

Aus der Klinik für Psychiatrie und Psychotherapie
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Erstellung und Erprobung zweier Parallelversionen des
Multidimensionalen Empathietests (MET) zur longitudinalen
Erfassung des Empathievermögens

zur Erlangung des akademischen Grades
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Karin Diane Müller

aus Schwäbisch Gmünd

Datum der Promotion: 04.06.2021

Inhaltsverzeichnis

ABKÜRZUNGSVERZEICHNIS	V
ABBILDUNGSVERZEICHNIS	VII
TABELLENVERZEICHNIS	VIII
1. ABSTRACT	1
1.1. Deutsche Version	1
1.2. Englische Version	3
2. EINLEITUNG	4
2.1. Was ist Empathie?	4
2.1.1. Das multidimensionale Konzept der Empathie	4
2.1.2. Neurowissenschaftliche und evolutionäre Grundlagen der Empathie	5
2.1.3. Notwendigkeit der longitudinalen Erfassung des Empathievermögens	7
2.2. Testverfahren zur Messung des Empathievermögens	9
2.2.1. Reading the Mind in the Eyes Test (RMET)	9
2.2.2. Movie for the Assessment of Social Cognition (MASC).....	10
2.2.3. Baron-Cohen's Empathy Quotient (EQ)	10
2.2.4. Interpersonal Reactivity Index (IRI)	11
2.2.5. Multidimensionaler Empathietest (MET).....	12
2.3. Erstellung von Parallelversionen	16
2.3.1. Notwendigkeit der Erstellung zweier Parallelformen des MET	16
2.3.2. Verfahren zur Erstellung von Parallelformen eines Tests.....	17
2.4. Zielsetzung	19
3. METHODIK	20
3.1. Erstellung der Parallelversionen des Multidimensionalen Empathietests ..	20
3.1.1. Erstellung von Itempaaren	20

3.1.2. Erstellung der Frageblöcke.....	21
3.1.4. Programmierung der Parallelversionen	23
3.2. Erprobung der Parallelversionen.....	24
3.2.1. Studienteilnehmer	25
3.2.2. Ablauf der Testung	25
3.2.3. Klinische und Neuropsychologische Diagnostik	27
3.2.3.1. Mini-International Neuropsychiatric Interview (M.I.N.I.)	27
3.2.3.2. Hamilton Depression Rating Scale (HDRS-24)	28
3.2.3.3. Psychotherapie Basisdokumentation (Psy-BaDo)	28
3.2.3.4. Beck-Depressions-Inventar (BDI-II).....	28
3.2.3.5. Childhood Trauma Questionnaire (CTQ).....	29
3.2.3.6. Toronto Alexithymie Skala (TAS-20)	30
3.2.3.7. Zahlennachsprechen.....	30
3.2.3.8. Regensburger Wortflüssigkeitstest.....	31
3.2.3.9. Trail-Making-Test (TMT).....	31
3.2.3.10. Mehrfach-Wortschatz-Intelligenztest (MWT-B).....	32
3.2.3.11. Verbaler Lern- und Merkfähigkeitstest (VLMT).....	32
3.3. Datenauswertung	33
3.4. Anpassung der Items nach erster Erprobung	34
4. ERGEBNISSE	35
4.1. Ergebnisse der ersten Stichprobe	35
4.1.1. Erstellung der Parallelversionen.....	35
4.1.2. Studienteilnehmer	38
4.1.2.1. Demographische Daten.....	38
4.1.2.2. Klinische und neuropsychologische Daten	40
4.1.3. Statistische Auswertung.....	42

4.1.3.1. Korrelation	42
4.1.3.2. Vergleich der Mittelwerte	42
4.1.3.3. Schwierigkeit	44
4.1.3.4. Trennschärfe	45
4.1.3.5. Reliabilität.....	49
4.2. Ergebnisse der zweiten Stichprobe	50
4.2.1. Erstellung der Parallelversionen.....	50
4.2.2. Studienteilnehmer	52
4.2.2.1. Demographische Daten.....	52
4.2.2.2. Klinische und neuropsychologische Daten	53
4.2.3. Statistische Auswertung.....	56
4.2.3.1. Korrelation	56
4.2.3.2. Vergleich der Mittelwerte	56
4.2.3.3. Schwierigkeit	58
4.2.3.4. Trennschärfe	59
4.2.3.5. Reliabilität.....	63
5. DISKUSSION	64
5.1. Korrelation zwischen den Parallelversionen	64
5.2. Vergleich der Mittelwerte	64
5.3. Itemschwierigkeit	65
5.4. Trennschärfe.....	66
5.5. Interne Konsistenz	67
5.6. Methodenkritische Überlegungen.....	68
5.6.1. Erstellung der Parallelversionen.....	68
5.6.2. Studienpopulation.....	69
5.6.3. Erprobung der Parallelversionen.....	70

