.83.1.3-10>
- Häcker, Thomas, Hilzensauer, Wolf, Reinmann & Gabi. (2008). Editorial zum Schwerpunktthema 'Reflexives Lernen'. *Bildungsforschung*, 5, 1–4. Zugriff am 18.10.2017.
- Händel, M. (2018). Intelligenz, kristalline und fluide. In M. A. Wirtz (Hrsg.), *Dorsch – Lexikon der Psychologie*. Zugriff am 08.10.2018. Verfügbar unter <https://m.portal.hogrefe.com/dorsch/intelligenz-kristalline-und-fluide/>
- Hartig, J. & Jude, N. Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (S. 17–36). Bundesministerium für Bildung und Forschung.
- Hashimoto, H. & Shiomi, K. (2002). The structure of empathy in Japanese adolescents: Construction and examination of an empathy scale. *Social Behavior and Personality*, 30, 593–602.
- Hatfield, E., Cacioppo, J. T. & Rapson, R. L. (1993). Emotional Contagion. *Current Directions in Psychological Science*, 2, 96–99. Zugriff am 16.11.2018.
- Hatfield, E., Cacioppo, J. T. & Rapson, R. L. (1994). *Emotional Contagion*. Cambridge, England: Cambridge university press.
- Hatfield, E., Hsee, C. K. & Costello, J. (1995). The Impact of Vocal Feedback on Emotional Experience and Expression. *Journal of Social Behavior and Personality*, 10, 293–312. Zugriff am 20.11.2018.
- Hatfield, E., Rapson, R. L. & Le Yen-Chi, L. (2009). Emotional Contagion and Empathy. In J. Decety & W. J. Ickes (Eds.), *The social neuroscience of empathy* (Social neuroscience series, pp. 19–30). Cambridge, Mass.: MIT Press; ProQuest.
- Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota multiphasic personality inventory*, Rev. Minneapolis, MN, US: University of Minnesota Press.
- Haugen, P. T., Welsh, D. P. & McNulty, J. K. (2008). Empathic accuracy and adolescent romantic relationships. *Journal of Adolescence*, 31, 709–727. <https://doi.org/10.1016/j.adolescence.2008.03.003>
- Heinrichs, M., Baumgartner, T., Kirschbaum, C. & Ehlert, U. (2003). Social support and oxytocin interact to suppress cortisol and subjective responses to psychosocial stress. *Biological Psychiatry*, 54, 1389–1398. [https://doi.org/10.1016/S0006-3223\(03\)00465-7](https://doi.org/10.1016/S0006-3223(03)00465-7)
- Helmke, A. (2008). *Unterrichtsbeobachtung - VERA*. Zugriff am 05.05.2019. Verfügbar unter <https://www.fdz-bildung.de/studiendetails.php?la=de&id=17> <https://doi.org/10.7477/20:1:1>
- Helmke, A. & Lenske, G. (2013). Unterrichtsdiagnostik als Voraussetzung für Unterrichtsentwicklung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 31, 214–233. Zugriff am 17.07.2017.

- Hepper, E. G., Hart, C. M. & Sedikides, C. (2014). Moving Narcissus: Can Narcissists Be Empathic? *Personality & Social Psychology Bulletin*, 40, 1079–1091. <https://doi.org/10.1177/0146167214535812>
- Hessische Lehrkräfteakademie. (2018). *CP Erste Staatsprüfung / Hessische Lehrkräfteakademie*. Zugriff am 04.12.2018. Verfügbar unter <https://lehrkraefteakademie.hessen.de/lehrausbildung/erste-staatspruefung>
- Hey, D. W., Slivovsky, L. A., Self, B. P., Widmann, J. & Taylor, J. K. (2014). Learning design through the lens of service-a qualitative study. *International Journal of Service Learning in Engineering*, 9(1), 1–23.
- Hickock, G. (2008). Eight Problems for the Mirror Neuron Theory of Action Understanding in Monkeys and Humans. *Journal of Cognitive Neuroscience*, 21, 1229–1243.
- Hickok, G. (2015). *Warum wir verstehen, was andere fühlen. Der Mythos der Spiegelneuronen*. München: Hanser.
- Hill, C. E. & Lent, R. W. (2006). A narrative and meta-analytic review of helping skills training: Time to revive a dormant area of inquiry. *Psychotherapy (Chicago, Ill.)*, 43, 154–172. <https://doi.org/10.1037/0033-3204.43.2.154>
- Hmelo-Silver, C. E., Duncan, R. G. & Chinn, C. A. (2007). Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42, 99–107. <https://doi.org/10.1080/00461520701263368>
- HMWK (Hessisches Ministerium für Wissenschaft und Kunst, Hrsg.). (2018). *Studium in Hessen*. Zugriff am 22.11.2018. Verfügbar unter <https://lehrkraefteakademie.hessen.de/lehrausbildung/studium-hessen>
- Hochschulnetzwerk BDV-Bildung durch Verantwortung. (2018a). *Kommentierung der Qualitätskriterien von Service Learning an Hochschulen*.
- Hochschulnetzwerk BDV-Bildung durch Verantwortung. (2018b). *Willkommen - Bildung durch Verantwortung*. Zugriff am 23.11.2018. Verfügbar unter <https://www.bildung-durch-verantwortung.de/>
- Höder, J., Tausch, R. & Weber, A. (1976). Förderliche Dimensionen des Lehrerverhaltens und ihr Zusammenhang mit der Qualität der Unterrichtsbeiträge der Schüler. *Im Manuskript*, 103, 120–134.
- Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, 33, 307–316. <https://doi.org/10.1037/h0027580>
- Hojat, M., Mangione, S., Gonella, J. S., Nasca, T., Veloski, J. J. & Kane, G. (2001). Empathy in medical education and patient care. *Academic Medicine*, 76, 669.
- Holt-Lunstad, J., Smith, T. B. & Layton, B. (2010). Social relationships and mortality risk: a metaanalytic review. *PLoS Medicine*, July(7), 1–20. <https://doi.org/10.1371/journal.pmed.1000316.g001>
- Hornak, J., Rolls, E. T. & Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioral changes following ventral frontal lobe damage. *Neuropsychologia*, 34, 247–261.
- HRK-Hochschulrektorenkonferenz (Hrsg.). (2014). *Service Learning – Lernen durch Engagement. Von der Idee zur Umsetzung* (Projekt nexus: Konzepte und gute Praxis für Studium und Lehre). Zugriff am 23.11.2018.
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C. & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, 3, 0529–0535. <https://doi.org/10.1371/journal.pbio.0030079>
- Ickes, W. (Hrsg.). (1997). *Empathic accuracy*. New York u.a.: Guilford Press.

- Ickes, W. (2003). *Everyday mind reading: Understanding what other people think and feel*. Amherst, NY: Prometheus Books.
- Ickes, W., Bissonette, V., Garcia, S. & Stinson, L. L. (1990). Implementing and using the dyadic interaction paradigm. In C. Hendrick & M. S. Clark (Hrsg.), *Research methods in personality and social psychology* (S. 16–44). Thousand Oaks, CA: Sage.
- Ickes, W., Gesn, P. R. & Graham, T. (2000). Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships*, 7, 95–109. Zugriff am 25.01.2016.
- Ireland, J. L. (1999). Provictim attitudes and empathy in relation to bullying behaviour among prisoners. *Legal and Criminological Psychology*, 4, 51–66. <https://doi.org/10.1348/135532599167789>
- Jackson, A. L. & Bonacker, N. (2006). The Effect of Victim Impact Training Programs on the Development of Guilt, Shame and Empathy Among Offenders. *International Review of Victimology*, 13(3), 301–324. <https://doi.org/10.1177/026975800601300304>
- Jahoda, G. (2005). Theodor Lipps and the shift from "sympathy" to "empathy". *Journal of the History of the Behavioral Sciences*, 41, 151–163. <https://doi.org/10.1002/jhbs.20080>
- Jarrell, K., Ozmy, J., Gallagher, J., Hagler, D., Corral, C. & Hagler, A. (2014). Constructing the foundations for compassionate care: how service-learning affects nursing students' attitudes towards the poor. *Nurse Education in Practice*, 14, 299–303. <https://doi.org/10.1016/j.nepr.2013.11.004>
- John, O. P., Donahue, E. M. & Kentle, R. L. (1991). *The Big Five Inventory -Versions 4a and 54*. Berkeley, CA: University of California, Berkeley.
- John, O. P. & Srivastava, S. (1999). The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality. Theory and research* (2nd ed., pp. 102–138). New York, NY: Guilford Press.
- Jolliffe, D. & Farrington, D. P. (2004). Empathy and offending. A systematic review and meta-analysis. *Aggression and Violent Behavior*, 9, 441–476. <https://doi.org/10.1016/j.avb.2003.03.001>
- Jolliffe, D. & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence*, 29, 589–611. <https://doi.org/10.1016/j.adolescence.2005.08.010>
- Jones, S. R. & Abes, E. S. (2004). Enduring Influences of Service-Learning on College Students' Identity Development. *Journal of College Student Development*, 45, 149–166. <https://doi.org/10.1353/csd.2004.0023>
- Jones, S. R. & Hill, K. (2001). Crossing High Street: Understanding Diversity through Community Service-Learning. *Journal of College Student Development*, 42, 204–216.
- Junghenn, A. (2016). *Eigenschaften von Service Learning Veranstaltungen an Hochschulen zur Steigerung von Empathiefähigkeit und Selbstwirksamkeit studentischer Teilnehmer - Eine Expertenbefragung*. Bachelorarbeit. Johann-Wolfgang-Goethe-Universität, Frankfurt am Main.
- Kagan, N. & Kratwohl, D. R. (1967). *Studies in Human Interactions*. East Lansing, Michigan: Michigan State University. Verfügbar unter <https://files.eric.ed.gov/fulltext/ED017946.pdf>
- Kanning, U. P. (2009). *ISK - Inventar Sozialer Kompetenzen*. Göttingen: Hogrefe.
- Kerkenbusch, M. (2012). *Empathie: Ein notwendiges Persönlichkeitsmerkmal bei Lehrern?, Psychologie - Lernpsychologie, Intelligenzforschung, GRIN*. Examensarbeit. Zugriff am 20.08.2015.
- Keyesers, C. & Gazzola, V. (2014). Dissociating the ability and propensity for empathy. *Trends in Cognitive Sciences*, 18, 163–166. <https://doi.org/10.1016/j.tics.2013.12.011>

- Klein, K. J. K. & Hodges, S. D. (2001). Gender Differences, Motivation, and Empathic Accuracy: When it Pays to Understand. *Personality and Social Psychology Bulletin*, 27, 720–730. <https://doi.org/10.1177/0146167201276007>
- Klieme, E. & Hartig, J. (2008). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. In M. Prenzel, I. Gogolin & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik. Zeitschrift für Erziehungswissenschaft* (Zeitschrift für Erziehungswissenschaft Sonderheft, Bd. 8, S. 11–29). Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH Wiesbaden. https://doi.org/10.1007/978-3-531-90865-6_2
- Konopka, M. (2017). *Empathiefähigkeit verschiedener Studiengänge im Vergleich*. Bachelorarbeit. Goethe-Universität, Frankfurt am Main.
- Kravas, C. H. (1975). Using Videotape Vignettes to Measure Teacher Affective Sensitivity. *Educational Technology*. Zugriff am 09.02.2018.
- Kullik, A. & Petermann, F. (2012). *Emotionsregulation im Kindesalter* (Klinische Kinderpsychologie, Band 14). Göttingen: Hogrefe. Verfügbar unter <http://www.socialnet.de/rezensionen/isbn.php?isbn=978-3-8017-2434-4>
- Kultusministerkonferenz. (2014). *Standards Lehrerbildung: Bildungswissenschaften*. Berlin, Bonn.
- Kunter, M. & Pohlmann, B. (2009). Lehrer. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (Springer-Lehrbuch, S. 261–282). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.
- Kunyk, D. & Olson, J. K. (2001). Clarification of conceptualizations of empathy. *Journal of Advanced Nursing*, 35, 317–325. <https://doi.org/10.1046/j.1365-2648.2001.01848.x>
- Lamm, C., Decety, J. & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage*, 54, 2492–2502. <https://doi.org/10.1016/j.neuroimage.2010.10.014>
- Langstraat, L. & Bowdon, M. (2011). Service-Learning and Critical Emotion Studies: On the Perils of Empathy and the Politics of Compassion. *Michigan Journal of Community Service Learning*, 17(2), 5–14.
- Lauterbach, O. & Hosser, D. (2007). Assessing Empathy in Prisoners - A Shortened Version of the Interpersonal Reactivity Index. *Swiss Journal of Psychology*, 66(2), 91–101. <https://doi.org/10.1024/1421-0185.66.2.91>
- Lee, K. & Ashton, M. C. (2004). Psychometric Properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research*, 39, 329–358. https://doi.org/10.1207/s15327906mbr3902_8
- Levenson, R. W. & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology*, 63, 234–246. <https://doi.org/10.1037//0022-3514.63.2.234>
- Leventhal, H. & Scherer, K. (1987). The Relationship of Emotion to Cognition. A Functional Approach to a Semantic Controversy. *Cognition & emotion*, 1, 3–28. <https://doi.org/10.1080/02699938708408361>
- Lewinski, P., den Uyl, T. M. & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, 7, 227–236. <https://doi.org/10.1037/npe0000028>
- Lipowsky, F. (2009). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (Springer-Lehrbuch, S. 74–101). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.

- Lipowsky, F. & Klieme, E. (2006). *Leistungstests* (Materialien zur Bildungsforschung, Bd. 14). Frankfurt am Main: GFFP. Verfügbar unter <http://nbn-resolving.de/urn:nbn:de:0111-opus-31078>
- Lorenz, C. (2011). *Diagnostische Kompetenz von Grundschullehrkräften: strukturelle Aspekte und Bedingungen*: University of Bamberg Press.
- Lundy, B. L. (2007). Service Learning in Life-Span Developmental Psychology: Higher Exam Scores and Increased Empathy. *Teaching of Psychology*, 34, 23–27. <https://doi.org/10.1080/00986280709336644>
- MacDonald, D. K. (2018, 2. November). *Carkhuff & Truax 1967 - Empathy Training Lit Review*. Zugriff am 09.11.2018. Verfügbar unter <https://sites.google.com/site/empathytraininglitreview/measurements/carkhuff-truax-1967>
- Manera, V., Grandi, E. & Colle, L. (2013). Susceptibility to emotional contagion for negative emotions improves detection of smile authenticity. *Frontiers in Human Neuroscience*, 7, 6. <https://doi.org/10.3389/fnhum.2013.00006>
- Marshall, J. H., Lawrence, E. C., Lee Williams, J. & Peugh, J. (2015). Mentoring as service-learning. The relationship between perceived peer support and outcomes for college women mentors. *Studies in Educational Evaluation*, 47, 38–46. <https://doi.org/10.1016/j.stueduc.2015.07.001>
- Martin, L. J., Hathaway, G., Isbester, K., Mirali, S., Acland, E. L., Niederstrasser, N. et al. (2015). Reducing social stress elicits emotional contagion of pain in mouse and human strangers. *Current Biology : CB*, 25, 326–332. <https://doi.org/10.1016/j.cub.2014.11.028>
- Matthews, K. A., Batson, C. D., Horn, J. & Rosenman, R. H. (1981). "Principles in his nature which interest him in the fortune of others...": The heritability of empathic concern for others. *Journal of Personality*, 49, 237–247. <https://doi.org/10.1111/j.1467-6494.1981.tb00933.x>
- Maurer, R. E. & Tindall, J. H. (1983). Effect of Postural Congruence on Client's Perception of Counselor Empathy. *Journal of Counseling Psychology*, 30, 158–163.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *The American Psychologist*, 59, 14–19. <https://doi.org/10.1037/0003-066X.59.1.14>
- Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (Beltz Pädagogik, 12., überarb. Aufl.). Weinheim: Beltz.
- McCrae, R. R., Kurtz, J. E., Yamagata, S. & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review : an Official Journal of the Society for Personality and Social Psychology, Inc*, 15, 28–50. <https://doi.org/10.1177/1088868310366253>
- Mead, G. H. (1934). *Mind Self and Society: from the Standpoint of a Social Behaviorist*. (Edited by Charles W. Morris). Chicago: University of Chicago.
- Mehrabian, A. (2000). *Manual for the Balanced Emotional Empathy Scale (BEES)*. Unveröffentlichtes Manuskript, Monterey, CA.
- Mehrabian, A. & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality*, 40, 525–543.
- Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations*, 18, 57–76.
- Moosbrugger, H. (2002). *Lineare Modelle. Regressions- und Varianzanalysen ; mit einem Anhang über Matrixalgebra, 22 Tabellen, Beispielen und Prüfverteilungen* (Methoden der Psychologie, Bd. 14, 3., gründlich überarb. und erg. Aufl.). Bern: Huber.

- Moosbrugger, H. & Kelava, A. (2007). *Testtheorie und Fragebogenkonstruktion* (Springer-Lehrbuch). Heidelberg: Springer.
- Moshagen, M., Hilbig, B. E. & Zettler, I. (2014). Faktorenstruktur, psychometrische Eigenschaften und Messinvarianz der deutschsprachigen Version des 60-Item HEXACO Persönlichkeitsinventars. *Diagnostica*, 60, 86–97. <https://doi.org/10.1026/0012-1924/a000112>
- Myers, I. B. (1962). *Manual for the Myers-Briggs Type Indicator*. Princeton.
- Neuberg, S. L., Cialdini, R. B., Brown, S. L., Luce, C. & et al. (1997). Does empathy lead to anything more than superficial helping? Comment on Batson et al. (1997). *Journal of Personality and Social Psychology*, 73, 510–516. <https://doi.org/10.1037/0022-3514.73.3.510>
- 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(1), 1-21.
- Nickerson, R. S. (1999). How we know - and sometimes misjudge - what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, 125, 737–759.
- Novak, J. M., Markey, V. & Allen, M. (2007). Evaluating Cognitive Outcomes of Service Learning in Higher Education. A Meta-Analysis. *Communication Research Reports*, 24, 149–157. <https://doi.org/10.1080/08824090701304881>
- Office 2013 [Computer software]: Microsoft.
- Oswald, P. (1996). The effects of cognitive and affective perspective taking on empathic concern and altruistic... *The Journal of Social Psychology*, 136, 613–623.
- Palagi, E., Norscia, I. & Demuru, E. (2014). Yawn contagion in humans and bonobos: emotional affinity matters more than species. *PeerJ*, 2, 1-17. <https://doi.org/10.7717/peerj.519>
- Palmer, D. J., Stough, L. M., Burdenski, J. T. K. & Gonzales, M. (2005). Identifying Teacher Expertise: An Examination of Researchers' Decision Making. *Educational Psychologist*, 40, 13–25. https://doi.org/10.1207/s15326985ep4001_2
- Papenfuß, U., Redlich, M., Steinhauer, L. & Friedländer, B. (2015). *Forschend und engagiert lernen im Public Management: Befunde und Gestaltungsanregungen eines Service Learning Lehrforschungsprojektes* (Nr. 140). Leipzig: Universität Leipzig, Wirtschaftswissenschaftliche Fakultät.
- Patterson, C. H. (1985). *The Therapeutic Relationship*. Monterey, Calif.: Brooks/Cole.
- Paulus, C. (2009). *Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie*. Universität des Saarlandes.
- Perrone-McGovern, K. M., Oliveira-Silva, P., Simon-Dack, S., Lefdahl-Davis, E., Adams, D., McConnell, J. et al. (2014). Effects of empathy and conflict resolution strategies on psychophysiological arousal and satisfaction in romantic relationships. *Applied Psychophysiology and Biofeedback*, 39, 19–25. <https://doi.org/10.1007/s10484-013-9237-2>
- Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J. et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, 389, 495–498. <https://doi.org/10.1038/39051>
- Piaget, J. & Inhelder, B. (1969). *The psychology of the child*. London: Routledge & Keagan Paul.
- Piaget, J. & Inhelder, B. (1999). *Die Entwicklung des räumlichen Denkens beim Kinde*. Stuttgart: Klett-Cotta.
- Potthoff, D. E., Dinsmore, J. A., Stirtz, G., Walsh, T., Ziebarth, J. & Eifler, K. (2000). Preparing for Democracy and Diversity: The Impact of a Community-Based Field Experience on

- Preservice Teachers' Knowledge, Skills, and Attitudes. *Action in Teacher Education*, 22, 79–92. <https://doi.org/10.1080/01626620.2000.10462995>
- Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515–526. <https://doi.org/10.1017/S0140525X00076512>
- Prentice, M. & Robinson, G. (2010). *Improving Student Learning Outcomes with Service-Learning*. American Association of Community Colleges.
- Preston, S. D. & Waal, F. B. M. de. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25, 1–72.
- Pritchard, I. A. (2002). Community Service and Service-Learning in America: The state of the Art. In A. Furco & S. Billig (Eds.), *Service-learning. The essence of the pedagogy* (Advances in service-learning, pp. 3–21). Greenwich, Conn.: Information Age.
- Putnam, R. T. & Borko, H. (2000). What Do New Views of Knowledge and Thinking Have to Say About Research on Teacher Learning? *Educational Researcher*, 29(3), 4–14.
- Rameson, L. T. & Lieberman, M. D. (2009). Empathy. A Social Cognitive Neuroscience Approach. *Social and Personality Psychology Compass*, 3, 94–110. <https://doi.org/10.1111/j.1751-9004.2008.00154.x>
- Rammstedt, B. & Danner, D. (2016). Die Facettenstruktur des Big Five Inventory (BFI). *Diagnostica*, 63, 70–84. <https://doi.org/10.1026/0012-1924/a000161>
- Rammstedt, B. & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K). *Diagnostica*, 51, 95–206. <https://doi.org/10.1026/0012-1924.51.4.195>
- Reinders, H. (2010). Lernprozesse durch Service Learning an Universitäten. *Zeitschrift für Pädagogik*, 56, 531–547.
- Reinders, H. (2016). *Service Learning - Theoretische Überlegungen und empirische Studien zu Lernen durch Engagement*. Weinheim: Beltz Juventa.
- Rempala, D. M. (2013). Cognitive strategies for controlling emotional contagion. *Journal of Applied Social Psychology*, 43, 1528–1537. <https://doi.org/10.1111/jasp.12146>
- Renkl, A. (2009). Wissenswerb. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (Springer-Lehrbuch, S. 4–27). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.
- Reynolds, W. (2000). *The measurement and development of empathy in nursing*. Aldershot, England: Ashgate.
- Richter, D. (2009). *Empathie: Gewinne und Verluste im Erwachsenenalter*. Dissertation. Bremen.
- Rindermann, H. (2009). *EKF: Emotionale-Kompetenz-Fragebogen. Einschätzung emotionaler Kompetenzen und emotionaler Intelligenz aus Selbst- und Fremdsicht*. Göttingen: Hogrefe Verl. für Psychologie.
- Rizzolatti, G. & Sinigaglia, C. (2008). *Empathie und Spiegelneurone. Die biologische Basis des Mitgefühls*. Frankfurt am Main: Suhrkamp.
- Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, 21, 95–103. <https://doi.org/10.1037/h0045357>
- Rogers, C. R. (1975). Empathic: An Unappreciated Way of Being. *The Counseling Psychologist*, 5(2), 2–10.
- Rogers, C. R. & Carmichael, L. (Hrsg.). (1951). *Client--Centered Therapy: Its Current Practice, Implications, and Theory*. Boston: Houghton Mifflin Company.
- Rogers, C. R. & Dymond, R. F. (Hrsg.). (1954). *Psychotherapy and personality change*. Chicago, IL,US: University of Chicago Press.
- Rokeach, M. (1960). *The open and closed mind*. New York: Basic Books.
- Root, S. (2012). School-Based Service: A Review of Research for Teacher Educators. In J. Erickson & J. B. Anderson (Eds.), *Learning With the Community. Concepts and Models for*