5.7. Vergleich der initialen und finalen Parallelversionen	72
5.8. Fazit und Ausblick.....	74
6. LITERATURVERZEICHNIS	76
7. ANHANG.....	84
8. EIDESSTATTLICHE VERSICHERUNG.....	103
9. LEBENSLAUF	104
10. DANKSAGUNG.....	106

Abkürzungsverzeichnis

ASD	Autismus-Spektrum-Störung
BDI-II	Beck-Depressions-Inventar II
CBASP	Cognitive Behavioral Analysis System of Psychotherapy
CTQ	Childhood Trauma Questionnaire
DAS	Dysfunctional Attitude Scale
DAS-40A/B	40-Item Parallelversion A/B der Dysfunktional Attitude Scale
DAS-18A/B	18-Item Parallelversion A/B der Dysfunktional Attitude Scale
EQ	Empathy Quotient
fMRT	Funktionelle Magnetresonanztomographie
HAWIE	Hamburg-Wechsler-Intelligenztest für Erwachsene
HDRS-24	24-Item Hamilton Depression Rating Scale
IRI	Interpersonal Reactivity Index
MASC	Movie for the Assessment of Social Cognition
MET	Multidimensionaler Empathietest
MET-A/B	Multidimensionaler Empathietest Parallelversion A/B
MET-A*/B*	Multidimensionaler Empathietest Parallelversion A*/B*
MET-J	Multidimensionaler Empathietest - Jugendliche
MET-KE	Multidimensionaler Empathietest – Kinder/Erwachsene
M.I.N.I	Mini-International Neuropsychiatric Interview
MW	Mittelwert
MWT	Mehrfach-Wortschatz-Intelligenztest
MWT-A	Mehrfach-Wortschatz-Intelligenztest Parallelversion A
MWT-B	Mehrfach-Wortschatz-Intelligenztest Parallelversion B
PsyBado	Psychotherapie Basisdokumentation

ABKÜRZUNGSVERZEICHNIS

RMET	Reading the Mind in the Eyes Test
SD	Standardabweichung
TAS	Toronto Alexithymie Skala
VLMT	Verbaler Lern- und Merkfähigkeitstest

Abbildungsverzeichnis

Abbildung 1: Modell der Bottom-Up und Top-Down Prozesse der menschlichen Empathie nach Decety & Lamm (2016).	5
Abbildung 2: Beispiel für ein Augenpaar aus dem RMET (Baron-Cohen et al., 2001)...	9
Abbildung 3: Beispielitem der kognitiven Empathie im MET.....	13
Abbildung 4: Beispielitem der emotionalen Empathie im MET.	14
Abbildung 5: Abfolge der Frageblöcke in den Parallelversionen A und B.....	22
Abbildung 6: Beispiel für die Messung der kognitiven Empathie aus MET-B.....	23
Abbildung 7: Beispiel für die Messung der emotionalen Empathie aus MET-B.	24
Abbildung 8: Ablauf der Testung an Tag 1 und Tag 2.	26
Abbildung 9: Ablauf des Verbalen Lern- und Merkfähigkeitstest (VLMT).....	32
Abbildung 10: Vergleich der Mittelwerte in Stichprobe 1.....	43
Abbildung 11: Häufigkeitsverteilung der Itemschwierigkeiten in Stichprobe 1.	45
Abbildung 12: Häufigkeitsverteilung der Trennschärfekoeffizienten in Stichprobe 1. ..	46
Abbildung 13: Vergleich der Mittelwerte in Stichprobe 2.....	57
Abbildung 14: Häufigkeitsverteilung der Itemschwierigkeiten in Stichprobe 2.	59
Abbildung 15: Häufigkeitsverteilung der Trennschärfekoeffizienten in Stichprobe 2. ..	60

Tabellenverzeichnis

Tabelle 1: Übersicht der Testverfahren zur Messung von Empathie.....	16
Tabelle 2: Itemblöcke von MET-A und MET-B.....	36
Tabelle 3: Übersicht von Schwierigkeit und Trennschärfe von MET-A und MET-B.	37
Tabelle 4: Itemmerkmale von MET-A und MET-B.....	38
Tabelle 5: Demographische Daten von Stichprobe 1.....	39
Tabelle 6: Übersicht klinische und neuropsychologische Diagnostik in Stichprobe 1. ..	41
Tabelle 7: Vergleich der Mittelwerte in Stichprobe 1.....	44
Tabelle 8A: Itemkennwerte von MET-A.	47
Tabelle 8B: Itemkennwerte von MET-B.	48
Tabelle 9: Mittelwerte der Schwierigkeiten und Trennschärfen in Stichprobe 1.....	50
Tabelle 10: Itemblöcke von MET-A* und MET-B*.....	51
Tabelle 11: Übersicht von Schwierigkeit und Trennschärfe von MET-A* und MET-B*..	52
Tabelle 12: Demographische Daten von Stichprobe 2.....	53
Tabelle 13: Übersicht klinische und neuropsychologische Diagnostik in Stichprobe 2.	55
Tabelle 14: Vergleich der Mittelwerte in Stichprobe 2.....	58
Tabelle 15A: Itemkennwerte von MET-A*.....	61
Tabelle 15B: Itemkennwerte von MET-B*.....	62

1. Abstract

1.1. Deutsche Version

Hintergrund: Empathie ist ein multidimensionales Konstrukt bestehend aus einer kognitiven und einer emotionalen Komponente. Ein Empathiedefizit wurde für einige psychiatrische Krankheitsbilder bereits nachgewiesen. Es gibt Hinweise darauf, dass die Empathiefähigkeit auch bei affektiven Störungen vermindert ist, wobei aktuell noch keine konsistenten Daten existieren. Es wäre zudem notwendig, das Empathievermögen im longitudinalen Verlauf beurteilen zu können, beispielsweise in verschiedenen Stadien der Depression oder nach einer spezifischen Intervention, um auf diese Weise die Ursachen und aufrechterhaltenden Faktoren der Erkrankung besser beurteilen und therapeutisch nutzen zu können. Ein geeignetes Messinstrument fehlt jedoch bislang. Der Multidimensionale Empathietest (MET) ist ein etabliertes Testverfahren zur Messung von Empathie, welches die kognitive und die emotionale Empathie getrennt und dennoch zeitgleich erfasst. Das Ziel dieser Arbeit war es demnach, zwei Parallelversionen des MET zu erstellen und auf ihre Vergleichbarkeit zu prüfen.

Methodik: Anhand von Itemschwierigkeit und Trennschärfe bereits bestehender Datensätze wurden zwei Parallelversionen MET-A und MET-B erstellt und an 20 gesunden Probanden erprobt. Anschließend wurden die Versionen anhand von Korrelationen und Mittelwertvergleichen auf ihre Parallelität überprüft. Nach einer notwendigen Anpassung der Tests zu den Parallelversionen MET-A* und MET-B* erfolgte die erneute Erprobung an 15 gesunden Probanden sowie die konsekutive Überprüfung auf Parallelität.

Ergebnisse: In den zunächst erstellten Versionen MET-A und MET-B gab es keine signifikante Korrelation zwischen den Summen der richtigen Antworten (kognitive Empathie). Vielmehr zeigten sich signifikante Mittelwertsunterschiede. Nach Anpassung der Parallelversionen korrelierten MET-A* und MET-B* sowohl im Hinblick auf die Ergebnisse der kognitiven als auch der emotionalen Empathie sowie der Antwortlatenzen im mittleren bis hohen Bereich. Zwischen den berechneten Mittelwerten gab es keine statistisch signifikanten Unterschiede. Jedoch zeigte sich in der hier verwendeten Stichprobe einzig für die Version MET-A* eine zufriedenstellende Reliabilität von $\alpha = 0,76$.

Schlussfolgerung: Anhand der bisherigen Ergebnisse kann von einer Parallelität der Versionen MET-A* und MET-B* ausgegangen werden. Die Erprobung an einer größeren

ABSTRACT

Stichprobe wird empfohlen, um die Ergebnisse im Hinblick auf die Parallelität zu bestätigen, sowie die Reliabilität erneut zu überprüfen.