- Service Learning in Teacher Education* (Service Learning in the Disciplines Series, 42-72). Herndon: Stylus Publishing.
- Rosenberg, M. B. (2016). *Gewaltfreie Kommunikation. Eine Sprache des Lebens*. Paderborn: Junfermann Verlag; ProQuest.
- Rosengren, A., Hawken, S., Ôunpuu, S., Sliwa, K., Zubaid, M., Almahmeed, W. A. et al. (2004). Association of psychosocial risk factors with risk of acute myocardial infarction in 11 119 cases and 13 648 controls from 52 countries (the INTERHEART study): case-control study. *The Lancet*, 364, 953–962. [https://doi.org/10.1016/S0140-6736\(04\)17019-0](https://doi.org/10.1016/S0140-6736(04)17019-0)
- Rump, J. & Eilers, S. (2006). Managing Employability. In T. Sattelberger, H. Fischer & J. Rump (Hrsg.), *Employability Management. Grundlagen, Konzepte, Perspektiven* (1. Aufl., S. 13–73). Wiesbaden: Gabler.
- Sabers, D. S., Cushing, K. S. & Berliner, D. C. (1991). Differences among Teachers in a Task Characterized by Simultaneity, Multidimensionality, and Immediacy. *American Educational Research Journal*, 28, 63-88. <https://doi.org/10.2307/1162879>
- Schluß, H. & Jehle, M. (2013). *Quellensicherung und Zugänglichmachung von Videoaufzeichnungen von Unterricht der Freien Universität Berlin*. Zugriff am 05.05.2019. Verfügbar unter <https://www.fdz-bildung.de/erhebung.php?id=34&la=de> <https://doi.org/10.7477/19:1:0>
- Schmidt-Atzert, L. & Amelang, M. (2012). *Psychologische Diagnostik* (5., vollst. überarb. und erw. Aufl.). Berlin u.a.: Springer.
- Schönefeld, V., Deckers, M., Altmann, T. & Roth, M. (2016). *Deutsche Version des Toronto Empathy Questionnaire (TEQ-D) nach R. N. Spreng, M. C. McKinnon, R. A. Mar, & B. Levine.*, Universität Duisburg-Essen.
- Schönefeld, V. & Roth, M. (2017). *Essener Fragebogen zur Erfassung des subjektiven Empathiekonzeptes - Version für LehrerInnen (ESEK-L)*, Essen, Deutschland.
- Schönefeld, V. & Roth, M. (2018, September). *Entwicklung und Validierung einer deutschen Version des Toronto Empathy Questionnaire (TEQ-D)*. Vortrag auf dem 51. Kongress der Deutschen Gesellschaft für Psychologie (DGPs), Frankfurt am Main.
- Schubarth, W. & Speck, K. (2013). *Employability und Praxisbezüge im wissenschaftlichen Studium* (HRK-Hochschulrektorenkonferenz, Hrsg.). Zugriff am 23.03.2015. Verfügbar unter http://www.hrk-nexus.de/fileadmin/redaktion/hrk-nexus/07-Downloads/07-02-Publikationen/Fachgutachten_Employability.pdf
- Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D., Cooper, J. T., Golden, C. J. et al. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences*, 25, 167–177.
- Seidel, T., Blomberg, G. & Stürmer, K. (2010). "Observer" - Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt OBSERVE. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (Leitschrift für Pädagogik, Beiheft, Bd. 56). Weinheim, Basel: Beltz.
- Seidel, T. & Prenzel, M. (2007). Wie Lehrpersonen Unterricht wahrnehmen und einschätzen – Erfassung pädagogisch-psychologischer Kompetenzen bei Lehrpersonen mit Hilfe von Videosequenzen. In M. Prenzel, I. Gogolin & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik. Zeitschrift für Erziehungswissenschaft* (Sonderheft 8, S. 201–218). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Seifert, A. (2011). *Resilienzförderung an der Schule. Eine Studie zu Service-Learning mit Schülern aus Risikolagen* (Schule und Gesellschaft, Bd. 52). Trier: VS Verlag.

- Seifert, A., Zentner, S., Nagy, F., Baltes, A. M. & Sliwka, A. (2012). *Praxisbuch Service-Learning. "Lernen durch Engagement" an Schulen*. Weinheim: Beltz.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (Hrsg.). (2004). *Bildungsstandards der Kultusministerkonferenz. Erläuterungen zur Konzeption und Entwicklung*. München, neuwied: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Wolters Kluwer Deutschland GmbH.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Sherin, M. G. (2002). When Teaching Becomes Learning. *Cognition and Instruction*, 20, 119–150.
- Shirtcliff, E. A., Vitacco, M. J., Graf, A. R., Gostisha, A. J., Merz, J. L. & Zahn-Waxler, C. (2009). Neurobiology of empathy and callousness: implications for the development of antisocial behavior. *Behavioral Sciences & the Law*, 27, 137–171. <https://doi.org/10.1002/bsl.862>
- Sigmon, R. L. (1996). *Journey to Service-Learning. Experiences from Independent Liberal Arts Colleges and Universities*. Washington, DC: The Council of Independent Colleges.
- Sigmon, R. L. (1997). *Linking Service with Learning in Liberal Arts Education*. Washington, DC: Council of Independent Colleges.
- Siu, A. M. H. & Shek, D. T. L. (2005). Validation of the Interpersonal Reactivity Index in a Chinese Context. *Research on Social Work Practice*, 15, 118–126. <https://doi.org/10.1177/1049731504270384>
- Soukup, P. A. (1999). *Soukup, Paul A. Assessing Service-Learning in a Communication Curriculum*. Paper presented at the Annual Meeting of the National Communication Association, Chicago, Illinois.
- Spreng, R. N., McKinnon, M. C., Mar, R. A. & Levine, B. (2009). The Toronto Empathy Questionnaire: scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of Personality Assessment*, 91, 62–71. <https://doi.org/10.1080/00223890802484381>
- SPSS Statistics (Version 24) [Computer software]: IBM. Verfügbar unter <https://www.ibm.com/de-de/marketplace/spss-statistics>
- Steibe, S. C., Boulet, D. B. & Lee, D. C. (1979). Trainee Trait Empathy, Age, trainer Functioning, CLient Age and Training Time as Disrminators of Successful Empathy Training. *Canadian Counsellor*, 14, 41–46.
- Steins, G. (2005). Empathie. In H. Weber, T. Rammsayer & J. Bengel (Hrsg.), *Handbuch der Persönlichkeitspsychologie und differentiellen Psychologie* (Bd. 2, S. 467–475). Göttingen: Hogrefe.
- Stepien, K. A. & Baernstein, A. (2006). Educating for empathy. A review. *Journal of General Internal Medicine*, 21, 524–530. <https://doi.org/10.1111/j.1525-1497.2006.00443.x>
- Stiftung Lernen durch Engagement. (2018, 27. November). *Netzwerk Lernen durch Engagement*. Verfügbar unter <https://www.servicelearning.de/>
- Stone, V. E. (2006). Theory of mind and the evolution of social intelligence. In J. T. Cacioppo, P. S. Visser & C. L. Pickett (Hrsg.), *Social Neuroscience: People thinking about thinking people* (S. 103–129). Cambridge, MA: MIT Press.
- Stone, V. E. & Gerans, P. (2006). What's domain-specific about theory of mind). *Social Neuroscience*, 1, 309–319.

- Stürmer, K., Seidel, T. & Schäfer, S. (2013). Changes in professional vision in the context of practice. *Gruppendynamik und Organisationsberatung*, 44, 339–355. <https://doi.org/10.1007/s11612-013-0216-0>
- Tamblyn, R., Abrahamowicz, M., Dauphinee, D., Wenghofer, E., Jacques, A., Klass, D. et al. (2007). Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *JAMA*, 298, 993–1001. <https://doi.org/10.1001/jama.298.9.993>
- Tausch, R. (2008). Personzentriertes Verhalten von Lehrern in Unterricht und Erziehung. In Schweer, Martin K. W. (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (Schule und Gesellschaft, Bd. 24, 2., vollst. überarb. Aufl, 155). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Tausch, R. & Tausch, A.-M. (1998). *Erziehungspsychologie. Begegnung von Person zu Person* (11., korrigierte Aufl.). Göttingen u.a.: Hogrefe, Verl. für Psychologie.
- Taylor, J. A. (1953). A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology*, 48, 285–290.
- Teichler, U. (2013). Hochschule und Arbeitswelt. Theoretische Überlegungen, politische Diskurse und empirische Befunde. In G. Hessler, M. Oechsle & I. Scharlau (Hrsg.), *Studium und Beruf: Studienstrategien – Praxiskonzepte – Professionsverständnis: Perspektiven von Studierenden und Lehrenden nach der Bologna-Reform* (S. 21–38). Bielefeld: Transcript.
- Titchener, E. B. (1912). The Schema of Introspection. *The American Journal of Psychology*, 23, 485–508..
- Truax, C. B. (1967). The training of nonprofessional personnel in therapeutic interpersonal relationships. *American Journal of Public Health and the Nations Health*, 57, 1778–1791.
- Truax, C. B. (1972). The Meaning and Reliability of Accurate Empathy Ratings. A Rejoinder. *Psychological Bulletin*, 77, 397–399.
- Twenge, J. M., Baumeister, R. F., DeWall, C. N., Ciarocco, N. J. & Bartels, J. M. (2007). Social exclusion decreases prosocial behavior. *Journal of Personality and Social Psychology*, 92, 56–66. <https://doi.org/10.1037/0022-3514.92.1.56>
- U.S. Department of Education. (2001). No Child Left Behind (the Elementary and Secondary Education Act). Verfügbar unter <https://www2.ed.gov/policy/elsec/leg/esea02/index.html>. Zugriff am 12.05.2019.
- Ulich, D. & Volland, C. (1998). Erfassung und Korrelate von Mitgefühl bei Erwachsenen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 30, 89–97.
- Underwood, B. & Moore, B. (1982). Perspective-taking and altruism. *Psychological Bulletin*, 91, 143–173. <https://doi.org/10.1037/0033-2909.91.1.143>
- Unipark [Computer software]: Questback.
- Van Baaren, R. B., Decety, J., Dijksterhuis, A., van der Leij, A. & van Leeuwen, M. L. (2009). Being Imitated: Consequences of Nonconsciously Showing Empathy. In J. Decety & W. J. Ickes (Eds.), *The social neuroscience of empathy* (Social neuroscience series, pp. 31–42). Cambridge, Mass.: MIT Press; ProQuest.
- Voss, T., Kunter, M. & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge. Test construction and validation. *Journal of Educational Psychology*, 103, 952–969. <https://doi.org/10.1037/a0025125>
- Vries, R. E. de. (2013). The 24-item Brief HEXACO Inventory (BHI). *Journal of Research in Personality*, 47, 871–880. <https://doi.org/10.1016/j.jrp.2013.09.003>
- Wade, R. C. (1995). *Developing Active Citizens: Community Service Learning in Social Studies Teacher Education* (University of Nebraska Omaha, Hrsg.) (General Paper 36).

- Wang, Y., Davidson, M., Yakushko, O. F., Savoy, H. B., Tan, J. A. & Bleier, J. K. (2003). The scale of ethnocultural empathy: Development, validation, and reliability. *Journal of Counseling Psychology*, *50*, 221–234.
- Warren, J. (2012). Does Service-Learning Increase Student Learning?: A Meta-Analysis. *Michigan Journal of Community Service Learning*, *18*, 56–61.
- Weiß, R. H. (2006). *CFT-20R mit WS/ZF-R. Grundintelligenztest Skala 2 - Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgetest - Revision (WS/ZF-R)*. Göttingen u.a.: Hogrefe.
- Wied, M. de, Branje, S. J. T. & Meeus, Wim H J. (2007). Empathy and conflict resolution in friendship relations among adolescents. *Aggressive Behavior*, *33*, 48–55. <https://doi.org/10.1002/ab.20166>
- Wild, E. & Möller, J. (Hrsg.). (2009a). *Pädagogische Psychologie* (Springer-Lehrbuch). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-88573-3>
- Wild, E. & Möller, J. (2009b). *Pädagogische Psychologie. Mit ... 27 Tabellen* (Springer-Lehrbuch). Heidelberg: Springer.
- Wilson, J. C. (2011). Service-learning and the development of empathy in US college students. *Education + Training*, *53*, 207–217. <https://doi.org/10.1108/00400911111115735>.
- Wirtz, M. & Kutschmann, M. (2007). Analyse der Beurteilerübereinstimmung für kategoriale Daten mittels Cohens Kappa und alternativer Masse. *Die Rehabilitation*, *46*, 370–377. <https://doi.org/10.1055/s-2007-976535>
- Wirtz, M. A. (Hrsg.). (2018). *Dorsch – Lexikon der Psychologie*. Zugriff am 08.10.2018 unter <https://m.portal.hogrefe.com/dorsch/intelligenz-kristalline-und-fluide/>.
- Wispé, L. (1986). The distinction between sympathy and empathy: To call forth a concept, a word is needed. *Journal of Personality and Social Psychology*, *50*, 314–321. <https://doi.org/10.1037//0022-3514.50.2.314>
- Wolter, A. & Banscherus, U. (2012). Praxisbezug und Beschäftigungsfähigkeit im Bologna-Prozess – “A never ending story“? In W. Schubarth, K. Speck, A. Seidel, C. Gottmann, C. Kamm & M. Krohn (Hrsg.), *Studium nach Bologna: Praxisbezüge stärken?! Praktika als Brücke zwischen Hochschule und Arbeitsmarkt* (S. 21–36). Wiesbaden: Springer VS.
- Yorio, P. L. & Ye, F. (2012). A Meta-Analysis on the Effects of Service-Learning on the Social, Personal, and Cognitive Outcomes of Learning. *Academy of Management Learning & Education*, *11*, 9–27. <https://doi.org/10.5465/amle.2010.0072>
- Young, R. A. & Collin, A. (2004). Introduction: Constructivism and social constructionism in the career field. *Journal of Vocational Behavior*, *64*, 373–388. <https://doi.org/10.1016/j.jvb.2003.12.005>
- Zahn-Waxler, C., Radke-Yarrow, M. & Wagner, E. (1992). Development of Concern for Others. *Developmental Psychology*, *28*, 126–136.
- Zahn-Waxler, C., Schiro, K., Robinson, J. L., Emde, R. N. & Schmitz, S. (2001). Empathy and Prosocial Patterns in Young MZ and DZ Twins. Development and Genetic and Environmental Influences. In J. K. Hewitt, J. Kagan & R. N. Emde (Eds.), *Infancy to early childhood. Genetic and environmental influences on developmental change* (pp. 141–162). Oxford: Oxford University Press.
- Zaman, B. & Shrimpton-Smith, T. (2006). The FaceReader: measuring instant fun of use. In A. I. Mørch (Hrsg.), *NordiCHI 2006. Proceedings of the 4th Nordic Conference on Human-computer Interaction : changing roles, 14-18 October, 2006, Oslo, Norway* (ACM international conference proceeding series, S. 457–460). New York, NY: ACM Press.
- Zillmann, D. (2004). Emotionspsychologische Grundlagen. In R. Mangold, P. Vorderer & G. Bente (Hrsg.), *Lehrbuch der Medienpsychologie* (S. 101–128). Göttingen: Hogrefe.

- Zucker, P. J., Worthington, E. L., JR. & Forsyth, D. R. (1985). Increasing empathy through participation in structured groups Some attributional evidence. *Human Relations*, 38, 247–255.
- Zygmunt, E., Cipollone, K., Tancock, S., Clausen, J., Clark, P. & Mucherah, W. (2018). Loving Out Loud: Community Mentors, Teacher Candidates, and Transformational Learning Through a Pedagogy of Care and Connection. *Journal of Teacher Education*, 69, 127–139. <https://doi.org/10.1177/0022487117751640>

Anhang

Anhang A

Anhang A1: K-12 Kriterien

K-12 Standard	Indikatoren
1. Duration and Intensity (Dauer und Intensität)	<ol style="list-style-type: none"> 1. service-learning experiences include the processes of investigating community needs, preparing for service, action, reflection, demonstration of learning and impacts, and celebration. 2. service-learning is conducted during concentrated blocks of time across a period of several weeks or months. 3. service-learning experiences provide enough time to address identified community needs and achieve learning outcomes.
2. Link to curriculum (Verbindung zum Curriculum)	<ul style="list-style-type: none"> • service-learning has clearly articulated learning goals. • service-learning is aligned with the academic and/or programmatic curriculum. • service-learning helps participants learn how to transfer knowledge and skills from one setting to another. • service-learning that takes place in schools is formally recognized in school board policies and student records.
3. Partnerships (Partnerschaften)	<ol style="list-style-type: none"> 1. service-learning involves a variety of partners, including youth, educators, families, community members, community-based organizations, and/or businesses. 2. service-learning partnerships are characterized by frequent and regular communication to keep all partners well-informed about activities and progress. 3. service-learning partners collaborate to establish a shared vision and set common goals to address community needs. 4. service-learning partners collaboratively develop and implement action plans to meet specified goals. 5. service-learning partners share knowledge and understanding of school and community assets and needs, and view each other as valued resources.
4. Meaningful Service (Bedeutungsvoller Service)	<ol style="list-style-type: none"> 1. service-learning experiences are appropriate to participant ages and developmental abilities. 2. service-learning addresses issues that are personally relevant to the participants. 3. service-learning provides participants with interesting and engaging service activities. 4. service-learning encourages participants to understand their service experiences in the context of the underlying societal issues being addressed. 5. service-learning leads to attainable and visible outcomes that are valued by those being served.
5. Youth Voice (Mitspracherecht)	<ol style="list-style-type: none"> 1. service-learning engages youth in generating ideas during the planning, implementation, and evaluation processes. 2. service-learning involves youth in the decision-making process throughout the service-learning experiences. 3. service-learning involves youth and adults in creating an environment that supports trust and open expression of ideas. 4. service-learning promotes acquisition of knowledge and skills to enhance youth leadership and decision-making. 5. service-learning involves youth in evaluating the quality and effectiveness of the service-learning experience.

6. Diversity (Diversität)	<ol style="list-style-type: none">1. service-learning helps participants identify and analyze different points of view to gain understanding of multiple perspectives.2. service-learning helps participants develop interpersonal skills in conflict resolution and group decision-making.3. service-learning helps participants actively seek to understand and value the diverse backgrounds and perspectives of those offering and receiving service.4. service-learning encourages participants to recognize and overcome stereotypes.
7. Reflection (Reflexion)	<ol style="list-style-type: none">1. service-learning reflection includes a variety of verbal, written, artistic, and nonverbal activities to demonstrate understanding and changes in participants' knowledge, skills, and/or attitudes.2. service-learning reflection occurs before, during, and after the service experience.3. service-learning reflection prompts participants to think deeply about complex community problems and alternative solutions.4. service-learning reflection encourages participants to examine their preconceptions and assumptions in order to explore and understand their roles and responsibilities as citizens.5. service-learning reflection encourages participants to examine a variety of social and civic issues related to their service-learning experience so that participants understand connections to public policy and civic life.
8. Progress Monitoring (Beobachten des Fortschritts)	<ol style="list-style-type: none">1. service-learning participants collect evidence of progress toward meeting specific service goals and learning outcomes from multiple sources throughout the service-learning experience.2. service-learning participants collect evidence of the quality of service-learning implementation from multiple sources throughout the service-learning experience.3. service-learning participants use evidence to improve servicelearning experiences.4. service-learning participants communicate evidence of progress toward goals and outcomes with the broader community, including policymakers and education leaders, to deepen service-learning understanding and ensure that high quality practices are sustained.

Anhang A2: Anschreiben Empathie-ExpertInnen

Sehr geehrter Herr/Frau xy,

ich möchte Sie gerne als Experten dazu einladen, an einer Studie im Rahmen meiner Dissertation teilzunehmen. Im Rahmen meines Promotionsprojektes befasse ich mich zurzeit mit der Entwicklung eines Videovignettentestes zur Messung von Empathie. Da Sie sich in Ihrer Forschung mit dem Konstrukt „Empathie“ beschäftigen, würde ich mich über Ihre Unterstützung und Expertise sehr freuen. Über den Link https://ww3.unipark.de/uc/f_uni_sozpsy_LS_Paed_Horz/9c7e/ gelangen Sie zu einer Online-Befragung, die etwa 40 Minuten Ihrer Zeit in Anspruch nehmen wird.

Es werden Ihnen insgesamt 16 kurze Videosequenzen präsentiert, die Sie nach verschiedenen Gesichtspunkten bewerten sollen. Ihre Daten werden anonymisiert erhoben und ein Rückschluss auf Ihre Person ist nicht möglich. Die Befragung ist bis zum 04.09.16 geöffnet, so dass Sie gerne innerhalb dieses Zeitraumes zu einem Zeitpunkt Ihrer Wahl teilnehmen können.

Sollten Sie Fragen haben oder mehr über das Dissertations-Projekt wissen wollen, können Sie sich gerne mit mir in Verbindung setzen.

Im Voraus bereits vielen Dank für Ihre Teilnahme und Ihre Unterstützung.