1.2. Englische Version

Background: Empathy is a multidimensional construct that contains both cognitive and affective components. Impaired empathic abilities have already been demonstrated for some psychiatric disorders. There is evidence to suggest that empathic abilities are impaired in affective disorders as well. However, consistent data is lacking. Furthermore, it would be necessary to assess empathy in the longitudinal course, for example in different stages of depression or after an intervention, in order to evaluate and therapeutically utilize possible causes and sustaining factors. So far there is no suitable measure for this. The Multifaceted Empathy Test (MET) is a well-established measure that assesses cognitive and affective empathy separately but simultaneously. Thus, the aim of this work was to develop two parallel versions of the MET and to examine their comparability.

Methods: Based on item difficulty and item-total-correlation of existing data sets, two parallel versions MET-A and MET-B were developed and tested in 20 healthy subjects. The comparability of the versions was examined using correlations and comparison of means. Subsequently, the tests had to be adjusted to the parallel versions MET-A* and MET-B*, that were re-tested in 15 healthy subjects and their comparability consecutively examined.

Results: In the initially developed versions MET-A and MET-B, there was no significant correlation between the sums of the correct answers (cognitive empathy). On the contrary, the corresponding mean values differed significantly. After adjusting the parallel versions, MET-A* and MET-B* correlated moderately to highly with regard to cognitive and affective empathy as well as the response time. There were no statistically significant differences between the calculated mean values. However, in this sample only version MET-A* showed a satisfactory reliability of $\alpha = 0,76$.

Conclusions: Based on the existing results, MET-A* and MET-B* can be assumed to be parallel. Re-testing on a larger sample of subjects is recommended in order to confirm the results with regard to parallelism and to re-examine reliability.

2. Einleitung

2.1. Was ist Empathie?

2.1.1. Das multidimensionale Konzept der Empathie

Empathie bezeichnet die Fähigkeit Gedanken, Empfindungen und Emotionen einer anderen Person zu verstehen und nachzuempfinden (Altmann, 2014) und ist ein multidimensionales Konstrukt, welches aus einer kognitiven und einer emotionalen bzw. affektiven Komponente besteht (Decety & Lamm, 2006; Dziobek et al., 2008). Unter kognitiver Empathie versteht man hierbei die Fähigkeit, die Perspektive einer anderen Person einzunehmen und daraus Rückschlüsse auf ihre Gefühle, Gedanken und Absichten zu ziehen („Theory of Mind“, „Perspective Taking“) (ebd.). Emotionale oder affektive Empathie beschreibt dagegen die Fähigkeit den emotionalen Zustand einer anderen Person „mitzufühlen“ („Empathic Concern“) (ebd.).

Decety & Lamm beschreiben darüber hinaus in ihrem Empathiemodell die wechselseitige Beziehung beider Komponenten über das Vorliegen eines automatisch, unbewusst ablaufenden „Bottom-Up“ Prozesses (affektives Mitfühlen/Resonanz), sowie eines bewussten, kognitiven „Top-Down“ Prozesses (kognitive Empathie, „Theory of Mind“), welcher auch für die Emotionsregulation und -kontrolle von Bedeutung ist (Decety & Lamm, 2006; siehe **Abbildung 1**).

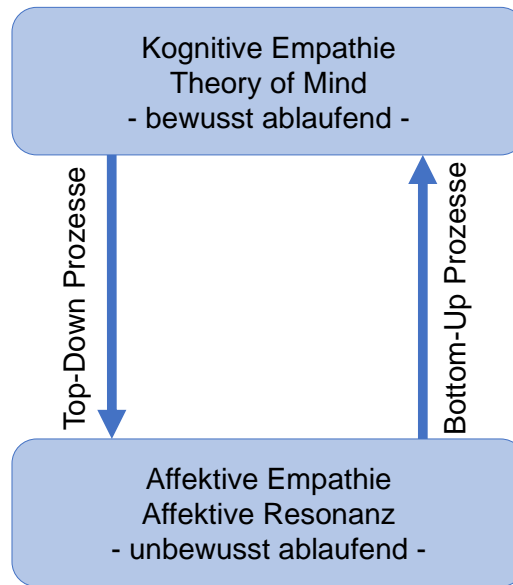


Abbildung 1: Modell der Bottom-Up und Top-Down Prozesse der menschlichen Empathie nach Decety & Lamm (2006).

Kognitive „Top-Down“ Prozesse sind im Rahmen der Emotionsregulation in der Lage, Emotionen zu verstärken oder zu vermindern (ebd.) Zudem hat sich gezeigt, dass Menschen, die in der Lage sind ihre Emotionen zu regulieren, ein höheres Maß an Empathie empfinden (Eisenberg et al., 1994).

2.1.2. Neurowissenschaftliche und evolutionäre Grundlagen der Empathie

Psychologische Konstrukte, wie beispielsweise das der Empathie, können nicht zwangsläufig einem bestimmten neuronalen Prozess zugeordnet werden. Daher ist es für das Verständnis der biologischen Prozesse notwendig, diese auf mehreren Ebenen zu betrachten (Decety & Lamm, 2006). Der Mensch ist das einzige Lebewesen, das in der Lage ist, willentlich mit anderen Menschen mitzufühlen, deren Erfahrungen wesentlich von den eigenen abweichen, und dann nach diesem Gefühl zu handeln (Batson et al., 1991; Decety & Lamm, 2006). Man kann daher annehmen, dass sich Empathie wie der Mensch im Laufe der Evolution entwickeln musste. Der erste Schritt, welcher in der kindlichen Entwicklung Empathie vorausgeht, ist die unbewusste Nachahmung von Gesichtsausdruck, Haltung, Bewegungsmustern, Lauten und allgemeinen Verhaltensweisen einer

anderen Person, um sich auf diese Weise emotional anzunähern (Hatfield et al., 1993). Diese unbewusste Nachahmung wird auch als „Chamäleon-Effekt“ bezeichnet (Chartrand & Bargh, 1999). Der „Chamäleon-Effekt“ ist ausgeprägter bei Menschen, welche eine höhere Empathiefähigkeit aufweisen (ebd.). Unbewusste Nachahmung steigert das Gefühl von Zugehörigkeit und fördert damit den Beziehungsaufbau zu anderen Menschen (Lakin et al., 2003). In einer Studie konnte mittels funktioneller Magnetresonanztomographie (fMRT) gezeigt werden, dass Probanden, die Gesichtsausdrücke verschiedener Emotionen betrachten oder nachahmen sollten, eine erhöhte neuronale Aktivität im Sulcus temporalis superior, der vorderen Inselrinde, der Amygdala sowie in Gebieten des prämotorischen Kortex aufwiesen (Carr et al., 2003). Solch eine neuronale Reaktion auf die Emotionen anderer Menschen könnte evolutionsbiologisch betrachtet Vorteile für das Überleben des Einzelnen haben. So zeigte eine andere fMRT Studie, dass beim Beobachten ängstlicher Körperhaltungen sowohl die Aktivität in Hirnregionen für emotionale Prozesse als auch in Hirnregionen für Aktivität und Bewegung anstieg (De Gelder et al., 2004). Diese wiederum bereiten das Gehirn auf eine Handlung, beispielsweise Flucht, vor (ebd.).

Im Gegensatz zur beschriebenen Nachahmung von Gefühlsausdrücken setzt die Fähigkeit sich in andere Menschen hineinzusetzen bzw. das Verstehen mentaler Zustände anderer Personen (Theory of Mind) die Entwicklung von exekutiven Funktionen (z. B. Selbstregulation) voraus (Perner & Lang, 1999). Beides entwickelt sich erst im Alter von etwa vier Jahren (ebd.). Darüber hinaus beeinflussen diverse intra- und interpersonelle Faktoren das Ausmaß empathischer Reaktionen. So kann beispielsweise die Art der Beziehung zu einem anderen Menschen beeinflussen, wie wir auf ihn reagieren (Preston & De Waal, 2002). In einer fMRT Studie aus dem Jahr 2003 konnte gezeigt werden, dass Männer und Frauen, je nachdem ob sie bereits Eltern waren oder nicht, unterschiedlich auf weinende und lachende Kinder reagierten (Seifritz et al., 2003). Frauen reagierten, unabhängig davon ob sie bereits Mutter waren, sowohl auf weinende als auch auf lachende Kinder mit einer neuronalen Deaktivierung im vorderen Gyrus cinguli (ebd.). Bei Männern konnte diese Deaktivierung nicht beobachtet werden (ebd.). Bei Probanden mit eigenen Kindern veränderten sich zusätzlich die Antwortmuster in Amygdala und den damit verbundenen limbischen Strukturen bei beiden Geschlechtern (ebd.). Männer und Frauen, die noch nicht Eltern waren, zeigten eine stärkere Aktivität bei lachenden Kindern, Eltern dagegen eine stärkere Aktivität gegenüber den weinenden Kindern (ebd.).