Herzliche Grüße

Kathrin Kuchta

Anhang A3: Fragebogen Befragung Empathie-Experten

Instruktion/ Seite 1

Liebe Versuchsteilnehmerin, lieber Versuchsteilnehmer,

vielen Dank, dass Sie an der folgenden Studie zur Bewertung von Videosequenzen teilnehmen. Die Studie erfolgt im Rahmen meiner Dissertation zum Thema "Förderung von Empathie durch Service-Learning-Veranstaltungen an Beispielen aus der Lehrerbildung" an der Goethe-Universität in Frankfurt, pädagogische Psychologie. Wenn Sie nähere Informationen wünschen, finden Sie meine Kontaktdaten nochmals am Ende der Studie.

Im Folgenden werden Ihnen insgesamt 16 kurze Videosituationen gezeigt, zu denen Sie einige Fragen zur Bewertung des Gezeigten beantworten sollen. Es gibt bei der Beantwortung kein richtig oder falsch, da es um Ihre persönliche Einschätzung und Expertise in diesem Bereich geht. Ihre Antworten werden dazu dienen, zwei Tests zur Messung von Empathiefähigkeiten zu entwickeln. Die Bearbeitung wird etwa 30 Minuten in Anspruch nehmen und Sie können die Videosequenzen gerne mehrmals anschauen.

Um Ihre Angaben anonymisiert zu speichern, ist es notwendig, dass Sie einen persönlichen Code erstellen. Ein Rückschluss auf Ihre Person ist nicht möglich.

Ihr persönlicher Code wird folgendermaßen gebildet:

1. Buchstabe Vorname der Mutter

2. Buchstabe eigener Vorname

Geburtstag Mutter mit führender Null

Eigener Geburtsmonat mit führender Null

Beispiel:

Vorname der Mutter: Erika

Ihr Vorname: Daniel

Geburtstag der Mutter: 05. Juli

Eigener Geburtsmonat: 12.03.1985

E A 0 5 0 3

Bitte geben Sie nun Ihren persönlichen Code ein:

Seite 2

Hinweis: Da die meisten Aufnahmen im Rahmen von groß angelegten Unterrichtsbeobachtungen entstanden sind, kann es vorkommen, dass der besseren Zuordnung halber einige SchülerInnen Schilder mit Zahlen um den Hals haben. Das hat keinen wertenden Charakter und braucht Sie nicht weiter zu irritieren. Ebenfalls kann es von Videosequenz zu Videosequenz aufgrund verschiedener Aufnahmebedingungen Lautstärkeschwankungen geben - stellen Sie die Lautstärke also nicht zu laut ein, sondern justieren Sie lieber nach.

Definition Empathie/Seite 3

Bevor die Präsentation der Videosequenzen beginnt, beschreiben Sie bitte noch kurz wie Sie Empathie definieren würden:

Demografie/Seite 4

- 1. Ihr Alter (in Jahren):**

- 2. Sie sind:**
 - a. weiblich
 - b. männlich

- 3. Ihr Beruf (Mehrfachnennung möglich):**
 - a. PsychologIn (Master oder Diplom)
 - b. PsychotherapeutIn (in Ausbildung)
 - c. PsychotherapeutIn (approbiert)
 - d. TrainerIn für emotionale Kompetenzen
 - e. KommunikationstrainerIn
 - f. StudentIn
 - g. Anderer
 - h.

Bspw. Vignette 16/Seite 5

Der Lehrer versucht der Schülerin mit dem schwarzen Oberteil zu helfen. Er erläutert ihr den kritischen Punkt, woraufhin eine andere Schülerin des Tisches einwendet, dass sie das bereits gesagt hätte. Der Lehrer setzt mit erhobener Stimme zu einer Entschuldigung an. Die folgenden Fragen beziehen sich auf die Schülerin mit dem schwarzen Oberteil.

Vignette 16 abspielen

- 1. In welchem Maße konnten Sie hauptsächlich die folgenden Emotionen bei dem Mädchen beobachten:**
1= Gar nicht, 2=Sehr schwach, 3= Schwach, 4=Mittelmäßig, 5=Stark, 6=Sehr stark
 - a. Freude
 - b. Ärger
 - c. Angst
 - d. Traurigkeit
 - e. Überraschung/Neugier
 - f. Ekel

- 2. Wie würden Sie die Emotion nennen?**

- 3. Haben Sie während des Beobachtens der Sequenz bei sich Emotionen bemerken können?**
 - a. Ja
 - b. Nein

- 3.1 Wenn ja: Wie haben Sie sich gefühlt?**

- 4. Würden Sie sagen, dass diese Videosequenz dazu geeignet ist, empathische Empfindungen beim Zuschauenden auszulösen?**
 - a. Ja
 - b. Nein

- 5. Stellen Sie sich vor Sie wären im Klassenzimmer gewesen. Wie würden Sie sich dem Mädchen gegenüber empathisch verhalten?**

- 6. Wie fanden Sie die Länge der Sequenz?**

- 7 Zu kurz
- 8 Genau richtig
- 9 Zu lang
- 10 Weiß nicht

7. Gibt es sonst noch etwas, das Sie dazu sagen möchten?

Endseite

Vielen Dank für Ihre Unterstützung!

Sollten Sie Fragen haben oder mehr Informationen über das Dissertationsprojekt haben wollen, wenden Sie sich gerne an:

Dipl.-Psychologin Kathrin Kuchta

Goethe Universität Frankfurt

Institut für Psychologie - Pädagogische Psychologie

Theodor-W.-Adorno-Platz 6

60629 Frankfurt

kuchta@psych.uni-frankfurt.de

Anhang A4: Bewertung „Empathieauslösend“

Item	Vignetten	Ja-Antworten (Empathieauslösend)	Nein-Antworten	Personen
1.	Vignette 1	2 (28,6%)	5 (71,4%)	7
2.	Vignette 2	3 (42,9%)	4 (57,1%)	7
3.	Vignette 4	10 (50,0%)	10 (50,0%)	20
4.	Vignette 5	2 (33,3%)	4 (66,7%)	7
5.	Vignette 6	6 (75%)	2 (25%)	8
6.	Vignette 7	2 (28,6%)	5 (71,4%)	7
7.	Vignette 8	5 (62,5%)	3 (37,5%)	9
8.	Vignette 9	12 (63,2%)	7 (36,8%)	19
9.	Vignette 10	12 (70,6%)	5 (29,4%)	17
10.	Vignette 11	1 (16,7%)	5 (83,3%)	6
11.	Vignette 12	0 (0%)	5 (100%)	5
12.	Vignette 13	3 (42,9%)	4 (57,1%)	7
13.	Vignette 14	4 (66,7%)	2 (33,3%)	6
14.	Vignette 15	1 (8,3%)	11 (91,7%)	12
15.	Vignette 16	6 (46,2%)	7 (53,8%)	13
16.	Vignette 17	3 (42,9%)	4 (57,1%)	7
17.	Vignette 18	2 (25%)	6 (75%)	8
18.	Vignette 19	3 (23,1%)	10 (76,9%)	13
19.	Vignette 20	5 (38,5%)	8 (61,5%)	13
20.	Vignette 21	16 (84,2%)	3 (15,8%)	19
21.	Vignette 22	3 (42,9%)	4 (57,1%)	7
22.	Vignette 23	8 (44,5%)	10 (55,5%)	18
23.	Vignette 24	4 (22,8%)	14 (77,2%)	18
24.	Vignette 25	10 (71,4%)	4 (28,6%)	14
25.	Vignette 26	2 (33,3%)	4 (66,7%)	6
26.	Vignette 27	8 (53,3%)	7 (46,7%)	15
27.	Vignette 28	2 (28,6%)	5 (71,4%)	7
28.	Vignette 29	6 (42,9%)	8 (57,1%)	14
29.	Vignette 30	1 (16,7%)	5 (83,3%)	6

Anhang A5: Auswahlkriterien für Vignetten (EmpathieexpertInnenbefragung)

Schwarz= nicht ausgewählt

Schwarz= wegen schlecht erkennbarer Gesichter nicht ausgewählt

Grün=für nächste Entwicklungsrunde ausgewählt

Markiert=für nächste Entwicklungsrunde ausgewählt

Vignette	Empathieauslösend (Ja/Nein)?	Gefühlsqualitäten:	Gefühlsintensität (MW→ Min/Max- Angabe):	Anmerkungen
V1	Nein (2/5)	Divers, Einzelnennung	nur Gar nicht bis sehr schwach (1-2)	
V2	Nein (3/4)	Divers, Mehrfachnennungen	aber Gar nicht- Mittelmäßig (1-4)	
V4	Jein (10/10)	Divers, Mehrfachnennung	aber Gar nicht – schwach (1-3)	Gesicht schlecht sichtbar
V5	Nein (2/4)	Kompakt Mehrfachnennung	mit Gar nicht – sehr schwach (1-2)	
V6	Ja (6/2)	Divers, Mehrfachnennung	aber Gar nicht – sehr schwach (1-2)	
V7	Nein (2/5)	Eher kompakt Mehrfachnennung	mit Gar nicht – schwach (1-3)	
V8	Ja (5/3)	Divers, Mehrfachnennung	mit Gar nicht – mittelmäßig (1-4)	
V9	Ja (12/7)	Divers, Mehrfachnennung	mit Gar nicht – schwach (1-3)	
V10	Ja (12/5)	Sehr divers, mit vielen Mehrfachnennungen	Gar nicht – mittelmäßig (1-4)	
V11	Nein (1/5)	Kompakt Mehrfachnennung	mit Gar nicht – Mittelmäßig (1-4)	
V12	Nein (0/5)	Divers, Mehrfachnennung	mit Gar nicht – sehr schwach (1-2)	Gesicht kaum sichtbar und schlechte Quali!
V13	Nein (3/4)	Divers Mehrfachnennung	mit Gar nicht – schwach (1-3)	
V14	Ja (4/2)	Divers Mehrfachnennung	mit Gar nicht – Stark (1-5)	
V15	Nein (1/12)	Divers, mit vielen Mehrfachnennungen	Gar nicht – mittelmäßig (1-4)	Gesicht kaum sichtbar!
V16	Nein (6/7)	Divers Mehrfachnennung	mit Gar nicht – sehr schwach (1-2)	
V17	Nein (3/4)	Divers Mehrfachnennung	mit Gar nicht – schwach (1-3)	
V18	Nein (2/6)	Divers Mehrfachnennung	mit Gar nicht – mittelmäßig (1-4)	Gesicht kaum sichtbar!
V19	Nein (3/10)	Sehr divers, mit vielen Mehrfachnennungen	Gar nicht – mittelmäßig (1-4)	Freude als Mehrfachnennung
V20	Nein (5/8)	Sehr divers, mit vielen Mehrfachnennungen	Gar nicht – schwach (1-3)	

V21	Ja (16/3)	Divers mit Mehrfachnennung	Gar nicht – mittelmäßig (1-4)	
V22	Nein (3/4)	Divers mit nur einer Mehrfachnennung	Gar nicht – sehr schwach (1-2)	
V23	Nein (8/10)	Sehr divers, mit vielen Mehrfachnennungen	Gar nicht – schwach (1-3)	
V24	Nein (4/14)	Sehr divers, mit Mehrfachnennungen und vielen Einzelnennungen	Gar nicht – mittelmäßig (1-4)	Freude als häufigste Nennung, aber sehr divers
V25	Ja (10/4)	Sehr divers, mit vielen Mehrfachnennungen	Gar nicht – Mittelmäßig (1-4)	
V26	Nein (2/4)	Divers, mit Mehrfachnennungen und vielen Einzelnennungen	Gar nicht – sehr schwach (1-2)	
V27	Ja (8/7)	Sehr divers, mit vielen Mehrfachnennungen	Gar nicht – schwach (1-3)	
V28	Nein (2/5)	Sehr divers, mit Mehrfachnennungen	Gar nicht – mittelmäßig (1-4)	Frust und Ärger als häufigste Nennung
V29	Nein (6/8)	Sehr divers, mit vielen Mehrfachnennungen	Gar nicht – schwach (1-3)	
V30	Nein (1/5)	Divers mit einer Mehrfachnennung	Gar nicht – schwach (1-3)	

Anhang A6: Wertebereich der Intensitätsantwort des Vignettentests

	Freude	Ärger	Angst	Traurigkeit	Überraschung/Neugierde	Ekel
Vignette 2 (<i>n</i> =6)	3	4	2	2	2	1
(<i>SD</i>)	(1.05)	(1.64)	(1.21)	(0.55)	(1.03)	(0.00)
Konfidenzintervall	1-4	2-5	1-3	1-2	1-3	1
95%						
Bereich +/- 1 SD	1-4	2-5	1-3	1-2	1-3	1
Vignette 6 (<i>n</i> =8)	2	1	2	2	1	1
(<i>SD</i>)	(0.92)	(0.35)	(1.31)	(1.49)	(0.46)	(0.00)
Konfidenzintervall	1-2	1	1-3	1-3	1-2	1
95%						
+/- 1 SD	1-3	1	1-3	1-4	1-2	1
Vignette 8 (<i>n</i> =8)	4	1	2	1	3	1
(<i>SD</i>)	(1.16)	(0.00)	(1.19)	(0.00)	(2.13)	(0.00)
Konfidenzintervall	3-5	1	1-3	1	2-5	1
95%						
+/- 1 SD	3-5	1	1-3	1	1-6	1
Vignette 9 (<i>n</i> =19)	1	2	3	2	2	1
(<i>SD</i>)	(0,23)	(1,2)	(1,65)	(1,39)	(1,12)	(0,00)
Konfidenzintervall	1	1-2	2-4	2-3	1-2	1
95%						
+/- 1 SD	1	1-3	1-5	1-4	1-3	1
Vignette 10 (<i>n</i> =19)	4	2	2	1,	3	1
(<i>SD</i>)	(1.05)	(1.33)	(1.17)	(1.08)	(1.32)	(0.62)
Konfidenzintervall	3-4	1-3	1-2	1-2	2-3	1
95%						
+/- 1 SD	3-5	1-3	1-3	1-2	1-4	1-2
Vignette 11 (<i>n</i> =5)	1	4	2	2	2	3
(<i>SD</i>)	(0.00)	(1.82)	(1.34)	(1.64)	(0.55)	(0.84)
Konfidenzintervall	1	1-6	1-3	1-4	1-2	2-4
95%						
+/- 1 SD	1	2-5	1-3	1-4	1-2	2-4
Vignette 14 (<i>n</i> =6)	1	5	2	2	2	2
(<i>SD</i>)	(0.00)	(1.05)	(1.21)	(1.22)	(1.67)	(1.21)
Konfidenzintervall	1	3-6	1-4	1-3	1-4	1-3
95%						
+/- 1 SD	1	3-6	1-4	1-3	1-4	1-3
Vignette 19 (<i>n</i> =12)	4	2	2	1	2	1
(<i>SD</i>)	(1.16)	(1.29)	(0.62)	(0.89)	(0.94)	(0.00)
Konfidenzintervall	3-4	1-3	1-2	1-2	1-2	1
95%						
+/- 1 SD	2-5	1-4	1-2	1-2	1-3	1
Vignette 21 (<i>n</i> =18)	2	1	3	1	4	1
(<i>SD</i>)	(0.94)	(0.85)	(1.18)	(0.55)	(1.94)	(0.00)
Konfidenzintervall	1-2	1-2	2-3	1	3-5	1
95%						
+/- 1 SD	1-3	1-2	2-4	1-2	2-6	1
Vignette 24 (<i>n</i> =18)	4	2	2	1	3	1
(<i>SD</i>)	(1.19)	(0.98)	(1.04)	(0.55)	(1.71)	(1.04)
Konfidenzintervall	3-5	1-2	1-2	1	2-4	1-2
95%						
+/- 1 SD	3-5	1-3	1-3	1-2	1-4	1-2

Vignette 25 (<i>n</i> =13)	1	2	4	2	2	1
(<i>SD</i>)	(0.00)	(0.78)	(1.57)	(1.07)	(1.19)	(0.28)
Konfidenzintervall 95%	1	1-2	3-5	2-3	1-2	1
+/- 1 <i>SD</i>	1	1-2	2-5	1-3	1-3	1
Vignette 27 (<i>n</i> =15)	1	2	2	2	3	1
(<i>SD</i>)	(0.26)	(1.15)	(1.37)	(0.99)	(1.55)	(0.82)
Konfidenzintervall 95%	1	1-2	1-3	1-2	2-3	1-2
+/- 1 <i>SD</i>	1	1-3	1-4	1-3	1-4	1-2
Vignette 28 (<i>n</i> =7)	1	3	2	3	4	2
(<i>SD</i>)	(0.00)	(1.38)	(1.21)	(1.63)	(1.46)	(0.98)
Konfidenzintervall 95%	1	2-5	1-3	1-5	3-5	1-2
+/- 1 <i>SD</i>	1	2-5	1-3	1-5	2-5	1-3
Vignette 30 (<i>n</i> =6)	2	2	2	1	2	1
(<i>SD</i>)	(1.14)	(0.89)	(0.89)	(0.00)	(0.89)	(0.00)
Konfidenzintervall 95%	1-4	1-3	1-4	1	1-4	1
+/- 1 <i>SD</i>	1-4	1-2	2-3	1	2-3	1

Gerundeter Mittelwert (SD)/Wertebereich +/- 1 SD

Anhang A7: Gefühlswort-Liste

Beschreiben die folgenden Wörter Gefühlszustände? Wenn nicht, welches Wort könnte den dazu passenden Gefühlszustand besser ausdrücken?

0=nein; 1=Ja; -99=nicht gegebene Antwort

	Gefühlswort?	Rater1	Rater2
1.	Ablehnung	0	0
2.	Abneigung	1	1
3.	Alleinsein	0	0
4.	Amüsement	0	0
5.	Amüsiert	1	1
6.	Angst	1	1
7.	Anspannung	0	1
8.	Ärger	1	1
9.	Aufregung	1	1
10.	Auftrumpfend	0	1
11.	Ausgelassenheit	1	1
12.	Belastung	1	1
13.	Belustigung	1	1
14.	Beschwingt	1	1
15.	Besorgtheit	1	1
16.	Bloßstellung	0	1
17.	Desinteresse	1	1
18.	Distanziertheit	1	1
19.	Entspannung	1	1
20.	Enttäuschung	1	1
21.	Erleichterung	1	1
22.	Erstaunen	1	1
23.	ertappt	0	1
24.	Erwartungsvoll	-99	-99
25.	Fokussiert	1	0
26.	Fragend	0	1
27.	Freude	1	1
28.	Frust	1	1
29.	Frustration	1	1
30.	Genervtheit	1	1
31.	Genugtuung	1	1
32.	Geringschätzung	0	1
33.	Gleichgültigkeit	1	1
34.	Heiterkeit	1	1
35.	Hilflosigkeit	1	1
36.	Interesse	1	1
37.	Irritation	1	1
38.	Langeweile	1	1
39.	Nachdenklich	1	1

40.	Neid	1	1
41.	Neugier	1	1
42.	Neugierde	1	1
43.	Niedergeschlagenheit	1	1
44.	Offenheit	1	0
45.	Peinlichkeit	0	0
46.	Provokant	0	0
47.	provozieren	0	1
48.	Ratlosigkeit	1	1
49.	Resignation	1	1
50.	Rücksichtslosigkeit	0	1
51.	Ruhig	1	1
52.	Schadenfreude	0	1
53.	Scham	1	1
54.	Schreck	1	1
55.	Schüchternheit	1	1
56.	Selbstbestätigung	0	0
57.	Selbstsicherheit	1	1
58.	Selbstzufriedenheit	1	1
59.	Sorge	1	1
60.	Spaß	1	0
61.	Spott	0	0
62.	Stärke	0	1
63.	Stolz	1	1
64.	streitsüchtig	1	1
65.	Stress	1	1
66.	Traurig	1	1
67.	Traurigkeit	1	1
68.	Triumph	1	0
69.	Überforderung	1	1
70.	Überheblichkeit	0	0
71.	Überlegenheit	0	1
72.	Übermut	1	1
73.	überprüfend	0	0
74.	Überraschung	1	1
75.	Überrumpelt	1	1
76.	Unbeholfenheit	1	1
77.	Ungeduld	1	1
78.	Ungläubiges Staunen	0	1
79.	Uninteressiert	1	1
80.	Unlust	0	1
81.	Unruhe	1	1
82.	Unsicherheit	1	1
83.	Verachtend	0	1
84.	Verblüffung	1	1
85.	Verlegenheit	1	1

86.	Verloren	1	1
87.	Verschämt	1	1
88.	Verspieltheit	-99	-99
89.	Verwirrung	1	1
90.	Vorfreude	1	1
91.	Wachsamkeit	1	1
92.	Wertlosigkeit	1	1
93.	Widerwille	0	1
94.	Wut	1	1
95.	Zufriedenheit	1	1
96.	Zurückgenommen	0	0
97.	Zuversicht	1	1

Anhang A8: Freitextantworten/Verhaltensoption

Unbearbeitete Rohdaten

Vignetten	Gegebene Antworten	Extrahierte Antwortoptionen
02	<ol style="list-style-type: none"> 1. Der Schüler an der Tafel scheint selbstbewusst zu sein, deshalb würde ich hier noch gar nicht eingreifen. 2. Ihn zu erkennen geben, dass man weiß, dass er es schon kann. ihm vielleicht eine anspruchsvollere Aufgabe geben? 3. Seine Ungeduld, seinen Ärger spiegeln. Dann würde er sich zumindest verstanden fühlen. Evtl. herausfinden, ob er sich in seinen Fähigkeit nicht ausreichend gesehen fühlt - siehe warum immer der... 4. Ich verstehe, dass Du genervt bist. Aber bitte bringe die Geduld auf, den Lösungsweg von X anzuhören. Es gibt sicherlich auch Situationen, in denen es anderen langweilig ist, wenn sie Dir zuhören. 5. Mir scheint, es handelt sich hier nicht um Empathie. Ich würde ähnlich kurz zeigen, dass die Anmerkung angekommen ist, aber nicht weiter sich in die Interaktion vertiefen. 	<ol style="list-style-type: none"> 1. Das gezeigte Verhalten des Lehrers in dem Clip. 2. Den Schüler an der Tafel zurecht weisen und die Klasse zur Ruhe mahnen. 3. Dem Schüler (im hellgrauen Oberteil) zeigen, dass man die Genervtheit und den Ärger versteht, aber ihn gleichzeitig um Geduld bitten, den Lösungsweg des Schülers an der Tafel anzuhören - auch wenn man weiß, dass er es kann.
06	<ol style="list-style-type: none"> 1. Ich würde dem Mädchen freistellen sich zu setzen oder es solange sie vorne steht immer wieder mit ins Gespräch mit einbinden. Z.B. könnte sie sich selbstgesteuert Unterstützung aus der Klasse holen. 2. wenn es sich hinsetzt oder zu mir schaut, aufmunternde Gesten + Blicke zusenden 3. (ich gehe davon aus, als dritte Person, nicht als Lehrerin oder Mitschülern, und im Anschluss an die Situation) Abklären der Perspektive und Emotion des Mädchens, im Sinne von so etwas wie Das ist ganz schön doof, wenn alle anderen was sagen dürfen, und man selbst nicht...? um so herauszufinden, hat sie sich unsicher gefühlt, hat sie sich über die Lehrerin geärgert, oder war sie vielleicht sogar erleichtert, dass sie nichts weiter sagen musste..? 	<ol style="list-style-type: none"> 1. Der Schülerin an der Tafel einen Moment Zeit geben und ihre Gedanken zu der gestellten Frage anhören. 2. Das Mädchen an der Tafel weiterhin mit einbeziehen, sich ihr immer wieder zuwenden und auch loben. 3. Das gezeigte Verhalten der Lehrerin in dem Clip