Das Empathievermögen bzw. die individuelle empathische Reaktion kann sich also in bestimmten Situationen, abhängig von persönlichen Erfahrungen, verändern.

2.1.3. Notwendigkeit der longitudinalen Erfassung des Empathievermögens

Studien deuten darauf hin, dass das Empathievermögen auch bei bestimmten psychiatrischen Krankheitsbildern, beispielsweise dem Autismus oder der Depression, vermindert bzw. verändert ist (Bora & Berk, 2016; Dziobek et al., 2008; Schreiter et al., 2013; Wolkenstein et al., 2011). So gibt es Hinweise darauf, dass bei Patienten mit einer Depression im Vergleich zu gesunden Kontrollen ein Empathiedefizit vorliegt (Schreiter et al., 2013; Wolkenstein et al., 2011). Darüber hinaus scheint bezüglich der Depression ein geringes Empathievermögen ein Risikofaktor für das erneute Auftreten einer depressiven Episode zu sein: In einer Studie von 2015 wurden 100 von einer Depression remittierte Patienten bezüglich ihres Empathievermögens untersucht und anschließend über ein Jahr begleitet (Yamada et al., 2015). Es zeigte sich, dass Patienten, welche Defizite im Empathievermögen aufwiesen, auch signifikant häufiger an einer depressiven Episode erkrankten (ebd.). Ebenso zeigten sich in einer Studie Defizite im Bereich der kognitiven Empathie bei Patienten mit chronischer Depression (Mattern et al., 2015). Die chronische Depression ist eine stark beeinträchtigende psychische Störung, bei der die Symptome einer Depression über mehrere Jahre, meist schon im frühen Lebensalter beginnend, vorliegen (Brakemeier et al., 2012). Die chronische Depression gilt als besonders therapieresistent (Gilmer et al., 2008; Köhler et al., 2016) und weist ein höheres Suizidrisiko auf als die episodische Depression (Blanco et al., 2010). Zudem berichten die Patienten häufig über traumatisierende Beziehungserfahrungen in der Kindheit, beispielsweise emotionale oder körperliche Vernachlässigung und Missbrauch (Hölzel et al., 2011; Wiersma et al., 2009). Hierdurch entwickelt sich ein hohes Misstrauen in zwischenmenschlichen Beziehungen, das unter anderem in sozialem Rückzug und der Entkopplung der eigenen Wahrnehmung von der Umwelt resultiert (McCullough, 2000). Betroffene verbleiben in einem kognitiv-emotionalen Zustand, der dem von Kindern zwischen vier und sieben Jahren ähnlich ist und der letztlich zu dysfunktionalen Denk- und Verhaltensweisen führt, wodurch Patienten mit chronischer Depression ihre Ziele und Bedürfnisse immer weniger erreichen können (ebd.) Ein Defizit in der kognitiven

Empathie könnte auf diese Weise zu einer Aufrechterhaltung der chronischen Depression führen, da sich hierdurch negative Beziehungserfahrungen wiederholen und korrigierende positive Erfahrungen andererseits erschwert werden (ebd.). Die einzige speziell für die chronische Depression entwickelte Therapieform, CBASP (Cognitive Behavioral Analysis System of Psychotherapy), beruht unter anderem auf der Annahme, dass durch eine Verbesserung der kognitiven Empathieprozesse auch eine Verbesserung der dysfunktionalen Kommunikationsstrategien und somit der Erkrankung insgesamt herbeigeführt werden kann (McCullough, 2000). Bislang existieren jedoch keine konsistenten Daten zu Unterschieden in der Empathiefähigkeit von chronisch depressiven, episodisch depressiven und Gesunden (Derntl et al., 2012; Thoma et al., 2011; van Randenborgh et al., 2012; Wilbertz et al., 2010). Derntl et al. (2012) fanden Defizite bei Patienten mit schwerer Depression ausschließlich im Bereich der emotionalen Empathie. Thoma et al. (2011) fanden keine signifikanten Unterschiede im Empathievermögen zwischen Patienten mit unipolarer Depression und einer Kontrollgruppe, wobei sich die depressiven Patienten selber dennoch als weniger empathisch einschätzten. In der Studie von van Randenborgh et al. (2012) zeigten sich keine signifikanten Unterschiede im Bereich der kognitiven Empathie zwischen chronisch und episodisch depressiven Patienten. Wilbertz et al. (2010) konnten ebenfalls keine signifikanten Unterschiede zwischen chronisch depressiven Patienten und einer Kontrollgruppe finden. Gleichwohl schätzten sich die chronisch depressiven Patienten als signifikant weniger empathisch ein, was auch von deren Therapeuten so empfunden wurde (ebd.). Hieraus kann geschlossen werden, dass durchaus der klinische Eindruck eines Empathiedefizits besteht, welches bislang jedoch noch nicht ausreichend belegt werden konnte. Anhand bisheriger Studien war es zudem nicht möglich, eine Aussage zu treffen, ob eine mögliche Verminderung der Empathiefähigkeit bei der Depression eine Ursache oder Folge derselben ist (Schreiter et al., 2013). Möglicherweise ist ein vermindertes Empathievermögen also als aufrechterhaltender oder sogar ursächlicher Faktor anzusehen, so dass eine Zunahme bzw. Veränderung der Empathiefähigkeit in die therapeutische Arbeit miteinbezogen werden könnte. Eine weitere Theorie besteht darin, dass die beschriebenen Defizite des Empathievermögens nur dann auftreten, wenn zuvor eine Aktivierung emotional traumatisierender Beziehungserfahrungen erfolgt ist (ebd.). Es besteht daher die Notwendigkeit weiterer Forschung in diesem Bereich. Schreiter et al. (2013) empfehlen hierfür die Durchführung longitudinaler Kohortenstudien. Zudem soll die Empathiefähigkeit von depressiven Probanden nach einer spezifischen emotionalen Aktivierung getestet werden (Schreiter et al., 2013; Wilbertz et al.,

2010). Bora & Berk weisen darauf hin, dass zukünftige Studien Follow-up Messungen nutzen könnten, um die Empathiefähigkeit während verschiedener Stadien der Depression zu messen (Bora & Berk, 2016). Damit könnten die bisher beschriebenen Defizite möglicherweise besser verstanden und genutzt werden (ebd.). Voraussetzung hierfür ist das Empathievermögen objektiv und in seiner Mehrdimensionalität messen zu können.

2.2. Testverfahren zur Messung des Empathievermögens

Im Folgenden werden zunächst einige bereits etablierte Testverfahren zur Messung von Empathie bei Erwachsenen betrachtet und bezüglich ihrer Eignung für die geforderten Messungen beurteilt. Eine Übersicht über die vorgestellten Testverfahren gibt **Tabelle 1**.

2.2.1. Reading the Mind in the Eyes Test (RMET)

Der “Reading the Mind in the Eyes Test” wurde von Simon Cohen-Baron entwickelt (Baron-Cohen et al., 2001). Beim RMET werden die Probanden gebeten, sich nacheinander 36 Bilder von Augenpaaren anzusehen und dabei nur anhand dieser zu entscheiden, was die abgebildete Person gerade fühlt (siehe **Abbildung 2**). Hierfür werden in der ursprünglichen Version zwei, in der überarbeiteten Version jeweils vier Antwortmöglichkeiten vorgegeben, von denen nur eine zutrifft (ebd.).



Abbildung 2: Beispiel für ein Augenpaar aus dem RMET (Baron-Cohen et al., 2001).

Der RMET ist ein Testverfahren, welches ausschließlich die kognitive Komponente der Empathie erfasst (ebd.). Darüber hinaus werden statische Bilder der Augenpartie beurteilt, wogegen uns im realen Leben viel mehr Eindrücke zur Verfügung stehen (ebd.). Für die deutsche Version wurde in einer Studie von 2013 eine Retest-Reliabilität von $r = 0,66$ angegeben (Pfaltz et al., 2013).