	<ol style="list-style-type: none"> 4. Am Ende der Videosequenz, als die Lehrerin die Pole an der Tafel gezeigt hatte, wäre ein guter Zeitpunkt gewesen, sich wieder dem Mädchen zuzuwenden und es wieder mit einzubeziehen. 5. Zeit lassen, Anschauen, Loben, weiter mit einbeziehen auch den Kontakt zu den Mitschülern und ihr gestalten 6. Zunächst ihre Gedanken zu der gestellten Frage anhören, Zeit geben. 	
08	<ol style="list-style-type: none"> 1. empathisch abholen 2. Wie die Lehrerin 3. kleiner Remppler/ Knuff in die Rippen - so nach dem Motto: eh, lach niemanden aus... 4. Versuch eines Abklärens der Empfindungen: Schon lustig, wenn jemand so richtig daneben liegt...? (gesetzt den Fall, ich würde für den jungen Empathie aufbringen und nicht eher mit dem Schüler, der ausgelacht wird - vgl. auch mit dem Video, in der der Schüler mit dem Base-Cap die Schülerin an der Tafel anrempelt) 5. Validation dass sich die Aussage erstmal lustig anhört, im Anschluss Erklärung über Bananenbau die im tropischen Gewächshaus möglich sei, sowie, dass nicht alle Kinder wissen können wo verschiedene Obstsorten her kommen, vielleicht sei er sich auch bei der ein oder anderen Sorte unsicher, dann könne er das jetzt gerne erfragen 	<ol style="list-style-type: none"> 1. Das gezeigte Verhalten der Lehrerin in dem Clip. 2. Ansprechen, dass sich die Aussage erstmal lustig anhört, aber im Anschluss erklären, dass Bananen-Anbau im Gewächshaus auch bei uns möglich ist. 3. Das Gelächter und Geflüstere unterbinden und den Schüler, der mit „Bananenbaum“ geantwortet hat, fragen, ob er sich bei der ein oder anderen Sorte unsicher ist und jetzt gerne Fragen stellen könne.
09	<ol style="list-style-type: none"> 1. Ihre Wissenslücke aufgefüllt, ohne sie dafür zu verurteilen, und ihr eventuell erklärt, wie die Formel für die Flächenberechnung zustande kommt, damit sie sich diese für die Zukunft leichter merken kann. Außerdem hätte ich leise mit ihr gesprochen, damit sie sich nicht vor der Klasse vorgeführt vorkommt. 2. Sie an die Formel erinnern und den Unterricht fortführen. 3. Ihr sagen, dass ich das auch nicht gewusst hätte 4. Als Mitschüler könnte man das Mädchen direkt ansprechen, um nicht auf die abwertende Art des Lehrers einzugehen. 5. Versuch, einen vorsichtigen Einstieg in die Situation zu bekommen, um als ersten 	<ol style="list-style-type: none"> 1. Das gezeigte Verhalten des Lehrers in dem Clip. 2. Ihr mit der Formel weiterhelfen, damit sie sich die Formel zukünftig besser merken und ohne große Verunsicherung weiter am Lösungsweg arbeiten kann. Außerdem leise mit ihr sprechen, damit sie sich nicht vor der Klasse vorgeführt vorkommt. 3. Ihre Frage ohne weitere Kommentare beantworten.

-
- Schritt die Emotion abzuklären. Na das hätte der Lehrer ja jetzt nicht unbedingt vor allen sagen müssen...?
6. ich wuerde ihr sagen dass sie sich von dem Lehrer nicht beeindrucken lassen soll, und sich besser mit der Aufgabe beschaefigten
 7. Nachfragen
 8. Oh deine Frage überrascht mich. Mir ist wichtig, dass ihr die Grundkenntnisse darüber habt bzw. wisst, wo ihr nachschlagen könnt.
 9. Hast du Angst die Aufgabe nicht richtig verstanden zu haben?
 10. Du wolltest einfach ein Antwort auf deine Frage, oder? Du hast dem Lehrer vertraut, als du fragtest, oder? Wie geht's dir denn nachdem das passiert ist? ihre Antwort abwarten, darauf reagieren, mit einer Einföhlung anbietenden Frage
 11. kurze Pause machen und durchatmen lassen, sie auf eine andere Aufgabe stoßen, andere Wortwahl wie Scheitern nutzen
 12. Die kenntnis nicht laut verlauten lassen
 13. ihr zu verstehen geben, dass es ok ist, mal einen Hänger zu haben, jeder steht mal auf dem Schlauch.Untersützen, die Blockade zu kösen. Diskret bleiben, Verständnis zeigen, ruhig erklären und Formel geben, damit an Lösungsweg weitergemacht werden kann, bevor zu große Verunsicherung eintritt
 14. trösten: ...zumindest hat er nicht Deinen Namen vor der ganzen Klasse gesagt sondern nur hier vorne scheitert man...
 15. Mach Dir nichts draus. Das kann jeder mal vergessen.
 16. Als Mitschüler: beruhigen, trösten, aufmuntern
 17. Ich würde die Frage der Schülerin ohne zusätzliche Kommentare beantworten. Evtl. noch den Hinweis ergänzen, wann dieses Thema in der Klasse schon behandelt wurde.

10

1. Ich hätte ihn zuerst für das richtige Ergebnis gelobt und hätte dann mit ihm besprochen, wieso sein Rechenweg zwar auch zum Ziel führt, aber es wahrscheinlich einen besseren gibt.
1. Das gezeigte Verhalten des Lehrers in dem Clip.
 2. Den Schüler bestärken und für den originellen Lösungsweg loben und den Lösungsweg in der Klasse vorstellen lassen.
-

-
- | | |
|--|--|
| <ol style="list-style-type: none"> 2. Sehn Sie mal ist ein deutlicher Ausdruck von Enttäuschung und Frust über die geringschätzende, lächerlich machende, Rückmeldung des Lehrer (ich bin immer wieder überrascht ... (wie Du Blödmann auch mal ein Korn findest ...) Empathisch wäre, den Schüler in seiner Frustration wahrzunehmen und aufbauend und wertschätzend mit ihm umzugehen. Wow... ein anderer Lösungsweg, erklär mir mal, wie Du darauf gekommen bist...) 3. Ich würde ihn für seinen originellen Lösungsweg loben und den Lösungsweg in der Klasse vorstellen. 4. Und, glaubst Du, das Dein Lösungsweg richtig ist? ... weil es für mich eine Diskrepanz gibt zwischen flüssigem Vortragen und zurückziehender Körperhaltung. Das würde ich als Unsicherheit interpretieren und dort indirekt emotional ansetzen. Bist Du unsicher? 5. bestärken, sagen, dass er stolz sein kann 6. Ihm anerkennend auf die Schulter klopfen 7. Schulter klopfen, Augen petzen , superklasse, dass dir das gelungen ist, grinsen 8. Als Lehrer nocheinmal zugeben , dass der klassische Lösungsweg nicht immer der richtige ist und dieser sehr kreativ ist. 9. Versuch eines Abklärens der Emotion und der Perspektive Schon cool, wenn man den Lehrer auch mal beeindrucken kann..? 10. high five.... das der Schueler getrumpft hat 11. Von mir reden und ehrlich bei mir bleiben, anstatt ihn zu bewerten (Gehirnwindungen) weil ich selbst noch nicht darauf gekommen bin. Ich bin überrascht, da bin ich noch gar nicht drauf gekommen, bin begeistert. 12. wertschätzend verbal loben 13. ähnlich. Lob wichtig. 14. interessiert nachfragen, wie er das gelöst hat und wie er darauf kam, ihn darin bestärken 15. Freude stärken und Ärger abmildern evtl. mit einer Bemerkung, dass sein Lösungswg super war und der Lehrer heute ein A... ist. 16. Als Lehrer: Verstärken, loben | <ol style="list-style-type: none"> 3. Diesen kreativen Lösungsweg besonders hervorheben und interessiert nachfragen wie sie darauf gekommen sind. |
|--|--|

11

-
- | | |
|---|--|
| <ol style="list-style-type: none"> 1. Ein Gesprächsangebot machen. 2. lass Dich nicht von den Jungs ärgern, ist es nicht wert | <ol style="list-style-type: none"> 1. Sich aus der Situation raushalten. 2. Die Schüler zurechtweisen. |
|---|--|
-

	3. Als Mitschüler: unterstützen, Verständnis zeigen	3. Den Schülern ein Gesprächsangebot machen.
14	<ol style="list-style-type: none"> 1. fragen, wo die Schwierigkeiten sind, Ärger und Frustration wahrnehmen und ernst nehmen. ja, verstehe, Du bist genervt, weil Du das Gefühl hast, Du kriegst das nicht , ... lass uns nochmal an die Aufgabe gehen ... komm ich unterstütze Dich 2. Gar nicht. 3. Kommentieren, dass sie anscheinend keine Lust hat und sie bitten, sich konzentriert und ohne Handy an die Aufgabe zu setzen. 4. das Mädchen ist mir unsympatisch, ich würde nichts sagen 5. Dem Mädchen glauben, dass sie nicht weiß wie es geht. Bist du frustriert und brauchst du noch mehr Klarheit über die Aufgabe? 6. Bist du genervt, weil du eine klare Anweisung brauchst? 	<ol style="list-style-type: none"> 1. Das gezeigte Verhalten der Lehrerin in dem Clip. 2. Kommentieren, dass sie anscheinend keine Lust hat und sie bitten, sich konzentriert und ohne Handy an die Aufgabe zu setzen. 3. Der Schülerin glauben, dass Sie nicht weiß wie es geht und sie fragen, wo genau die Schwierigkeiten liegen. Ihre Gefühle ernst nehmen und ihr das Angebot machen, sie beim Finden des Lösungsweges zu unterstützen.
19	<ol style="list-style-type: none"> 1. Auch hier sehe ich noch keine Grund für eine detaillierte empathische Interaktion. 2. Angebot Na, langweilig? 3. (Es scheint mir keine besonders Empathie erfordernde Situation zu sein, aber gesetzt den Fall, ich würde versuchen wollen, auf den Jungen empathisch einzugehen): Versuch eines Abklärens der Emotion als 1. Schritt, ...ist schon lustig sowas, hm? (auch wenn ich jetzt noch keine Idee habe, wie das weitergeführt werden könnte, das käme auf die Situation an) 4. Auf sein Verhalten ansprechen und Grenzen setzen. 5. Seinen Ärger ernst nehmen, Hintergrund besprechen 6. Solange es nicht weiter eskaliert würde ich es laufen lassen. 7. hey, hör doch mal auf! Erzähl mir lieber, was Deine Hobbies sind. Machst Du gern Sport? (Aufmerksamkeit geben und von destruktivem Verhalten ablenken) 	<ol style="list-style-type: none"> 1. Sich aus der Situation raushalten und die Schüler das selbst klären lassen. 2. Den Schüler auf sein Verhalten ansprechen und Grenzen setzen. 3. Den Schüler in seiner Emotion ernst nehmen und den Hintergrund der Situation besprechen.
21	<ol style="list-style-type: none"> 1. Ich hätte ihr gesagt, dass sie Unterhaltungen mit ihren Freunden/Freundinnen erst nach der Stunde führen soll, ohne sie zur Schau zu stellen. 2. Kontakt aufnehmen, Anliegen äußern, Reaktion abwarten wenn paradoxe 	<ol style="list-style-type: none"> 1. Das gezeigte Verhalten des Lehrers in dem Clip. 2. Die Schülerin leise darauf hinweisen, dass sie private Angelegenheiten bitte erst nach der Stunde klären soll.

	<p>Intervention dann bitte auf Augenhöhe und nicht auf Kosten des Mädchens</p> <ol style="list-style-type: none"> 3. Ich würde sie fragen, ob es ein wichtiger Brief ist, wenn nein, würde ich sie auffordern ihn auf später zu verschieben, wenn ja, würde ich sagen, dass ich Verständnis dafür habe, dass sie ihn lesen will, dass der Unterricht aber die unpassende Situation dafür ist. 4. Sind gerade andere Dinge wichtiger...? 5. Ich würde ihr sagen, dass ihr Lehrer sich blöd verhalten hat 6. Unter vier Augen fragen, warum der Brief so wichtig ist und ob er weg muss, damit sie sich auf den Unterricht konzentriert. 7. (Ich gehe jetzt davon aus, in der Rolle einer dritten Person, nicht als Lehrer oder Mitschüler) Versucht eines Abklärens/Spiegelns der Emotion als ersten Schritt, um Verständnis zu signalisieren. Oh scheiße, das ist ganz schön peinlich, wenn der Lehrer einen beim Briefelesen im Unterricht erwischt? Und wenn dann auch noch alle lachen...? 8. Beschwichtigend 9. Die peinliche Situation vermieden. 10. Desire, ich sehe dass du wiederholt Briefe liest, statt die Aufgaben zu machen. Da mache ich mir echt Sorgen, wie du den Stoff dann gut lernen kannst. 11. Ihr Interesse ,durch Ansprechen der Ablenkung ,wieder auf den Unterricht zu lenken. 12. verständnis ausdrücken, dass andere Dinge vielleicht spannender erscheinen als der Unterricht, sie dezent merken lassen, dass man ihr Verhalten bemerkt hat, ohne sie bloßzustellen und Interesse am Inhalt des Unterrichts wecken 13. Ähnlich den Mädchen am Tisch - es gemeinsam weg lachen. 14. Hey, nicht so schlimm. Jeder lässt sich mal erwischen 15. Als Mitschüler: aufmuntern, zum Mitlachen bringen, beruhigen 16. Humor einsetzen als empathische Reaktion, finde ich eine gute Lösung. 	<ol style="list-style-type: none"> 3. Verständnis dafür ausdrücken, dass andere Dinge vielleicht spannender erscheinen als der Unterricht und sie damit dezent merken lassen, dass man ihr Verhalten bemerkt hat. Im weiteren Verlauf versuchen ihr Interesse am Inhalt des Unterrichts zu wecken.
24	<ol style="list-style-type: none"> 1. Anweisung freundlich formulieren, aufmunternd lächeln, dass die Anweisung 	<ol style="list-style-type: none"> 1. Das gezeigte Verhalten der Lehrerin in dem Clip.

	<p>keine Geringschätzung oder Kritik an der Person ist, Atmosphäre wieder auflockern,</p> <ol style="list-style-type: none"> 2. direkt ansprechen, Kontakt abwarten, dann mein Anliegen nennen, Reaktion abwarten 3. Gar nicht 4. wie die Lehrerin 5. (Dies scheint mir keine so stark Empathie erfordernde Situation zu sein, aber gesetzt den Fall, ich würde versuchen wollen, empathisch auf das Mädchen einzugehen:) Vorschlag einer Situation, um die Perspektive und die Empfindungen abzuklären, vielleicht so etwas wie Na, weiter quatschen wäre jetzt auch netter als Unterricht? als 1. Schritt 6. gar nicht 7. Sie bereiten sich gerade für die Stunde vor. Wenn ich als Lehrer anfangen möchte, dann spreche ich von mir, was mir wichtig ist. Ich möchte jetzt anfangen, damit ich die Zeit gut mit euch nutzen kann. Ist jemand noch nicht bereit dazu? 8. Ich hätte keine Notwendigkeit gesehen darauf zu reagieren. 9. Zunächst Zuwenden und begrüßen 10. Freundlich aber bestimmt (ohne Zeigefinger) 11. Erwachsene sind häufig anstrengend mit ihren Regeln, Lehrer natürlich besonders. Kann ich verstehen. Wenn ihr keinen Ärger wollt, stellt die Taschen doch trotzdem runter. Oder gibt es einen wichtigen Grund, warum sie oben bleiben sollten? 12. Die Schülerin und die Lehrerin interagieren nicht mit einander. Daher sehe ich keinen Anlass, sich als Lehrkraft empathisch zu verhalten. 	<ol style="list-style-type: none"> 2. Die Anweisung freundlich formulieren. Zusätzlich aufmunternd lächeln, und erklären, dass die Anweisung keine Geringschätzung oder Kritik an der Person ist, um die Atmosphäre wieder aufzulockern. 3. Erklären, dass Sie gleich anfangen möchten, um die Zeit mit den SchülerInnen gut zu nutzen. Nachfragen, ob jemand noch nicht bereit dazu ist.
25	<ol style="list-style-type: none"> 1. Dem Mädchen gegenüber genauso wie gegenüber dem Rest der Klasse, ich würde fragen wo die Schwierigkeiten mit der Aufgabe liegen und evtl. nochmal einen didaktischen Schritt zurück gehen. 2. freundlich Ansprache, Atmosphäre entspannen, Druck nehmen, zu Antworten ermuntern, 3. motivieren die Antwort zu versuchen 4. Ich würde ihr sagen, dass ich auch nicht die Lösung wusste 5. Aktiver zur Lösung hinführen 	<ol style="list-style-type: none"> 1. Das gezeigte Verhalten der Lehrerin in dem Clip. 2. Dem Mädchen genauso wie gegenüber dem Rest der Klasse fragen, wo die Schwierigkeiten mit der Aufgabe liegen und eventuell nochmal einen didaktischen Schritt zurückgehen. 3. Durch eine freundliche Ansprache die Atmosphäre entspannen und den Druck

	<ol style="list-style-type: none"> 6. ermutigende Blicke zuwerfen 7. Gefühle anbieten und versuchen zu klären, um was es ihr geht. Bist du ratlos, brauchst du noch etwas Zeit oder ist der letzte Schritt noch unklar? 8. Na, was geht in dir vor, du wirkst ein bisschen ängstlich. 9. Verständnis zeigen, wenn Aufgabe schwer fällt, angstfreie Atmosphäre schaffen (keiner wird bloßgestellt), zum Mitdenken anregen, ohne abzufragen 10. Einen Teil der Aufgabe selbst lösen und sie ermutigen, den Rest zu ergänzen 11. Aufmunternd 12. Als Lehrerin: freundliches Aufmuntern, durch leitendes Fragen zur Lösung führen, loben 13. Das Video zu kurz und zeigt zunächst nur, dass die Lehrerin auf die Antwort wartet. Ich würde auch zunächst abwarten und dann evtl. paar Hinweise geben. 	<p>nehmen. Dann die Schülerin und den Rest der Klasse zu Antworten ermuntern und zum Mitdenken anregen.</p>
27	<ol style="list-style-type: none"> 1. empathisch Kontakt aufnehmen, Augenhöhe suchen (physisch und verbal), Zeit lassen 2. Ich würde mit ihm gemeinsam die Lösung versuchen herauszufinden. 3. ich würde ihn direkt ansprechen, zu Wort kommen lassen, seinen Fragen und Antworten zuhören, 4. abwartend, Zeit lassen 5. Versuch einer Spiegelung der Perspektive und Empfindungen des Jungen als ersten Schritt, evtl. so etwas wie Du bist Dir sicher, dass Du die Lösung doch irgendwo in Deinen Unterlagen haben musst? 6. Ihn fragen womit ich ihm helfen kann, da er offensichtlich ratlos nach einer Antwort auf die Fragen des Lehrers sucht. 7. Augenhöhe, Druck raus nehmen 8. Ich freue mich, wenn ich sehe, wie du Seite für Seite das Heft durchsuchst. 9. Ok. Ich seh, du möchtest das gerne verstehen, ich weiß gerade noch nicht wie ich dir weiterhelfen kann, was möchtest du wissen? 10. Ich würde mich zu ihm setzen und ihn beruhigen, dass er nochmal in Ruhe denken kann. 	<ol style="list-style-type: none"> 1. Das gezeigte Verhalten des Lehrers in dem Clip. 2. Versuchen gemeinsam mit ihm die Lösung herauszufinden. 3. Den Schüler fragen, womit man ihm helfen kann, da er offensichtlich nach einer Antwort auf die Fragen sucht. Man könnte sich dabei zu ihm setzen und ihn beruhigen, so dass er nochmal in Ruhe nachdenken kann.

	11. Eine ruhigere Atmosphäre suchen und mit dem Jungen auf Augenhöhe sprechen. Mehr Sicherheit vermitteln.	
28	<ol style="list-style-type: none"> 1. aufmunternd, fragend Ok, wie machen wir / Du jetzt das Beste aus dieser Situation? 2. Hilfe/Unterstützung anbieten (mit ins Buch schauen lassen) 3. Ihr anbieten, dass sie meine Unterlagen mitbenutzen kann. 4. Klären, ob die Materialien gebraucht werden und - wenn ja - nach einer Lösung des Problems suchen. 5. trösten, ihr helfen soweit ich kann 6. Bist du gerade enttäuscht, weil du gerne Bescheid wissen wolltest / mehr Klarheit gebraucht hättest? 7. Na, bist du verärgert? 	<ol style="list-style-type: none"> 1. Das gezeigte Verhalten des Lehrers in dem Clip. 2. Der Schülerin anbieten, dass sie die Unterlagen ihrer Sitznachbarin mitnutzen kann. 3. Aufmunternd lächeln und mit ihr gemeinsam eine Lösung besprechen.
30	<ol style="list-style-type: none"> 1. Ich sehe hier noch gar keinen Grund für eine detaillierte empathische Interaktion. 2. Er braucht gerade kein Verhalten von mir 3. Als Mitschüler: aufmunternd zulächeln 	<ol style="list-style-type: none"> 1. Das gezeigte Verhalten der Lehrerin in dem Clip. 2. Die Schüler in ihrem eigenen Arbeitsstil weitermachen lassen. 3. Aufmunternd lächeln und beide motivieren sich mit der Aufgabe zu beschäftigen. Anbieten sie bei Problemen gerne zu unterstützen.

Anhang A9: Übersicht der ICC für alle Vignetten ($\alpha=.05$)

Vignette (n=RaterAnzahl)	Reliabilität	ICC-Wert	Signifikanz
V02 (n=6)	.77	3.93	.002
V06 (n=8)	.65	2.62	.022
V08 (n=8)	.93	12.26	.000
V09 (n=19)	.93	12.12	.000
V10 (n=19)	.94	15,01	.000
V11 (n=5)	.74	3.46	.006
V14 (n=6)	.89	8.01	.000
V19 (n=12)	.92	11.49	.000
V21 (n=18)	.96	21.39	.000
V24 (n=18)	.96	19.83	.000
V25 (n=13)	.92	11.81	.000
V27 (n=15)	.84	5.58	.000
V28 (n=7)	.83	5.40	.000
V30 (n=5)	.78	4.21	.002

Anhang B

Anhang B1: Anschreiben Expertenbefragung LehrerInnen

Liebe Lehrperson,

im Rahmen meiner Dissertation zum Thema „Förderung von Empathie durch Service-Learning-Veranstaltungen an Beispielen aus der Lehrerbildung“ befasse ich mich zurzeit mit der Entwicklung eines Videovignettentestes zur Messung von Empathie. Da Sie in Ihrer Tätigkeit als LehrerIn eine professionelle Sicht auf Unterrichtssituationen haben, würde ich mich über das Einbringen Ihrer Expertise freuen und Sie einladen an meiner Studie teilzunehmen. Über den Link https://ww3.unipark.de/uc/f_uni_sozpsy_LS_Paed_Horz/3f66/ gelangen Sie zu einer Online-Befragung, die etwa 20-30 Minuten Ihrer Zeit in Anspruch nehmen wird.

Es werden Ihnen insgesamt 9 kurze Videosequenzen mit Unterrichtssituationen präsentiert, die Sie nach verschiedenen Gesichtspunkten bewerten sollen. Ihre Daten werden anonymisiert erhoben und ein Rückschluss auf Ihre Person ist nicht möglich. Die Befragung ist bis zum 29.01.17 geöffnet, so dass Sie gerne innerhalb dieses Zeitraumes zu einem Zeitpunkt Ihrer Wahl teilnehmen können. Sie können den Link auch sehr gerne an weitere deutschsprachige Lehrer und Lehrerinnen weiterleiten. Sollten Sie Fragen haben oder mehr über das Promotions-Projekt wissen wollen, können Sie sich gerne mit mir in Verbindung setzen.

Im Voraus bereits vielen Dank für Ihre Teilnahme und Ihre Unterstützung.

Herzliche Grüße

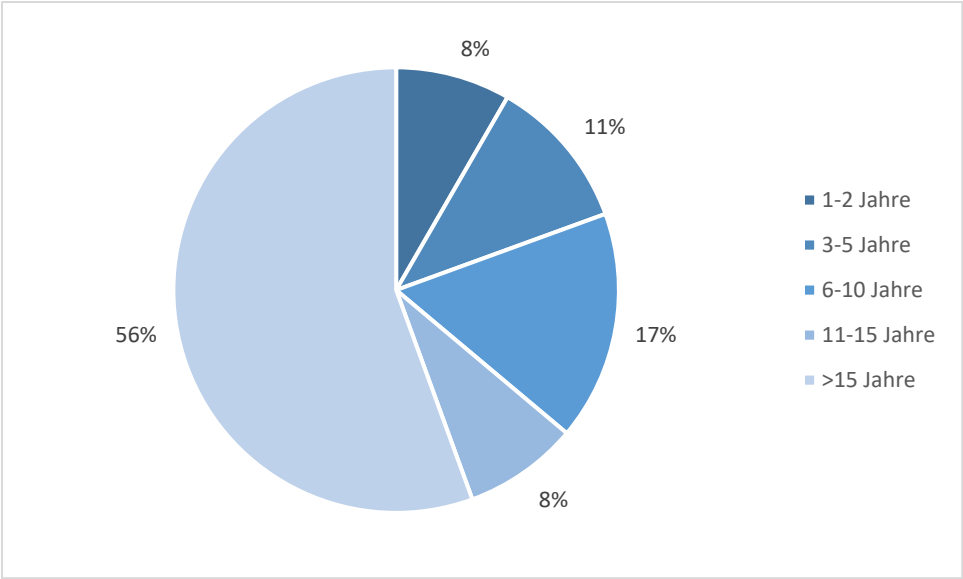
Kathrin Kuchta

Goethe-Universität Frankfurt | Campus Westend

Theodor-W.-Adorno-Platz 6 | D-60629 Frankfurt am Main

E-Mail: kuchta@psych.uni-frankfurt.de

Anhang B2: Lehrerfahrung der Teilnehmenden bei der Expertenbefragung LehrerInnen



Anhang B3: Fragebogen Expertenbefragung LehrerInnen

Instruktion/ Seite 1

Liebe Versuchsteilnehmerin, lieber Versuchsteilnehmer,

vielen Dank, dass Sie an der folgenden Studie zur Bewertung von Videosequenzen teilnehmen. Die Studie erfolgt im Rahmen meiner Dissertation zum Thema

"Förderung von Empathie durch Service-Learning-Veranstaltungen an Beispielen aus der Lehrerbildung" an der Goethe-Universität in Frankfurt, pädagogische

Psychologie. Wenn Sie nähere Informationen wünschen oder Rückfragen haben, finden Sie meine Kontaktdaten am Ende der Studie.

Im Folgenden werden Ihnen insgesamt 9 kurze Videosituationen gezeigt, zu denen Sie einige Fragen zur Bewertung des Gezeigten beantworten sollen. Es gibt bei der

Beantwortung kein "richtig" oder "falsch", da es um Ihre persönliche Einschätzung und Expertise als LehrerInnen geht. Ihre Antworten werden dazu dienen, zwei Tests

zur Messung von Empathie mitzuentwickeln. Die Bearbeitung wird etwa 20 bis 30 Minuten in Anspruch nehmen und Sie können die Videosequenzen gerne mehrmals

anschauen.

Um Ihre Angaben anonymisiert zu speichern, ist es notwendig, dass Sie einen persönlichen Code erstellen. Ein Rückschluss auf Ihre Person ist nicht möglich.

Ihr persönlicher Code wird folgendermaßen gebildet:

1. Buchstabe Vorname der Mutter

2. Buchstabe eigener Vorname

Geburtstag Mutter mit führender Null

Eigener Geburtsmonat mit führender Null

Beispiel:

Vorname der Mutter: Erika

Ihr Vorname: Daniel

Geburtstag der Mutter: 05. Juli

Eigener Geburtsmonat: 12.03.1985

E A 0 5 0 3

Bitte geben Sie nun Ihren persönlichen Code ein:

Seite 2

Hinweis: Da die meisten Aufnahmen im Rahmen von groß angelegten Unterrichtsbeobachtungen entstanden sind, kann es vorkommen, dass der besseren Zuordnung halber einige SchülerInnen Schilder mit Zahlen um den Hals haben. Das hat keinen wertenden Charakter und braucht Sie nicht weiter zu irritieren. Ebenfalls kann es von Videosequenz zu Videosequenz aufgrund verschiedener

Aufnahmebedingungen Lautstärkeschwankungen geben - stellen Sie die Lautstärke also nicht zu laut ein, sondern justieren Sie lieber nach.

Demografie/ Seite 3

1. Ihr Alter (in Jahren):

Sie sind:

- a. Weiblich
- b. Männlich
- c. Keine Angabe

2. Sind Sie noch aktiv im Schuldienst?

- a. Ja
- b. Nein

3. Wie lange unterrichten Sie bereits (oder haben Sie unterrichtet)?

- a. Gerade im Referendariat
- b. 1 - 2 Jahre
- c. 3 - 5 Jahre
- d. 6 - 10 Jahre
- e. 11 - 15 Jahre
- f. Mehr als 15 Jahre

4. Welche Fächer unterrichten Sie (oder haben Sie unterrichtet)?

5. In welchem Bundesland arbeiten Sie?

6. In welcher Schulform unterrichten Sie (oder haben Sie unterrichtet)?

Definition Empathie/Seite 4

Eine Definition von Empathie: „Als Empathie wird die Fähigkeit bezeichnet sich in die Gefühls- und Erlebenswelt einer anderen Person hineinzusetzen. Ganz charakteristisch ist auch das Mitempfinden der Gefühlsqualität der anderen Person, im Bewusstsein, dass das Erleben der anderen Person der Auslöser dafür ist.“

7. Wie wichtig ist es sich als LehrerIn empathisch gegenüber seinen/ihren SchülerInnen zu verhalten?

- a. Gar nicht wichtig
- b. Weniger wichtig
- c. Wichtig
- d. Sehr wichtig

8. Entspricht empathisches Verhalten „professionellem Lehrer-Verhalten“?

- a. Ja
- b. Nein

Bspw. Vignette 27/Seite 5

Der Lehrer spricht mit den Schülern über die Aufgabe. Der Schüler im blauen Oberteil im Hintergrund blättert in den Unterlagen. Er scheint etwas zu suchen. Die folgenden Fragen beziehen sich auf ihn.

Vignette 27 zum Abspielen

- 1. Welche Emotionen konnten Sie hauptsächlich bei dem Schüler sehen? (Mehrfachnennung möglich)**
 - a. Freude
 - b. Anspannung
 - c. Ratlosigkeit
 - d. Belustigung
 - e. Stolz
 - f. Interesse

- 2. In welcher Intensität haben Sie die folgenden Emotionen hauptsächlich erkennen können? 1= Gar nicht, 2=Sehr schwach, 3= Schwach, 4=Mittelmäßig, 5=Stark, 6=Sehr stark**
 - g. Freude
 - h. Ärger
 - i. Angst
 - j. Traurigkeit
 - k. Überraschung/Neugier
 - l. Ekel

- 3. Wie haben Sie sich selbst beim Anschauen gefühlt?**

- 4. Ein professionelles, empathisches Lehrerverhalten dem Schüler gegenüber hätte folgendermaßen ausgesehen:**

Bitte sortieren Sie die Antworten nach ihrer Relevanz (das Relevanteste auf Position 1).

 - a. Das gezeigte Verhalten des Lehrers in dem Clip.
 - b. Versuchen gemeinsam mit ihm die Lösung herauszufinden.
 - c. Den Schüler fragen, womit man ihm helfen kann, da er offensichtlich nach einer Antwort auf die Fragen sucht. Man könnte sich dabei zu ihm setzen und ihn beruhigen, so dass er nochmal in Ruhe nachdenken kann.
 - d. Anderes
- 4.1 Welches "Andere" Verhalten wäre angebracht gewesen?**

(falls Sie bei der Zuordnungsaufgabe "Anderes" Verhalten auf die ersten Positionen gesetzt haben)

- 5. Gibt es sonst noch etwas, das Sie dazu sagen möchten?**

Anhang B4: Fragen in der Expertenbefragung mit LehrerInnen

Item	Frage	Fragenformat	Scoring
1.	Welche Emotionen konnten Sie hauptsächlich bei dem Schüler sehen? (Mehrfachnennung möglich)	6 Antwortoptionen	1 Punkt/richtiger Antwort → Bilden eines Summenscores
2.	In welcher Intensität haben Sie die folgenden Emotionen hauptsächlich erkennen können?	6 Basisemotionen und Ratingskala von 1(gar nicht) bis 6 (sehr stark)	1 Punkt/richtiger Antwort (=Mittelwert der ExpertInnenbefragung); 0,5 Punkte für Werte +/- 1 SD → Bilden eines Summenscores
3.	Wie haben Sie sich selbst beim Anschauen gefühlt?	Freifeldantwort	%
4.	Wenn Sie in der Situation als Lehrperson dabei gewesen wären, welches Verhalten wäre professionell und empathisch gewesen?	4 Antwortoptionen, die von bestes Verhalten bis am wenigsten günstiges Verhalten und die Möglichkeit ein alternatives Verhalten zu benennen	%
5.	Welches "Andere" Verhalten wäre angebracht gewesen?	Freifeldantwort	%
6.	Gibt es sonst noch etwas, das Sie dazu sagen möchten?	Freifeldantwort	%

Anhang B5: Tabellarische Darstellung Schwierigkeiten LehrerInnenbefragung

Vignette	SchwierigkeitEscore	SchwierigkeitIscore	n
V2/T1	85,29	48,77	34
V2/T2			
V6/T1	60,00	64,05	35
V6/T2			
V8	66,67	75,00	12
V9	80,56	62,50	12
V10	52,27	61,36	22
V11	38,46	58,33	13
V14	84,62	66,03	13
V19	58,97	62,18	13
V21/T1	73,53	56,13	34
V21/T2			
V24	45,45	71,59	22
V25/T1	67,68	64,39	33
V25/T2			
V27	61,67	55,42	20
V28	53,97	55,16	21
V30	60,00	64,17	20

Anhang B6: Auswertung der zusätzlichen Verhaltensantwort

Unbearbeitete Rohdaten; *Vorschläge* thematisch unterschieden; 2.-4. *Nennung*: inhaltlich gleiche Antworten

Vignette 2

1. Dem Schüler (im hellgrauen Oberteil) zeigen, dass man die Genervtheit und den Ärger versteht, aber ihn gleichzeitig um Geduld bitten, den Lösungsweg des Schülers an der Tafel anzuhören - auch wenn man weiß, dass er es kann.
2. Das gezeigte Verhalten
3. Anderes

Vorschlag:	2. Nennung	3.Nennung	Komprimierte Antwort
Einen Moment bitte: Lass und erst schauen, wie x auf die Lösung kommt. Dann reden wir über dein Anliegen. Du fragst dich, warum x an der Tafel ist? Ist es dir wichtig, dass ich frage, wer an der Tafel rechnen möchte? Könnt ihr mir sagen, wie ihr am besten lernen könnt, wie merkt ihr euch Lösungswege am besten?(Methoden entwickeln in der alle Schüler mit eingezogen sind. Einer an der Tafel war schon vor 30 Jahren nur langweilig.)	Gruppenarbeit mit einer vorgegebenen Lösung zur Überprüfung?(Der Lehrer hat sich für einen Schülervortrag entschieden. Dieser Schüler hat sich gemeldet, vielleicht als einziger? Der Lehrer kann im Vorfeld besprechen, ob er jemanden dran nehmen kann, der sich nicht meldet? Es ist schwierig, immer die gleichen SuS vorn arbeiten zu lassen....)		
Den Schüler freundlich auf seine Überheblichkeit hinweisen			
Auf die angebotene Diskussion nicht einsteigen - wohl das Beste, was der L. machen konnte,...und vergessen, dass der Schüler gemosert hat (ihn nicht das nächste Mal zur Strafe an die Tafel holen.) Die S. scheinen den Fehler intern zu diskutieren. Super, laufen lassen. Zu lernen, nicht alles Gehörte, gehört zu haben, gehört zum Job.			
Ich hätte genauso reagiert. Es hätte in der Gesamtsituation keinen Sinn gehabt besonders	Probleme mit den Begriffen Ärger und Genervt : Nicht trennscharf, hier stark überschneidend. Wie soll man	Die Klasse wirkte im ganzen unruhig. Bereits vor dem HA-Vergleich ist zu empfehlen den	Im Vorfeld für eine ruhige Arbeitsatmosphäre sorgen.

empathisch zu reagieren, weil dadurch die Aufmerksamkeit auf den grauen Schüler gelenkt worden wäre. Insgesamt wäre es wahrscheinlich sinnvoller gewesen im Vorfeld für Ruhe zu sorgen, damit der Schüler an der Tafel konzentrierter arbeiten kann.	hier antworten? Verhaltensvorschläge: Den Schüler an der Tafel zurecht weisen. NEIN! Die Klasse zur Ruhe mahnen. Aber JA! Wieso sind diese beiden Verhaltensaspekte in einer Antwort verbunden? Die Klasse zur Ruhe mahnen. ist außerdem kein Lehrerverhalten dem Schüler gegenüber, denn der ist fast die ganze Zeit über ruhig und hört lange interessiert (!) zu. Diese Emotion wird aber übrigens auch in der 2. Frage nicht angeboten.	SchülerInnen das von ihnen erwartete Verhalten zu verbalisieren (Respektvoller Umgang, nicht die Kontrolle verlieren, Melderegeln einhalten,...). Durch eine andere Position in der Klasse kann der Lehrer schon im Vorfeld mögliche Kontrollverluste erkennen und präventiv eingreifen (z.B als sich der Schüler an den Kopf greift und mit seinen Nachbarn ein Gespräch anfängt).
--	---	---

Unterstützung anbieten
 ... leider kann ich den Schüler im grauen Oberteil nur zu gut verstehen! Aus seiner Sicht wird er beim lernen ausgebremst!

4. Den Schüler an der Tafel zurecht weisen und die Klasse zur Ruhe mahnen.

Vignette 6

1. Der Schülerin an der Tafel einen Moment Zeit geben und ihre Gedanken zu der gestellten Frage anhören.
2. Das Mädchen an der Tafel weiterhin mit einbeziehen, sich ihr immer wieder zuwenden und auch loben.
3. Anderes

Vorschlag:	2. Nennung	3.Nennung	4. Nennung	Kumulierte Antwort
DIE SCHÜLERIN AUF IHREN PLATZ ZURÜCKSCHICKEN UND MIT DER GESAMTEN GRUPPE REDEN.	DER SCHÜLERIN MEHR ZEIT GEBEN, SIE SOLL DANACH IHRE MITSCHÜLERINNEN AUFRUFEN UND SICH WEITERE TIPPS GEBEN LASSEN. DIE KLASSE GEHT AUF IHRE BEREITS VORHANDENEN ERGEBNISSE EIN, SAGT IHR, WO SIE ETWAS EINZEICHNEN SOLL. ABSCHLIEßEND KANN DIE	Die Schülerin darf sich selbst Hilfe holen (Publikums-Joker), indem sie Mitschüler/innen dran nimmt.	Es ist ok, die Klasse nach dem Begriff zu fragen, allerdings darf sich daraus keine Diskussion entwickeln. Man hätte eine weitere Meinung hören sollen, damit die Schülerin vorne	Die Schülerin darf sich selbst Hilfe holen, indem sie MitschülerInnen aufruft und sich weitere Tipps geben lässt. Gemeinsam wird das Tafelbild erarbeitet. Abschließend kann die Schülerin die Frage noch einmal zusammenfassend beantworten.

KLASSE SICH DAZU ÄUßERN. DER BALL BLEIBT BEI DER SCHÜLERIN. REGEL: WER AN DER TAFEL STEHT IST NUR VORNE - NICHT ALLEINE!	SCHÜLERIN DIE FRAGE NOCH EINMAL ZUSAMMENFASSEND BEANTWORTEN, SO DASS SIE BETEILIGT BLEIBT UND DAS GEFÜHL HAT, SIE HAT AN DER LÖSUNG MITGEWIRKT.	weitermachen kann.
DEM MÄDCHEN EINEN TIPP GEBEN, SIE ERMUNTERN		
AUCH WENN EIN SUS AN DER TAFEL STEHT MUSS DER REST DER KLASSE MITEINBEZOGEN WERDEN KÖNNEN!		
warum hat sie das Mädchen an die Tafel geholt, die Lehrerin sollte sich Notizen machen, die sie danach aufgreift		

4. Das gezeigte Verhalten

Vignette 8

1. Ansprechen, dass sich die Aussage erstmal lustig anhört, aber im Anschluss erklären, dass Bananen-Anbau im Gewächshaus auch bei uns möglich ist.
2. Das Gelächter und Geflüstere unterbinden und den Schüler, der mit „Bananenbaum“ geantwortet hat, fragen, ob er sich bei der ein oder anderen Sorte unsicher ist und jetzt gerne Fragen stellen könne.
3. Anderes:/ Ein anderes Verhalten

Vorschlag:

Den anderen Schülern klar machen, dass es Klassenregel ist, nicht über die Wortmeldungen anderer zu lachen. Danach die gleichen Fragen, wie die Lehrerin sie stellt.

Das Gelächter wäre unterbunden worden, wenn die Lehrerin gesagt hätte, dass ihre Frage wohl nicht für jeden eindeutig war und dass der Junge bezüglich ... (s.o.) Recht habe.

4. Das gezeigte Verhalten

Vignette 9

1. Ihr mit der Formel weiterhelfen, damit sie sich die Formel zukünftig besser merken und ohne große Verunsicherung weiter am Lösungsweg arbeiten kann. Außerdem leise mit ihr sprechen, damit sie sich nicht vor der Klasse vorgeführt vorkommt.
2. Ihre Frage ohne weitere Kommentare beantworten.
3. Anderes/Ein anderes Verhalten

Vorschlag:

Ich hätte versucht mit Fragen & Hilfestellungen die Schülerin dazu anzuleiten selbst auf die Lösung zu kommen (bzw. ihr gezeigt wie man an die Formel rankommt). Zum Beispiel über ein Blick in das Mathematikbuch (hier gibt es IMMER Beispielaufgaben) oder ich hätte eine Regelheft mit Beispielaufgaben. Damit wäre Sie nicht vorgeführt worden, hätte sich aber die Lösung (Formel) selber suchen müssen und damit hoffentlich besser gemerkt, als wenn ich die Formel genannt hätte!).

4. Das gezeigte Verhalten

Vignette 10

1. Aufmunternd lächeln und mit ihr gemeinsam eine Lösung besprechen.
2. Der Schülerin anbieten, dass sie die Unterlagen ihrer Sitznachbarin mitnutzen kann.
3. Das gezeigte Verhalten.
4. Anderes:

Vorschlag:	2. Nennung	3.Nennung	4. Nennung	Komprimierte Antwort
Bist bist neugierig, ob du die richtige Lösung gefunden hast? Antwort abwarten. Ja, die Lösung stimmt. Du hast einen anderen Weg genommen. Dieser Weg ist spannend, magst du ihn den anderen vorstellen?(Ich finde es wichtig, Schüler zu bestärken und ihnen Mut zu machen für alternative Lösungen.)	Ein klares Lob ohne sprachliche Abwertungen	evtl. den Lösungsweg auf einem Handout darstellen lassen, um es der Klasse zu einem späteren Zeitpunkt vorstellen zu lassen. (Der Lehrer hat den Schüler wahrgenommen und ihn gelobt. Das bestärkt den Schüler und ist in der Situation der freien Arbeit angemessen und sinnvoll.)	In die Hocke gehen, auf Augenhöhe das Geleistete wertschätzen, den Schüler in seiner Kreativität (= originelle Ansätze, die zu einem Ergebnis führen) loben. Ihn, quasi unter 4 Augen, nur für die Umstehenden hörbar. Ihn nicht erhöhen (Musterschüler).(Mir tat die gebückte Körperhaltung des S. weh - wie ein geprügelter Hund, nicht wie jemand, der auf dein Können stolz sein darf...)	Den originellen Lösungsweg klar loben.
Den Lösungsweg vorstellen zu lassen hat keinen Sinn, da sich die nicht gewünschten Wege dann in der Klasse verbreiten und der Standard/Soll verwässert wird. Das würde sich rächen.				
Die Bemerkung des Lehrers ist missverständlich und die raktion des Schülers zeigt auch, dass er denkt, dass der Lehrer ihm eine gute Lösung eher nicht zugetraut hat.	Die Äußerung Ich bin immer wieder überrascht, was in euren Gehirnzellen vor geht. ist ambivalent. Das meist richtig und ganz am Schluß das toll gehen dagegen unter.			

Vignette 14

1. Der Schülerin glauben, dass Sie nicht weiß wie es geht und sie fragen, wo genau die Schwierigkeiten liegen. Ihre Gefühle ernst nehmen und ihr das Angebot machen, sie beim Finden des Lösungsweges zu unterstützen.

2. Kommentieren, dass sie anscheinend keine Lust hat und sie bitten, sich konzentriert und ohne Handy an die Aufgabe zu setzen.
3. Das gezeigte Verhalten
4. Anderes:/ Ein anderes Verhalten

Vorschlag:

Wenn es sich um Mathematik handelt, ist es besonders wichtig, den Mädchen zu zeigen, dass auch sie das können. Wurde das schon von Kollegen zuvor versäumt oder von den Eltern unbewusst noch verstärkt, ist ein Unterrichten sicher auch schwer. Es kann jedoch sein, dass das Mädchen sich nichts zutraut und es deswegen gar nicht erst versucht und sofort in die Abwehrhaltung geht. Wäre das so, sollten vielleicht für eine Woche die Jungen und Mädchen separat unterrichtet werden, sodass das pubertäre Verhalten bezüglich des anderen Geschlechts nicht das Denken verhindert. Das Mädchen würde auch lernen, dass sie sich an der Kollegin nicht reiben muss, sondern von dieser Hilfe bekommen kann.

Leider ist es aber auch oft so, dass Schüler oben gezeigtes Verhalten an den Tag legen. Da kommen Fragen zu Aufgaben, diese Sie noch nicht einmal gelesen haben.

Vignette 21

1. Verständnis dafür ausdrücken, dass andere Dinge vielleicht spannender erscheinen als der Unterricht und sie damit dezent merken lassen, dass man ihr Verhalten bemerkt hat. Im weiteren Verlauf versuchen ihr Interesse am Inhalt des Unterrichts zu wecken.
2. Die Schülerin leise darauf hinweisen, dass sie private Angelegenheiten bitte erst nach der Stunde klären soll.
3. Anderes:

Vorschlag:	2. Nennung	Komprimierte Antwort
Antippen - Blick auf den Brief - auf die Schultasche zeigen. (Der Lehrer scheint sich sehr toff zu finden.)		
Da er ja wohl mehrfach gesagt hat, sie soll den Brief wegpacken ... klare Ansage, allerdings von vorne und ohne Schreien.	Besser von Vorne (Sichtbereich der S.) ansprechen. Eventuell darauf achten, dass nicht der Eindruck entstehen KANN, dass er der jungen Frau in den Ausschnitt schaut.	Die Schülerin von vorne und in einer normalen Lautstärke ansprechen und sie bitten, den Brief wegzupacken.
den Brief abnehmen und weiter im Programm... Gespräch nach der Stunde		
Wenig glaubwürdig, dass ein Lehrer seine Schülerin so anbrüllt und auf das Lachen der Klassenkameraden so indifferent reagiert. Das Lachen würde ich den Schülerinnen verbieten; sie machen sich über sie lustig. Das Dilemma ist, der Lehrer hat das Lachen provoziert und damit das Sich-Lustig-Machen .		
Die in Bezug auf die Emotionen hier vorgelegten Begriffe sind für mich oft nicht treffgenau (gilt auch für die vorhergehenden Videosituationen). In der obigen Sequenz sind z.B. für mich wesentlich auch Erleichterung und Entspannung (nach dem ersten Schreck) erkennbar. (--> Freude ???? Naja, vielleicht auch ein bisschen.) Und: Überraschung/Neugier . Die Überraschung beim ertappt-werden ist sehr stark zu erkennen (--> Schreck). Aber Neugier???? Das Begriffspaar ist für mich unpassend. In der Einleitung wurde die Information gegeben, dass die Schülerin bereits mehrfach (in dieser Stunde?) beim Briefe-lesen erwischt worden ist. Da halte ich die ersten beiden Verhaltensoptionen nicht mehr für angemessen. Andererseits formuliert der Lehrer Einmal ist ja schon		

peinlich ... In Bezug auf die Angemessenheit von Reaktionen stellt sich für mich auch hier wieder die Frage nach den Hintergrund-/Rahmenbedingungen: Was für ein Typ ist die Schülerin? Welche Stellung in der Klasse hat sie? Kann man ihr zumuten, dass alle anderen über sie lachen? Es gibt Schülerpersönlichkeiten und soziale Bezüge in der Klasse, bei denen ich diese Frage eindeutig mit JA und andere, bei denen ich sie eindeutig mit NEIN beantworten würde.

Auch Lehrer haben Emotionen. Ich kenne das nur zu gut, dass mich diese Nebentätigkeiten der Schüler massiv nerven und ich dann ein Ventil brauche. Für mich klang der Lehrer hier authentisch. Das Verhalten ist zwar grenzüberschreitend, aber nicht böse gemeint. Diese Absicht dahinter ist wichtig für das Verhältnis zu den Schülern.

Ich kann den Lehrer verstehen, weil sie wiederholt seiner Aufforderung nicht nachgekommen ist. Allerdings ist es nicht nötig, eine Schülerin so zu erschrecken. Wird sie jetzt arbeiten können?

Die Reaktion der Mitschülerinnen zeigt, dass der Lehrer die Sache nicht so ernst nimmt

4. Das gezeigte Verhalten.

Vignette 25

1. Durch eine freundliche Ansprache die Atmosphäre entspannen und den Druck nehmen. Dann die Schülerin und den Rest der Klasse zu Antworten ermuntern und zum Mitdenken anregen.
2. Dem Mädchen genauso wie gegenüber dem Rest der Klasse fragen, wo die Schwierigkeiten mit der Aufgabe liegen und eventuell nochmal einen didaktischen Schritt zurückgehen.
3. Anderes:

Vorschlag:	2. Nennung	3. Nennung	Komprimierte Antwort
Den SchülerInnen die Möglichkeit geben sich am Gruppentisch über die Formulierung auszutauschen und so die Angst zu nehmen vor der ganzen Klasse etwas Falsches zu sagen.	Die Schüler das Problem entwickeln lassen.		
Sie hätte das Mädchen neben der Angesprochenen ansprechen können, weil diese sich gemeldet hatte(Im Frontalunterricht gibt es solche Situationen immer wieder. Besser ist es, die Schüler in diese Lösungen so mit einbeziehen, dass sie selbst darauf kommen... Frage / Antwort kann dann evtl. schneller folgen)	Ich hatte meine wenig positiven Gefühle auch aus dem Grund, dass sich ein Schüler meldet, der also helfen könnte und weil es zuvor keinen Blickkontakt zwischen Lehrerin und Schülerin gab. Das Abfragen wirkte wie ein Schuss aus dem Hinterhalt auf mich		

<p>Alltägliche Situation... anstrengend (siehe oben) und manchmal kann man auch einfach nur hoffen, dass die Schüler nicht empathisch genug sind, diese Mischung aus Frust/Zorn/Verzweiflung, die in solchen Situation aufkommen kann, mitbekommen.</p>	<p>Eine Erklärung geben, einen Schritt zurück gehen...(wenn niemand es weiß, stimmt etwas nicht)</p>	<p>Wenn sich niemand in der Klasse meldet und auch die erste unfreiwillige Schülerin nicht weiterkommt, muss ich die Klasse fragen, woran dies liegt. Alternativ biete ich gleich ein kurzes Warm-up an und löse die Aufgabe im Anschluss gemeinsam mit der Klasse. Im Anschluss kann ich noch weitere Beispielaufgaben geben und die Klasse bitte sich selbst zu überprüfen ob sie den Lösungsweg verstanden haben.</p>	<p>Die Schüler drannehmen, die sich melden und ggf. die Klasse das Problem gemeinsam entwickeln lassen, da es vielleicht Verständnisprobleme gibt.</p>
<p>Die Kollegin hätte aktiver sein können, indem sie die offenbar nicht verstandene Frage noch einmal erläutert. Da sich sonst auch niemand meldet, haben offenbar alle Schüler/innen ein Verständnisproblem. Die für die Schülerin peinliche Situation des nicht antworten könnens, hätte man verkürzen müssen.</p>			
<p>Ich weiß nicht, ob das Schweigen der Klasse an der Schwierigkeit liegt. Ich kenne es leider zu oft, dass Klassen dem Unterrichtsthema kein Interesse bzw. Engagement entgegen bringen können. Bei Untätigkeit der Schüler ist Unterricht schwierig bis unmöglich. Ich kann die Klasse nicht zur Arbeit tragen. Die Schüler müssen sich schon selbst bewegen.</p>			

4. Das gezeigte Verhalten

Vignette 27

1. Den Schüler fragen, womit man ihm helfen kann, da er offensichtlich nach einer Antwort auf die Fragen sucht. Man könnte sich dabei zu ihm setzen und ihn beruhigen, so dass er nochmal in Ruhe nachdenken kann.
2. Versuchen gemeinsam mit ihm die Lösung herauszufinden.
3. Anderes:

Vorschlag:	2. Nennung	3.Nennung	Komprimierte Antwort
<p>Geduldig zugewandt(Das insistierende Beharren des Lehrers auf seiner Linie nervt)</p>	<p>Den Schüler nach seinen Problemen fragen und ihm mehr zeit zum Überlegen geben. später wieder zu ihm zurückkommen.</p>	<p>Weiter mit dem Schüler sprechen, mit dem er zuerst sprach. Dann kommt der zweite S. aus der Stresssituation und findet, was er eigentlich gesucht hat. (vielleicht) Dann kann man ja wieder</p>	<p>Dem Schüler mehr Zeit zum Suchen und Überlegen geben. Wenn er nicht weiterkommt, unterstützende Hilfe anbieten.</p>

	zurückschwenken, auf den suchenden S.
Oft suchen Schüler etwas, weil sie nicht zugehört haben. Ich könnte mir vorstellen, dass dieser Schüler etwas sucht, was ihm zur Lösung des derzeitigen Problems weiterhelfen kann. Deshalb bräuchte er hier bessere unterstützende Hilfe.	Eine Regel nennen
In dieser Kameraeinstellung hat der Lehrer den Schüler im Hintergrund nicht bemerkt. Dieser Schüler sucht zwar. Andere Schüler spielen das Spiel: Nur nicht bewegen, dann bemerkt er mich nicht. Diese Situation sollte ein Lehrer bemerken können und dann den Schüler direkt darauf ansprechen ohne ihn zu maßregeln. Ermuntern Motivation, Grund und Ziel seines Handelns zu nennen.(Der Lehrer hatte einen Grund, sich dem anderen Schüler zu zu wenden. Das Gespräch mit ihm fordert seine ganze Aufmerksamkeit und Energie. Lehrer sind auch nur Menschen und können weder alles wahrnehmen, noch auf alles reagieren.)	

4. Das gezeigte Verhalten des Lehrers in dem Clip.

Vignette 30

1. Aufmunternd lächeln und beide motivieren sich mit der Aufgabe zu beschäftigen. Anbieten sie bei Problemen gerne zu unterstützen.
2. Das gezeigte Verhalten
3. Die Schüler in ihrem eigenen Arbeitsstil weitermachen lassen.
4. Anderes:

Vorschlag:	2. Nennung	3.Nennung	Kumulierte Antwort
Die Lehrerin steht wie eine Wächterin hinter dem Schüler			
Ja, die Lösung ist richtig. Hast du verstanden, dass wir die Aufgabe erst mündlich besprechen? Antwort abwarten. Ich habe mir das anders vorgestellt, dass ihr erst die Chance bekommt alleine die Lösung zu finden. Dann werden wir es mündlich zusammen durchgehen und besprechen. Schau doch mal, wie die Aufgabe am besten/geschicktesten lösen kannst.(Ich glaube, dass diese Gespräche in der Schule heute noch genauso ablaufen. Auf diese Schlussfolgerung komme ich, da ich selbst 3 Kinder habe und ähnliches gehört habe. Schade, hätten wir mehr Empathie			

würden wir die Kinder stärken, in die Eigenverantwortung bringen und ihnen vorleben, wie Sprache Beziehungen gestaltet.)

wo sind denn eure Ergebnisse?
 ich habe die Erfahrung gemacht, dass es gut ist, in solchen Situationen über die Sache zu reden: Ergebnis und nicht über den eigenen Ärger: ihr müsst jetzt aber..... leider ertappe ich mich auch bei solchen Frustäußerungen)

i.A. der Häufigkeit des beobachteten Verhaltens, könnte man auch die Frage stellen, warum sich der S. so verhält. Man kann auch sich die Frage stellen, ob das Verhalten einer stärkeren Intervention Bedarf. M.E. versucht der Schüler cool zu sein - oder er hat einfach keine Lust, das zu tun, wozu er keine Lust hat - ein völlig normales, menschliches Verhaltensmuster...

Nachfragen, ob sie Hilfe brauchen (besonders intensiv scheinen die Schüler hier nicht zu arbeiten??)

Nachfragen, ob/weshalb er noch keinen Zugang zu der Aufgabe gefunden hat: Welche Informationen fehlen? Woran hapert es?

Fragen, warum der Schüler mit dem blauen Oberteil sich noch nicht mit der Frage beschäftigt

Nachfragen, ob/weshalb er noch keinen Zugang zu der Aufgabe gefunden hat: Welche Informationen fehlen? Woran hapert es? Wobei brauchen sie Hilfe?

Sich herauszuhalten als Lehrerin ist für mich oft schwer, da ich zu wissen meine, wie die Schüler es besser machen könnten. Ich will dann Zeit sparen. Das geht aber nicht! Lernzeit ist Erfahrungszeit. Die ist unterschiedlich lang für unterschiedliche Menschen.

Anhang B7: Pool an Emotionsworten für Item 1

1. Angst
2. Anspannung
3. Ärger
4. Belustigung
5. Freude
6. Frust
7. Genervtheit
8. Hilflosigkeit
9. Interesse
10. Irritation
11. Langeweile
12. Peinlichkeit
13. Ratlosigkeit
14. Schadenfreude
15. Scham
16. Schreck
17. Stolz
18. Überraschung
19. Unsicherheit
20. Verlegenheit

Anhang C

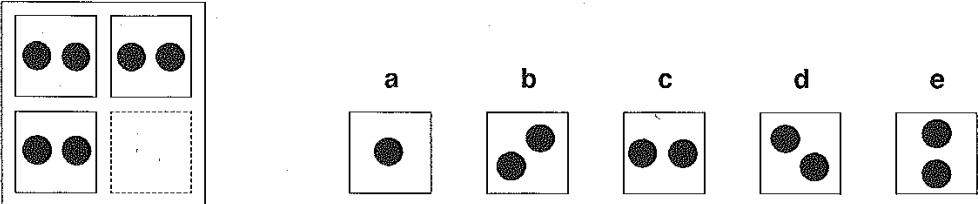
Anhang C1: Matrizentest CFT 20-R, Beispielitems

© Hogrefe Verlag GmbH & Co. KG, Göttingen

Im Folgenden sehen Sie **drei** **Beispielaufgaben**. Ihre Aufgabe ist es, rechts ein Kästchen mit **der** Zeichnung auszuwählen, die in das leere Kästchen links am besten hineinpasst, um den Kasten richtig zu vervollständigen.

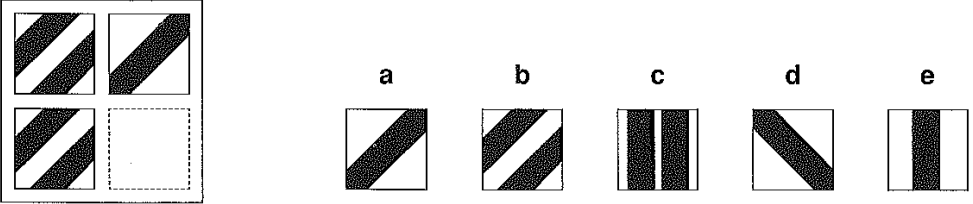
Die Beispielaufgaben dienen zum Verständnis. Nach den Beispielaufgaben folgen 15 Aufgaben, die auf dieselbe Weise gelöst werden sollen.

erstes Beispiel



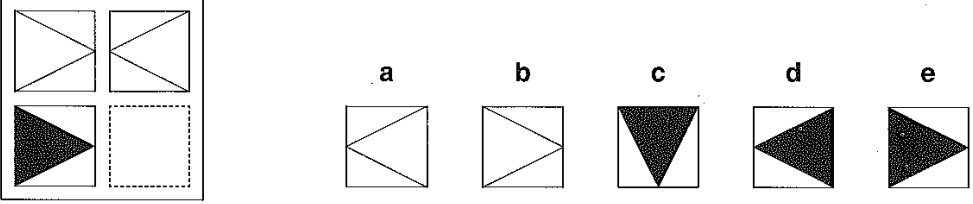
a b c d e

zweites Beispiel



a b c d e

drittes Beispiel



a b c d e

Im Folgenden werden Ihnen **15 Aufgaben** präsentiert, die auf den folgenden drei Seiten auf die vorher beschriebene Art gelöst werden sollen. Sie haben insgesamt **4 Minuten** Zeit alle 15 Aufgaben zu lösen.

Anhang C2: Big-Five-Kurzversion

Dauer: 2 Minuten

Inwieweit treffen die folgenden Aussagen auf Sie persönlich zu? Wählen Sie dazu bitte die entsprechende Kategorie und

kreuzen Sie sie rechts neben der jeweiligen Aussage an. Ein Beispiel:

Ich .. verbringe gerne Zeit mit anderen. 1=sehr unzutreffend; 2=eher unzutreffend; 3=weder noch; 4=eher unzutreffend; 5=sehr zutreffend

Kreuzen Sie z. B. das 2. Kästchen an, wenn Sie meinen, dass diese Aussage eher unzutreffend für Sie ist.

Ich ...

1. ... bin eher zurückhaltend, reserviert. (E)
2. ... neige dazu, andere zu kritisieren. (V)
3. ... erledige Aufgaben gründlich. (G)
4. ... werde leicht deprimiert, niedergeschlagen. (N)
5. ... bin vielseitig interessiert. (O)
6. ... bin begeisterungsfähig und kann andere leicht mitreißen. (E)
7. ... schenke anderen leicht Vertrauen, glaube an das Gute im Menschen. (V)
8. ... bin bequem, neige zur Faulheit. (G)
9. ... bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen. (N)
10. ... bin tief sinnig, denke gerne über Sachen nach. (O)
11. ... bin eher der „stille Typ“, wortkarg. (E)
12. ... kann mich kalt und distanziert verhalten. (V)
13. ... bin tüchtig und arbeite flott. (G)
14. ... mache mir viele Sorgen. (N)
15. ... habe eine aktive Vorstellungskraft, bin phantasievoll. (O)
16. ... gehe aus mir heraus, bin gesellig. (E)
17. ... kann mich schroff und abweisend anderen gegenüber verhalten. (V)
18. ... mache Pläne und führe sie auch durch. (G)
19. ... werde leicht nervös und unsicher. (N)
20. ... schätze künstlerische und ästhetische Eindrücke. (O)
21. ... habe nur wenig künstlerisches Interesse. (O)

Extraversion=E

Verträglichkeit=V

Gewissenhaftigkeit=G

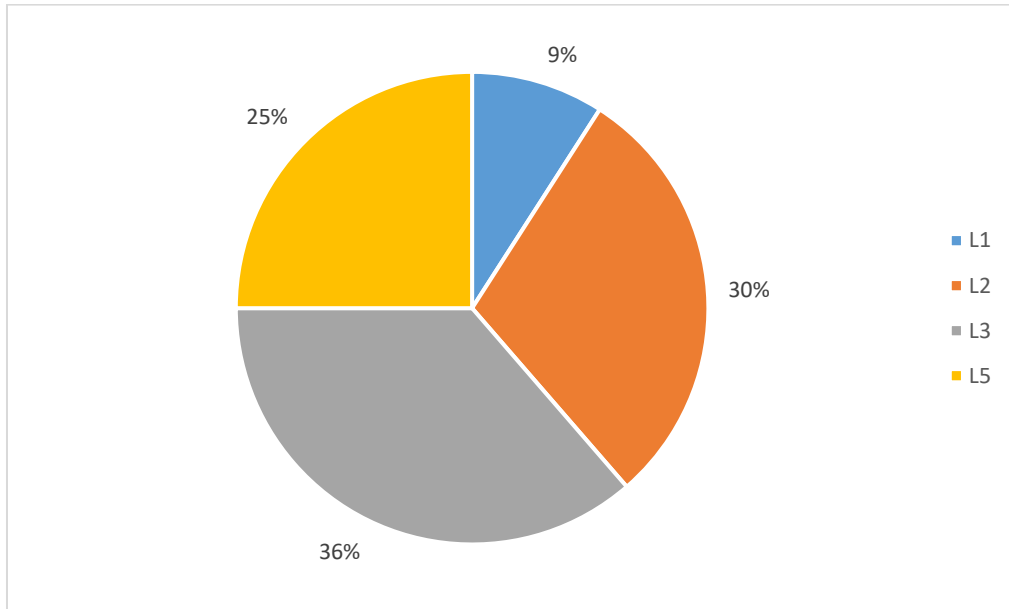
Neurotizismus=N

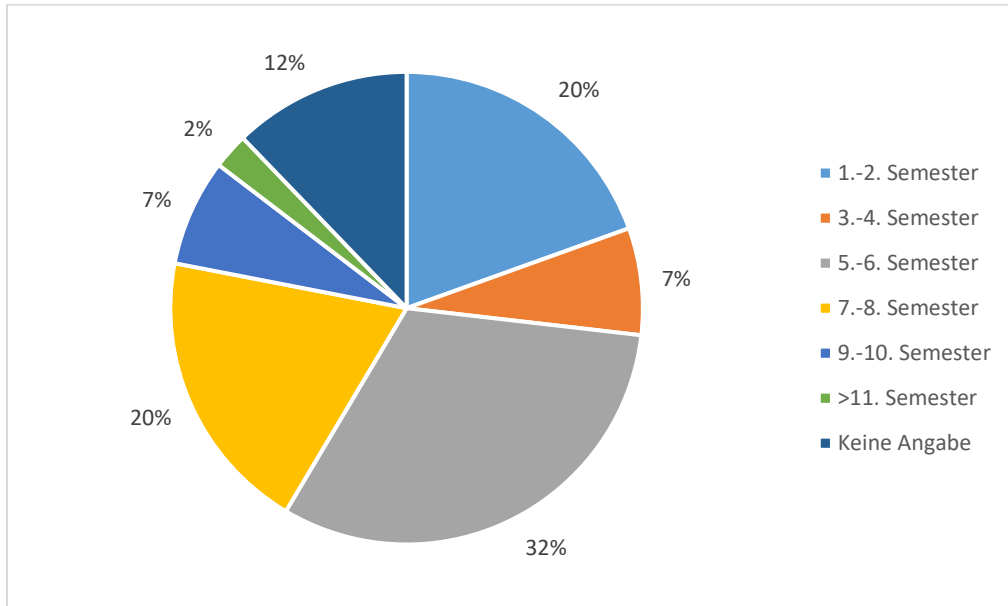
Offenheit für Erfahrungen=O

Anhang C3: KS-Test auf Normalverteilung (Validierungsstichprobe)

Variable	Normalverteilt?
ESEKfunktional	Ja (p=.200)
ESEKDysfunktional	Ja (p=.200)
Extraversion	Nein (p=.020)
Verträglichkeit	Ja (p=.168)
Gewissenhaftigkeit	Nein (p=.000)
Neurotizismus	Nein (p=.029)
Offenheit	Nein (p=.029)
PT	Ja (p=.200)
Fantasy	Ja (p=.191)
EC	Ja (p=.051)
PD	Ja (p=.200)
ScoreTEQ	Nein (p=.003)
EScore V2	Nein (p=.000)
IScore V2	Ja (p=.200)
VScore V2	Nein (p=.029)
EScore V6	Nein (p=.000)
IScore V6	Nein (p=.000)
VScore V6	Nein (p=.000)
EScore V8	Nein (p=.000)
IScore V8	Nein (p=.000)
VScore V8	Nein (p=.000)
EScore V9	Nein (p=.000)
IScore V9	Nein (p=.000)
VScore V9	Nein (p=.000)
EScore V10	Nein (p=.000)
IScore V10	Nein (p=.000)
VScore V10	Nein (p=.000)
EScore V14	Nein (p=.000)
IScore V14	Nein (p=.001)
VScore V14	Nein (p=.000)
EScore V21	Nein (p=.000)
IScore V21	Nein (p=.045)
VScore V21	Nein (p=.000)
EScore V25	Nein (p=.000)
IScore V25	Ja (p=.182)
VScore V25	Nein (p=.000)
EScore V27	Nein (p=.000)
IScore V27	Nein (p=.000)
VScore V27	Nein (p=.002)
EScore V30	Nein (p=.000)
IScore V30	Nein (p=.000)
VScore V30	Nein (p=.032)
GesamtEScoreT1	Nein (p=.000)
GesamtIScoreT1	Nein (p=.000)
GesamtVVT1	Nein (p=.001)
GesamtEScoreT2	Nein (p=.000)
GesamtIScoreT2	Ja (p=.156)
GesamtVVT2	Ja (p=.079)
ScoreCFT	Nein (p=.001)

Anhang C5: Studierte Lehramtsstudiengänge der teilnehmenden Studierenden der Pilotierung im SoSe17



Anhang C6: Besuchte Semester der teilnehmenden Studierenden der Pilotierung im SoSe17

Anhang C7: Korrelationen der Face Reader Analysen mit der eingeschätzten Intensität der zu beobachtenden Personen in den Videos und der angegebenen Intensität der eigenen Emotionen beim Anschauen des Videos

		Face-Reader Analysen													
		Freude V6	Freude V8	Traurig keit V8	Überr. / Neugier V8	Ekel V8	Freude V9	Ekel V9	Freude V21	Freude V25	Ärger V25	Ekel V25	Freude V27	Ärger V27	Ekel V27
EE V6	Angst	.59**													
IE V8	Traurig keit			.63**											
IE V8	Überr. /Neugi er				.56**										
EE V8	Freude		.41*	-.48*											
EE V8	Ärger			.42*											
IE V9	Freude						.68*								
IE V9	Ekel							.50*							

EE V9 Freude	.77**			
EE V9 Ärger	-.48*			
EE V9 Ekel		.42*		
EE V21 Freude			.62**	
EE V21 Ärger			-.43*	
IE V25 Ärger				.49*
IE V25 Angst				-.42*
IE V25 Überr. /Neugi er			.66**	
IE V25 Ekel				.44*
EE V25 Freude			.91**	

EE
V25
Überr.
/Neugi
er

.56**

EE
V27
Freude

.57**

.47*

EE
V27
Ekel

.56**

Anmerkungen. * $p < .05$, ** $p < .01$; $N=67$; EE=Eigene Emotion; IE=eingeschätzte Intensität der gesehenen Emotion; V=Vignette

Anhang D

Anhang D1: SL-Veranstaltungen, in denen für die Hauptstudie rekrutiert wurde

Informationen aus QIS/LSF oder den Projektwebseiten

1. **BW-E/Sb2: Motivieren in der Adoleszenz - Sich im Lehrerberuf erproben (L1-L5), Dr. Anita Püttmann:**

Inhalte: Die Adoleszenz ist eine Zeit der Auseinandersetzungen, sowohl mit Lehrern als auch mit Eltern und Gleichaltrigen. Besonders in der Schule kommt es immer wieder zu Motivationsproblemen.

Behandelt werden zunächst Themen wie: Grundbegriffe, Gegenstand, Aufgaben und Geschichte der Entwicklungspsychologie sowie Methoden der Entwicklungspsychologie (Erhebungsverfahren, Analyse-Auswertungsverfahren). Es werden Anlage- versus Umwelteinflüsse und andere Modellvorstellungen für die Erklärung von Entwicklung vorgestellt. Wir erarbeiten in diesem Zusammenhang weiterhin das Thema Motivation und Motive, besonders Leistungsmotivation.

Danach werden sowohl typische Entwicklungen im Jugendalter und frühem Erwachsenenalter als auch möglicherweise auftretende Adoleszenzprobleme und passende Reaktions- und Interventionsmöglichkeiten für Lehrer und Lehrerinnen besprochen und im Rollenspiel geübt. In diesem Seminar geht es nach der Reflektion der Theorie und des Erprobens im Seminar weiter darum, sich der Herausforderung der Heterogenität von Jugendlichen und deren schulischen Lernen in der Praxis zu stellen, die eigenen Kompetenzen und Überzeugungen zu überprüfen und sich bezüglich Lehren und der damit verbundenen Belastung und Verantwortung zu erproben.

Diese Lehrveranstaltung ist ein sogenanntes „Service Learning Seminar“. Service Learning oder auch „Lernen durch Verantwortung“ ist eine Lehr-/Lernform, die fachspezifische Inhalte mit ehrenamtlichem, gesellschaftlichem Engagement verbindet (Lerne und tue Gutes). Studierende arbeiten dabei mit meist gemeinnützigen Organisationen zusammen, die sich diese Hilfe sonst nicht leisten könnten. Zugleich soll ein realer Bedarf gedeckt werden, und inhaltlich relevantes Wissen zur Anwendung kommen. Dementsprechend profitieren beide Seiten von diesem Engagement.

Im Rahmen dieses Seminars unterstützen Sie im Schuljahr 2016/17 kontinuierlich einmal pro Woche für zwei Schulstunden (90 Minuten in den Fächern Deutsch, Mathe und falls möglich auch Fachtheoriewissen) Auszubildende, die sich allein mit den Anforderungen der Berufsschule schwertun. Es handelt sich hierbei also um Einzelfördermaßnahmen für Jugendliche zwischen 17 und 20 Jahren mit Lern-, Leistungs- und Sozialproblemen.

Das Seminar bietet Ihnen die Möglichkeit, sich in Ihrem späteren Tätigkeitsfeld (fachliche Kompetenz) zu erproben, überfachliche Kompetenzen auszubilden (Soft-Skills wie soziale Kompetenz, Empathiefähigkeit und Ausbildung eines sozialen Bewusstseins) und einen wertvollen Beitrag für die Gesellschaft zu leisten.

Für Ihr soziales Engagement stellt Ihnen der Träger des Jugendhilfezentrums am Ende des Schuljahres ein Zeugnis aus, was Sie bei späteren Bewerbungen Ihren Bewerbungsunterlagen beilegen können.

Die kooperierende Einrichtung befindet sich in Wiesbaden, bitte bedenken Sie das bei der Anmeldung.

2. BW-C/Sb2 (L1-L5): Umgang mit Diversität im Kontext von Flucht und Migration, Dr. Anne Seifert, Nora Iranee, Sophie Adler:

Inhalte: Das Service-Learning Seminar legt den Fokus auf wissenschaftsbezogene Projektreflexionen zur Thematik „Umgang mit Diversität im Kontext von Flucht und Migration“. Als zentrale Themen werden u.a. zu behandeln sein:

- Fremdheit und die Konstruktion des Fremden
- Machtverhältnisse und die Gefahr von Zuschreibungen im Kontext Schule
- Partizipation und Handlungsmächtigkeit von Kindern und Jugendlichen

Das Seminar bietet einigen TeilnehmerInnen die Möglichkeit, sich in schulischen Projekten zu engagieren und das Erlebte sowie das eigene Handeln im Seminar zu reflektieren. Die Kooperationsprojekte sind in den vergangenen Semestern von Studierenden in Zusammenarbeit mit Schulen und Vereinen im Raum Frankfurt entwickelt und erprobt worden. Beispiele sind die Auseinandersetzung mit dem Thema Zivilcourage vor der Fragestellung, wie dieses Thema zu einem „Thema der Schule“ gemacht werden kann (Mitarbeit in einer AG mit Schüler*innen der 9./10. Klasse) oder die Konzeption eines Wahlpflichtkurses unter dem Motto „Theater verbindet – Spielend Deutsch lernen“, für SchülerInnen der Intensiv- und Regelklassen.

Alle Studierenden, die sich nicht in den Projekten engagieren können oder möchten, begleiten die Projektentwicklung unterstützend und kritisch durch die Auseinandersetzung mit theoretischen Bezügen sowie ggf. Hospitationen. Eine mögliche Relationierung von Theorie und Praxis wird in den Seminarsitzungen gemeinsam interaktiv erarbeitet.

Das Seminar wird im SoSe 2018 im Modul BW-E angeboten. Eine Teilnahme über zwei Semester ist möglich und ausdrücklich erwünscht, da die Projekte an den Schulen über das gesamte Schuljahr laufen.

Teilnahmeanforderungen:

1. Teilnahme am Seminar ohne Engagement

Vorausgesetzt wird eine regelmäßige und aktive Teilnahme am Seminar, die intensive Auseinandersetzung mit wissenschaftlichen Texten zu obigen Themen sowie die Bereitschaft, sich kritisch und unterstützend in die Entwicklung der Projekte einzubringen.

2. Engagement in den schulischen Projekten

Für 16 Teilnehmende wird es die Möglichkeit geben, ausgehend von eigenen Interessen ein Projekt auszuwählen und dieses seminarbegleitend im Team mit anderen Studierenden und den AnsprechpartnerInnen vor Ort umzusetzen. Wer sich für ein Engagement in den Partnerprojekten entscheidet, sollte bereit sein, sich wöchentlich zwei Stunden in der Schule vor Ort einzubringen und zu engagieren. Die Zeitfenster hierfür erhalten Sie nach Anmeldung per Mail. Das Engagement ist freiwillig und erfolgt zusätzlich zum Seminar. Es ermöglicht einen Einblick in den Schulalltag und die unterschiedlichen Lebenswelten von Schülerinnen und Schülern mit und ohne Migrationserfahrung sowie die theoriebezogene Reflexion dieser Erfahrungen. Das Einbringen eigener Ideen, Interessen und Vorerfahrungen vor Ort ist ausdrücklich erwünscht. Alle Studierenden, die sich in den Kooperationsprojekten engagieren, erhalten am Ende ihres Engagements ein zusätzliches Zertifikat.

Falls Sie Interesse an einem seminarbegleitenden Engagement in den schulischen Projekten haben, wenden Sie sich bitte (zusätzlich zur Anmeldung in LSF) baldmöglichst per Mail an iranee@em.uni-frankfurt.de Sie erhalten daraufhin weitere Informationen. Bitte erläutern Sie in Ihrer Mail kurz Ihre Motivation für die Teilnahme an Seminar und Engagement und geben Sie an, ob Sie eine Teilnahme über zwei Semester in Erwägung ziehen.

3. Individuelle Förderung neu zugewanderter Kinder in Intensivklassen (L1 – Allgemeine Grundschuldidaktik), Prof'in Dr. Charlotte Röhner:

Inhalte: Der Sprachunterricht für neu zugewanderte Kinder wird in vielen Bundesländern in Form von Seiteneinsteigerklassen oder Intensivklassen erteilt. Der Unterricht in diesen Klassen weist ein Höchstmaß an Heterogenität im Hinblick auf Altersdifferenzen zwischen 6 bis 12 Jahren, vorausgegangene Schulerfahrungen, schulisches Lern- und Arbeitsverhalten, Leistungsdifferenzen und domänenspezifische Lernvoraussetzungen sowie im Hinblick auf die Alphabetisierung in anderem Sprach- und Schriftsystem bzw. keinerlei Alphabetisierung auf. Der Unterricht in diesen Klassen erfordert ein Höchstmaß an Individualisierung, um den heterogenen Lernvoraussetzungen gerecht zu werden. Im Seminar sollen Konzepte und Formen individueller Förderung neu zugewanderter Kinder in Intensivklassen untersucht und in Frankfurter Grundschulen praktisch erprobt werden. Dazu sollen Studierende einzelne Kinder in Intensivklassen begleiten und in ihrer Sprach- und Lernentwicklung unterstützen. Die diagnostischen und didaktisch-methodischen Grundlagen, die dazu erforderlich sind, werden im Seminar vermittelt.

4. BW-D/Sb2: Diagnostik auf individueller Ebene/ Diagnostik und Förderung der schriftsprachlichen bzw. mathematischen Kompetenz in der Grundschule und Sekundarstufe I (L1-L5), Prof. Dr. Ulrich Mehlem, Ulrike Suntheim, Constanze Fuchs, Aloysia Abraham:

Inhalte: Das über zwei Semester angelegte Seminar begleitet Projekte im Bereich der Schriftsprach-, und Lese- (Lerngruppen 1 und 3) sowie Rechenförderung (Lerngruppe 4) und wird an Grund- und Sekundarstufenschulen (Klasse 5/6) durchgeführt. Für ein Schuljahr fördern die Studierenden 2mal wöchentlich im gewählten Bereich Kleingruppen von 2-4 Schülern in der Grundschule oder Sekundarstufe I. Schwerpunkte der Seminararbeit sind neben der Erarbeitung der Stufenmodelle des Schriftspracherwerbs bzw. der mathematischen Kompetenzentwicklung das Kennenlernen von Fördermethoden und die Reflexion der Projekterfahrungen.

Im Schuljahr 2016-17 wurden an drei Partnerschulen erstmals auch Gruppen zur Förderung der Schriftsprache/Alphabetisierung (Lerngruppe 2) für Seiteneinsteiger-innen mit Fluchthintergrund, die mit der arabischen Schrift alphabetisiert wurden, gebildet. Studierende, die sich speziell für diesen Bereich interessieren, werden außerdem in das Gebiet der kontrastiven Alphabetisierung eingeführt. Dieses Teilprojekt wird von Herrn Mehlem, Frau Böhm (Universität Kassel) und Frau Suntheim gemeinsam betreut. Im Bereich der Alphabetisierung steht die Materialentwicklung im Vordergrund.

Die Veranstaltung richtet sich an Studierende aller Lehramter und Erziehungswissenschaftler, die gerne intensive Praxiserfahrungen sammeln möchten. Die Teilnahme an einem der Praxisprojekte setzt voraus, die Förderung in den Projektschulen in der Zeit von Oktober bis Juli verbindlich durchzuführen. Die Studierenden, die sich für das Projekt entscheiden, erhalten für jede geleistete Förderstunde ein Honorar von 10,00 Euro.

5. EW-BA11-BA12/PO2015: Praxisprojekt: Übergang Grundschule-Sekundarstufe I (L5), Dr. Robert Bernhard, Dr. Marian Kratz:

Inhalte: Das Projekt ist eine Fortsetzung des vormaligen FFM-Projektes unter veränderter Schwerpunktsetzung. Es besteht weiter in Kooperation mit der Peter Fuld-Stiftung, der Crespo Foundation, der ranstad stiftung, der Metzler Stiftung, der Stiftung Citoyen, dem Staatlichen Schulamt Frankfurt und der Stadt Frankfurt.

Der Übergang von der Grundschule in die Sekundarstufe I ist für Kinder eine große Herausforderung. Der Verlauf der Trennungen von der Grundschule und den sozialen Kontakten einerseits sowie das Ankommen und Gründen neuer Kontakte in der Sekundarstufe I andererseits können sich nachhaltig auf Bildungsverläufe auswirken. Im Projekt der Übergangsbegleitung stellen sich Studierende als Mentoren zur Verfügung um die Herausforderung des Übergangs zu begleiten.

In Einzelförderung und in Kleingruppen (max. 3 Personen) werden Schülerinnen und Schüler unterschiedlicher Schulformen (Grund-, Haupt-, Real-, und Gesamtschulen) unterstützt. Die Förderung umfasst 4 Wochenstunden und erfolgt in der Regel an zwei Wochentagen zu jeweils zwei Stunden. Die Vergütung beträgt 10 Euro pro Stunde.

Die Studierenden werden durch die Universität Frankfurt ausgewählt, für ihre Aufgabe fachlich vorbereitet und geschult. Sie werden während ihrer Tätigkeit von der Projektleitung in regelmäßigen Seminarveranstaltungen begleitet.

Die Studierenden lernen im Kontext der universitären Begleitung, individuelle Förderpläne zu entwickeln und sich mit Fragen der Erziehung und Bildung kritisch auseinander zu setzen.

Die Teilnahme an den Seminarveranstaltungen ist verbindlich. Es können Modulscheine für die Lehramtsstudiengänge (L1-L5 GWD II oder III, L5 FOE C I und II) sowie im Bachelor Erziehungswissenschaften (BA 11 oder 12) erworben werden.

6. EW-BA11-BA12: Schriftsprachliche und mathematische Diagnostik und -förderung in der Grundschule (auch EW-BA12-BA-13/PO2008/2013, L5-FOE-C:I / L5-FOE-C:II), Prof. Dr. Ulrich Mehlem, Ulrike Suntheim, Constanze Fuchs, Aloysia Abraham:

Inhalte: Das über zwei Semester angelegte Seminar begleitet Projekte im Bereich der Schriftsprach-, und Lese- (Lerngruppen 1 und 3) sowie Rechenförderung (Lerngruppe 4) und wird an Grund- und Sekundarstufenschulen (Klasse 5/6) durchgeführt. Für ein Schuljahr fördern die Studierenden 2mal wöchentlich im gewählten Bereich Kleingruppen von 2-4 Schülern in der Grundschule oder Sekundarstufe I. Schwerpunkte der Seminararbeit sind neben der Erarbeitung der Stufenmodelle des Schriftspracherwerbs bzw. der mathematischen Kompetenzentwicklung das Kennenlernen von Verfahren der individuellen Kompetenzdiagnostik sowie von Fördermethoden und die Reflexion der Projekterfahrungen.

Seit dem Schuljahr 2017-18 werden an fünf Partnerschulen auch Gruppen zur Förderung der Schriftsprache/Alphabetisierung (Lerngruppe 2) für Seiteneinsteiger-innen mit Fluchthintergrund, die mit der arabischen Schrift alphabetisiert wurden, gebildet. Studierende, die sich speziell für diesen Bereich interessieren, werden außerdem in das Gebiet der kontrastiven Alphabetisierung eingeführt. Dieses Teilprojekt wird von Herrn Mehlem und Frau Suntheim gemeinsam betreut.

Die Veranstaltung richtet sich an Studierende aller Lehrämter und Erziehungswissenschaftler, die gerne intensive Praxiserfahrungen sammeln möchten. Die Teilnahme an einem der Praxisprojekte setzt voraus, die Förderung in den Projektschulen in der Zeit von Oktober bis Juli verbindlich durchzuführen. Die Studierenden, die sich für das Projekt entscheiden, erhalten für jede geleistete Förderstunde ein Honorar von 10,00 Euro.

7. TELLUS – miteinander lehren, voneinander lernen (L3, L2, Erziehungswissenschaften, Psychologie, Wirtschaftspädagogik, Soziale Arbeit), Crespo-Foundation, Arbeitsstelle SL:

Inhalte: TELLUS|miteinander lehren – voneinander lernen ist ein Service Learning-Projekt, das sich einerseits an Studierende und andererseits an Schüler*innen richtet: Studierende verschiedener Fachrichtungen unterstützen dabei Schulklassen für Seiteneinsteiger*innen (Geflüchtete, Spätaussiedler*innen und Zugewanderte). Bei der Zielgruppe der Seiteneinsteiger*innen handelt es sich um Jugendliche im Alter von mind. 16 Jahren, die nicht mehr vom Regelschulsystem erfasst werden. Ihnen bleibt nur ein kurzes Zeitfenster, um Deutsch zu lernen und berufliche Perspektiven zu entwickeln. Daher hat diese Zielgruppe einen besonderen Bedarf, der die entsprechenden Lehr- und Fachkräfte an Schulen vor Herausforderungen stellt.

TELLUS dockt an das Sprachförderprogramm InteA sowie weitere Schulformen (z. B. PuSch) an, die für diese Zielgruppe an Beruflichen Schulen angeboten werden. Durch eine intensive Sprachförderung, allgemeinbildenden Unterricht sowie Angebote zur beruflichen Orientierung soll die Schülerschaft in die Lage versetzt werden, auf dem deutschen Arbeits- und Ausbildungsmarkt Fuß zu fassen. Diese Aufgabe ist angesichts der heterogenen Zusammensetzung der Gruppen und sehr unterschiedlicher Bildungsvoraussetzungen nicht leicht.

Daher hat die Crespo Foundation gemeinsam mit der Goethe-Universität Frankfurt und der Frankfurt University of Applied Sciences das Service Learning-Projekt TELLUS ins Leben gerufen, um Lehrkräfte und Fachkräfte der sozialpädagogischen Förderung in diesen Klassen zu unterstützen.

Studierende der Fachbereiche Lehramt, Wirtschaftspädagogik, Soziale Arbeit, Erziehungswissenschaften und Psychologie unterstützen die Schüler*innen in multiprofessionellen Teams und arbeiten eng mit den Lehr- und Fachkräften zusammen. Für ein Jahr engagieren sie sich rund elf Stunden pro Woche. Sie helfen im Unterricht, unterstützen und begleiten die Jugendlichen aber auch abseits der Schule, etwa beim Schreiben von Bewerbungen oder bei gemeinsamen Aktivitäten. Damit helfen sie den Jugendlichen, sich mit dem deutschen Schul- und Berufsbildungssystem auseinander zu setzen, eigene Perspektiven zu entwickeln und erste Kontakte herzustellen.

Die Studierenden werden in Schulungen auf den Umgang mit Menschen mit Fluchterfahrung und die damit verbundenen Anforderungen vorbereitet. Eine professionelle Supervision begleitet sie bei ihrer Arbeit. Zudem werden ihre Erfahrungen in Begleitseminaren an den jeweiligen Hochschulen besprochen. Auf diese Weise können sie die jeweils verlangten Pflichtpraktika absolvieren. Sie sammeln zudem didaktische und sozialpädagogische Erfahrungen für ihr späteres Berufsleben. So wird das Projekt zu einem Gewinn für alle Beteiligten: die Schüler*innen, die Studierenden und die Lehr- und Fachkräfte.

Weiterführende

Links

<http://www.crespo-foundation.de/tellus>

[https://kultusministerium.hessen.de/schule/individuelle-](https://kultusministerium.hessen.de/schule/individuelle-foerderung/sprachfoerderung/intensivklassen-beruflichen-schulen-intea)

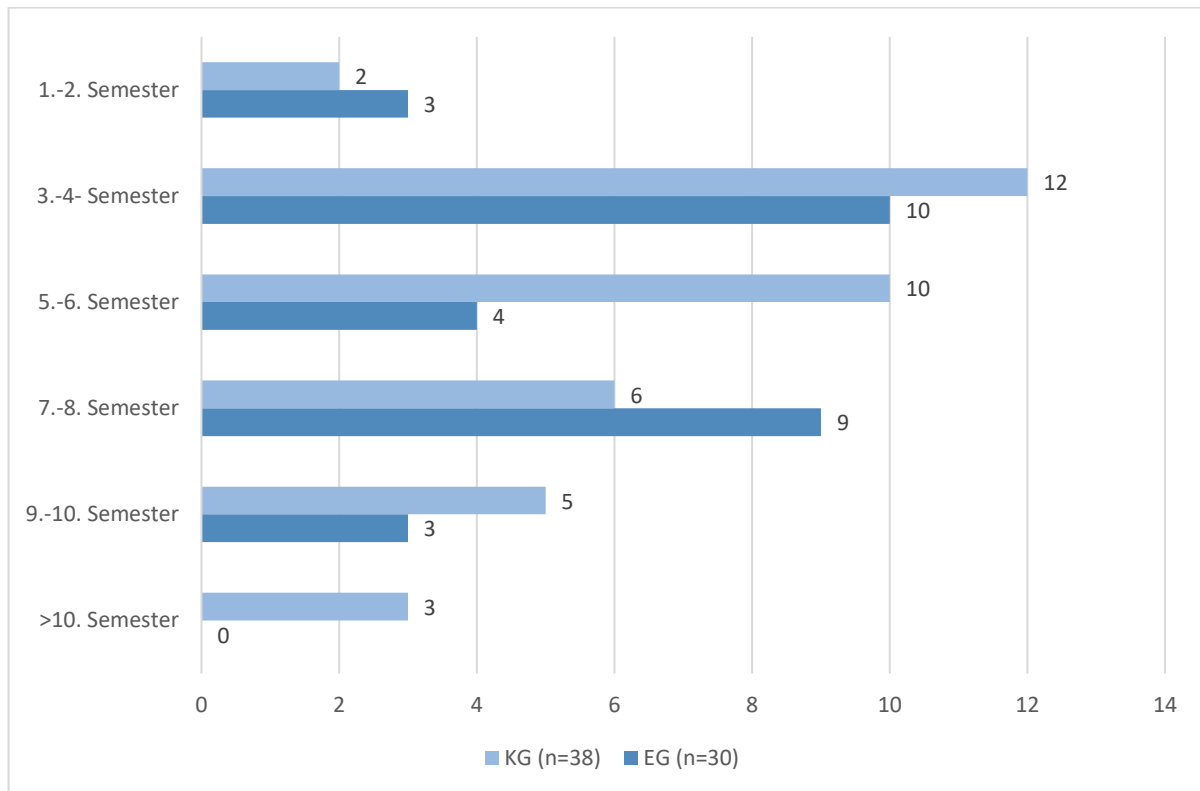
[foerderung/sprachfoerderung/intensivklassen-beruflichen-schulen-intea](https://kultusministerium.hessen.de/schule/individuelle-foerderung/sprachfoerderung/intensivklassen-beruflichen-schulen-intea)

<https://kultusministerium.hessen.de/foerderangebote/pusch-efre/praxis-und-schule-pusch>

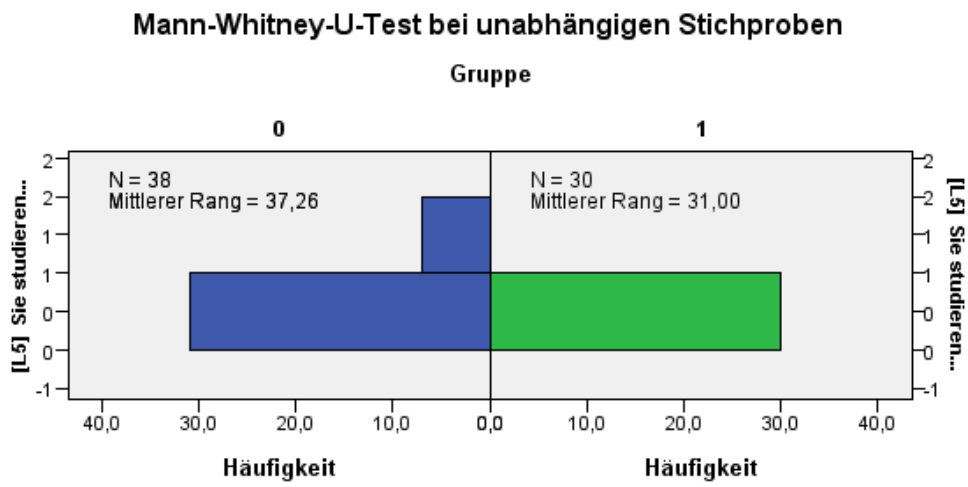
TELLUS vergibt jährlich Praktikumsplätze an Studierende. Die Bewerbungsprozesse finden im Frühjahr statt, Praktikumsbeginn ist im Spätsommer und richtet sich nach den Sommerferien in Hessen. Informationen zur Bewerbung um eine Praktikumsstelle in TELLUS finden Sie [hier](#).

8. Stadt als Lebensraum (Fachdidaktik Humangeographie; L2, L3, L5), Uwe Schulze, Oliver Wolff, Prof. Dr. Kanwischer:

Inhalte: Zielstellung der Veranstaltung ist es, bei angehenden Lehrer/-innen das Verständnis über die sozialräumliche Wahrnehmung, Konstruktion und Aneignung des Stadtraumes bei Kindern und Jugendlichen im Kontext der Nutzung digitaler Geomedien zu fördern. Die zentrale Fragestellung lautet, in wie fern digitale Geomedien Einfluss auf lebensweltlich Bereiche von Schüler/-innen haben und welche Kompetenzen den Umgang mit digitalen Geomedien positiv fördern. Neben technischen Aspekten bezieht sich dies vor allem auf Fähigkeitsbereiche wie Kommunikation, Reflektion und das Denken in unterschiedlichen räumlichen Bezügen. In der ersten Seminarphase im WS17/18 werden die theoretische Grundlagen räumlicher Sozialisationsforschung sowie notwendige soziotechnische und mediendidaktische Inhalte im Umfeld der Geoinformationsverarbeitung erarbeitet. Darau aufbauend, konzentriert sich die zweite Phase des Moduls im SoSe2018 auf die praktische Arbeit an und die Entwicklung von Unterrichtssequenzen, welche die Erkenntnisse aus der ersten Modulphase integrieren. In der dritten Modulphase „vor Ort“ im SoSe2018 sollen die entwickelten Lehr-Lernmaterialien durch die Teilnehmer/-innen auf deren praktische Umsetzung im Unterrichtsgeschehen hin überprüft und abschließend reflektiert werden.

Anhang D2: Fachsemesterzahl der teilnehmenden Studierenden

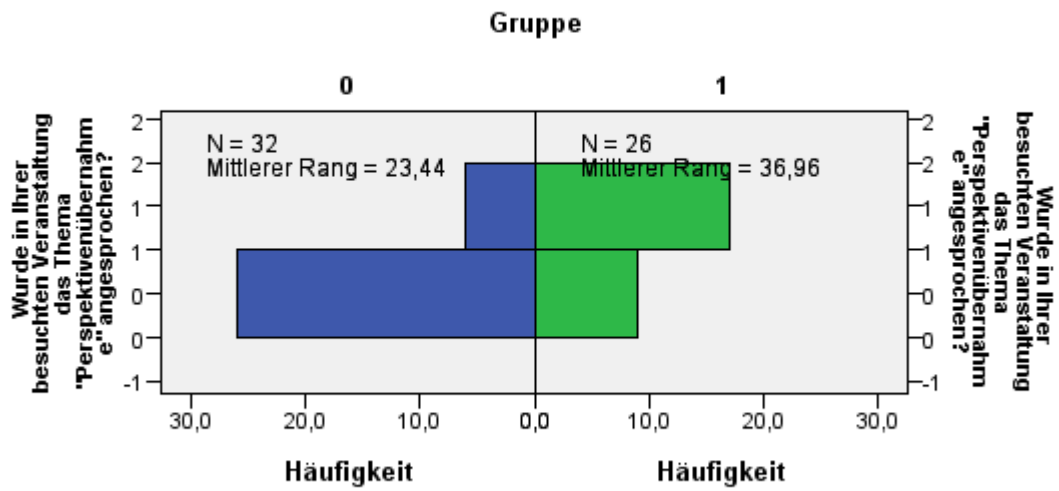
Anhang D3: Stichprobenanalysen



Gesamtanzahl	68
Mann-Whitney-U	465,000
Wilcoxon-W	930,000
Teststatistik	465,000
Standardfehler	42,619
Standardisierte Teststatistik	-2,464
Asymptotische Sig. (2-seitiger Test)	,014

Gruppenunterschiede bzgl. der Variablen „Wurde in Ihrer besuchten Veranstaltung das Thema "Perspektivenübernahme" angesprochen?“:

Mann-Whitney-U-Test bei unabhängigen Stichproben



Gesamtanzahl	58
Mann-Whitney-U	610,000
Wilcoxon-W	961,000
Teststatistik	610,000
Standardfehler	54,199
Standardisierte Teststatistik	3,579
Asymptotische Sig. (2-seitiger Test)	,000

Anhang D4: Instruktion vor der Testung

Liebe Versuchsteilnehmerin, lieber Versuchsteilnehmer,

vielen Dank, dass Sie an der folgenden Studie im Rahmen meiner Dissertation zum Thema "Förderung von Empathie durch Service-Learning-Veranstaltungen an Beispielen aus der Lehrerbildung" teilnehmen.

Dies ist der erste von zwei Messzeitpunkten, an denen Sie teilnehmen können. Können deswegen, weil Ihre Teilnahme freiwillig ist und Sie die Testung jederzeit beenden können.

Im Folgenden werden Ihnen einige Aussagen und insgesamt sechs kurze Videosituationen mit Fragen zur Bewertung und Einschätzung gezeigt. Die Videosequenzen sind Ausschnitte aus groß angelegten Unterrichtsstudien, weswegen die SchülerInnen teilweise Schilder mit Zahlen um den Hals haben (keine Bewertung, sondern Hilfe zur Auswertung).

Die Lautstärke der Vignetten wurde zwar angepasst, kann aber wegen der Aufnahmesituation trotzdem variieren, deswegen die Lautstärke individuell justieren → an Kopfhörer seitlich

Die Bearbeitung dauert etwa 45 Minuten und Sie können die Videosequenzen gerne mehrmals anschauen.

Um Empathie valide messen zu können, sind Ihre Reaktionen auf die Videosequenzen ein entscheidendes Kriterium. Aus diesem Grund werden während des Vignettentests Aufnahmen von Ihrem Gesicht gemacht, um diese mit einer speziellen Software auszuwerten.

Deswegen steht auch die Lampe vor Ihnen und wird von Anfang an angemacht, damit Sie sich etwas daran gewöhnen können. Die Aufnahmen werden nur während des Vignettentest-Teils gemacht und Sie bekommen davor und danach jeweils eine kurze Information, dass die Aufnahme begonnen hat bzw. beendet ist.

Die Aufnahmen werden getrennt von Ihren Fragebogendaten gespeichert und mit Ihrem Code versehen, um eine Zuordnung zu ermöglichen. Mit den Fragebogendaten werden nur die Ergebnisse der Videoauswertung direkt verknüpft.

Während des Vignettentests lassen Sie das Tablet bitte im Ständer stehen (optimaler Aufnahmeradius) und versuchen Sie währenddessen die Hand nicht ins Gesicht zu nehmen, nicht zu essen/trinken und kein Kaugummi zu kauen. Davor und danach können Sie das handhaben wie es Ihnen am besten passt.

Bei der Frage nach dem empathischsten LehrerInnenverhalten ziehen Sie die Antworten am besten über oben auf die andere Seite, weil sonst unter Umständen die Reihenfolge vertauscht wird (technisches Problem mit der App).

Bei weiteren Fragen oder wenn Sie Interesse an den Ergebnissen haben, setzen Sie sich gerne mit mir in Verbindung.

Anhang D5: Kovariatenanalyse und -auswahl

Shapiro-Wilk-Test auf Normalverteilung der Daten:

	Gruppe	Shapiro-Wilk		
		Statistik	df	Signifikanz
Semester_Cluster	0	,903	31	,009
	1	,883	22	,014
Ihr Alter (in Jahren)	0	,734	31	,000
	1	,746	22	,000
In welchem Fachsemester studieren Sie zur Zeit?	0	,900	31	,007
	1	,901	22	,031
Haben Sie sich in der Vergangenheit ehrenamtlich engagiert?	0	,607	31	,000
	1	,412	22	,000
Geschlecht	0	,519	31	,000
	1	,590	22	,000
Empathie/Perspektivenübernahme besprochen	0	,638	31	,000
	1	,474	22	,000
Praxisanteil in SL-Seminar	0	.	31	.
	1	,624	22	,000
Haben Sie im SoSe18 ein studienbezogenes Praktikum absolviert?	0	,591	31	,000
	1	,522	22	,000
Haben Sie Kinder?	0	,176	31	,000
	1	,332	22	,000

0 = KG; 1 = EG

Korrelationsmatrix, Kendals-Tau-B:

EG/KG

	Fachsemester	Ihr Alter (in Jahren)	Ehrenamtliches Engagement in Vergangenheit	Geschlecht	SoSe18 studienbezogenes Praktikum	Eigene Kinder	Naturwissenschaftliche Fächer	Sprachlich-literarisches Aufgabefeld	Gesellschaftswissenschaftliches Aufgabefeld	Sonstiges
ESEKfu	Korrelationskoeffizient	-,254*								
	Sig. (2-seitig)	0,039								
	N	38								
BESEKfunk	Korrelationskoeffizient		-,400*							

Anhang D6: Übersicht des Dropout der Analyseergebnisse von Face-Reader

Version Prä	Fehlende Videos	Version Inter	Fehlende Videos	Version Post	Fehlende Videos
Vignette 2	2	Vignette 6	4	Vignette 2	7
Vignette 6	2	Vignette 8	4	Vignette 6	6
Vignette 9	3	Vignette 10	4	Vignette 8	6
Vignette 21	3	Vignette 14	3	Vignette 21	6
Vignette 25	3	Vignette 27	3	Vignette 25	5
Vignette 30	3	Vignette 30	2	Vignette 30	5

Anhang D7: Übersicht: Prüfung auf Normalverteilung

	Gruppe	Statistik	Df	Signifikanz
ESEKfunk	0	.935	37	.032
(Inter)	1	.925	30	.037
ScoreTEQ	0	.929	37	.021
(Inter)				
Gesamt IScore	0	.926	37	.017
(Prä)	1	.894	30	.006
Gesamt IScore	0	.922	37	.013
(Post)	1	.903	30	.010
EKF EE	0	.931	37	.024
(Inter)				
ISK PS (Prä)	1	.919	30	.026
ISK ZU (Inter)	1	.887	30	.004
ISK KO (Inter)	0	.934	37	.029
ISK ZU (Post)	0	.912	37	.006
	1	.894	30	.006
ISK WP (Post)	0	.940	37	.045

Anmerkungen: Gruppe 0=KG, Gruppe 1=EG

Danksagungen

Am Ende dieser Arbeit möchte ich noch gerne meinen Dank zum Ausdruck bringen.

Zuerst einmal möchte ich gerne Professor Dr. Holger Horz, meinem Doktorvater, danken. Er hat das Thema Service Learning an die Goethe-Universität gebracht und damit den Grundstein für mein Promotionsvorhaben gelegt. Außerdem hat er mir freie Hand bei der Ausarbeitung der Fragestellung und Ausgestaltung des Forschungsvorhabens gelassen, so dass die vorliegende Arbeit entstehen konnte. Zusätzlich stattete er mich mit allen Materialien, finanziellen Mitteln und Unterstützung aus, die die Umsetzung erheblich erleichtert und erst möglich gemacht haben – vielen Dank, lieber Holger, für die Unterstützung!

Ebenfalls einen großen Dank an Dr. Mark Ullrich, der mir mit Beratung und Supervision regelmäßig, kurzfristig und auch während subjektiv dramatischer Phasen zur Seite stand und mit Ideen und Lösungsvorschlägen aufwartete, so dass es mit mehr Klarheit weitergehen konnte – danke, lieber Mark!

Ein großer Dank auch an meine (ehemalige) Kollegin und Freundin Dr. Katja Knuth-Herzig, die sich das Gejammerge, aber auch sämtliche Ideen mit viel Geduld angehört hat und mit Rat und Tat für mehr Klarheit sowie mit ihrer Art für mehr Gelassenheit gesorgt hat. Danke, Wifey!!!

Ein riesiger Dank geht auch an André Kerkhoff, ohne dessen Know-How die technische Umsetzung nicht möglich gewesen wäre. Wir haben herumprobiert, geschraubt, gelötet, verschiedene Glühlampen durchprobiert, viel gelacht und er hatte viel Geduld, wenn ich immer mit neuen „Baustellen“ hereinschneite. Und er machte mir mit vielen Skripten die Arbeit sehr viel leichter. Danke, lieber André!

Eine weitere wichtige Person im Gelingen dieser Arbeit ist Victoria Schönefeld, mein Partnerin-Crime. Kennengelernt haben wir uns im Rahmen der Empathieexpertenbefragung, nach der wir regelmäßig in Kontakt blieben und uns über Empathie austauschten. Im letzten September beschlossen wir dann ein Dissertations-Buddy-Programm, damit wir beide mit diesem Vorhaben schnell und zügig fertig werden könnten. Wir schickten uns Zeitpläne, mussten Rede und Antwort ob der Fortschritte stehen und motivierten uns in regelmäßigen Telefonaten und Nachrichten. Ohne Dich, liebe Vic, weiß ich nicht, ob ich das geschafft hätte. Danke für Dein Cheerleaden, Deinen fachlichen Rat, für den Austausch, Dein offenes Ohr, die Aufmunterung und Motivation und dass wir es tatsächlich gerockt haben!!!! Es ist vollbracht.

Ein weiteres Danke geht an meine Familie und meine Freunde. Alle mussten sich in den letzten Jahren viel Gejammer und Genöhle anhören und haben sich alles mit viel Gelassenheit und Geduld angehört. Danke für euren Zuspruch und sorry für die schlechte Laune in der letzten Phase! Jetzt wird alles besser :-P

Danke auch speziell an meinen Freund Andreas Mayer. Danke, Andy, für Deine unschätzbare Hilfe beim letzten Feinschliff beim Kampf mit Word und diversen Formatierungshürden. Ich glaube ohne Dich wäre die Publikation niemals möglich gewesen.

Ich möchte mich auch bei meiner ehemaligen Abteilung, den PsyLLiEs, bedanken. Danke für den Austausch und den Rat, den ich von allen über die Jahre in Gesprächen und nach Vorträgen immer wieder bekommen habe und die mir an vielen Stellen weitergeholfen haben. Danke euch!

Danke auch für die Unterstützung aller involvierter Dozierenden, die mich in ihren Veranstaltungen rekrutieren ließen und mit dem Anbieten von SL-Veranstaltungen solch einen wunderbaren Beitrag leisten. Bitte machen Sie weiter und vielen herzlichen Dank für die Unterstützung!

Und last, but not least, einen herzlichen Dank an die fleißigen Korrekturleser, die mir mit Feedback in der finalen Phase auch noch einmal sehr geholfen haben: Dr. Katja Knuth-Herzig, Charlotte Lehmann, Dr. Immanuel Ulrich, Sabine Jakubczyk und Edeltraud Kuchta-Korb.

Ich befürchte, dass ich bestimmt jemanden vergessen habe, was aber keiner bösen Absicht, sondern nur der Tatsache geschuldet ist, dass dieses Dissertationsprojekt eine so lange Zeit in Anspruch genommen hat, mein Gehirn am Ende dieses Schreibprozesses völlig erledigt ist und soooo viele Menschen zum Gelingen beigetragen haben. Danke an alle, die sich beteiligt haben, ein offenes Ohr hatten, Expertise einfließen ließen und mich motivierten weiterzumachen, um dieses Werk nun tatsächlich in den Händen zu halten. Danke euch allen!