2.2.2. Movie for the Assessment of Social Cognition (MASC)

Der Movie for the Assessment of Social Cognition (MASC) ist ein videobasierter Test, der ursprünglich für die Empathiemessung bei Patienten mit Asperger-Syndrom konstruiert wurde (Dziobek et al., 2006). Die Probanden sehen hierbei einen 15-minütigen Kurzfilm an, in welchem vier Personen im Rahmen einer Dinnerparty miteinander interagieren. Der Film wird insgesamt 46-mal angehalten, wobei den Probanden Fragen zu den Gefühlen, Gedanken und Absichten der Charaktere gestellt werden (ebd.). Der MASC besitzt mit $\alpha = 0,84$ eine gute Gesamtreliabilität (ebd.). Auch der MASC beurteilt jedoch nur die kognitive Empathie.

2.2.3. Baron-Cohen's Empathy Quotient (EQ)

Der Empathy Quotient (EQ), entwickelt von Simon Baron-Cohen, ist ein Fragebogen bestehend aus 60 Items (Baron-Cohen & Wheelwright, 2004). Er enthält 40 Empathie-Items sowie 20 Kontroll-Items. Jedes Item besteht aus einer Aussage, für welche die Versuchsperson den Grad ihrer Zustimmung auf einer vierstufigen Antwortskala angeben muss (ebd.). Bei jedem Empathie-Item kann die Versuchsperson 0, 1 oder 2 Punkte erreichen, sodass sich eine Gesamtpunktzahl von 0 (am wenigsten empathisch) bis 80 Punkten (maximal empathisch) ergibt. Beispiel: „I am good at predicting how someone will feel“ (ebd.). Der EQ weist mit $\alpha = 0,92$ eine gute Reliabilität auf (Lawrence et al., 2004). Er beruht jedoch, wie alle Selbstbeurteilungsfragebögen, ausschließlich auf der eigenen, subjektiven Einschätzung der Empathiefähigkeit und weist damit eine geringe ökologische Validität auf. Zudem wird im EQ hauptsächlich die kognitive Komponente der Empathie abgefragt (ebd.).

2.2.4. Interpersonal Reactivity Index (IRI)

Der Interpersonal Reactivity Index (IRI) ist ein Fragebogen zur Selbstbeurteilung des Empathievermögens, welcher 28 Items auf vier Skalen enthält (Davis, 1980):

- Fantasy (die Fähigkeit, sich in Roman- oder Filmfiguren hineinzusetzen)
Beispielitem: „I really get involved with feelings of characters in a novel.”
- Perspective-Taking (kognitive Empathie, Hineinversetzen in die Perspektive anderer Personen)
Beispielitem: „When I'm upset at someone, I usually try to “put myself in his shoes” for a while.”
- Empathic Concern (emotionale/affektive Empathie, Mitgefühl gegenüber anderen Personen empfinden)
Beispielitem: „I am often quite touched by things that I see happen”
- Personal Distress (persönliche Betroffenheit/unangenehme Empfindungen, wenn andere Personen negative oder angstausslösende Situationen erleben).
Beispielitem: „I tend to lose control during emergencies.”

Der IRI berücksichtigt insbesondere die Mehrdimensionalität der Empathie und ermöglicht die getrennte Erfassung von kognitiver und emotionaler Komponente (ebd.). In der ursprünglichen Validierung wurde eine akzeptable Reliabilität von $\alpha = 0,70$ bis $0,78$ angegeben (Davis, 1960). Jedoch weisen einige Folgestudien dem widersprechend auf eine unzureichende Reliabilität hin (Beven et al., 2004; Ireland, 1999). Darüber hinaus sollte neben der geringen ökologischen Validität (Dziobek et al., 2008) auch der Aspekt der sozialen Erwünschtheit berücksichtigt werden, welcher ebenfalls einen Einfluss auf die Zuverlässigkeit des Tests haben könnte (Curwen, 2003). Die überarbeitete Version des deutschen IRI, der Saarbrücker Persönlichkeitsfragebogen, weist zwar bessere Reliabilitätswerte auf, jedoch konnte die geringe ökologische Validität aufgrund des Testdesigns als Fragebogen nicht verbessert werden (Paulus, 2009).

Eine weitere Schwierigkeit, welche bei der Empathiemessung mittels Fragebögen generell besteht, ergibt sich daraus, dass die Erfassung dieses Konstruktes mittels Fragebögen voraussetzt, dass die Versuchsperson die Fähigkeit zu Abstraktion und Introspektion besitzt. Diese ist jedoch beispielsweise bei Menschen mit Autismus-Spektrum-Störung (ASD) vermindert (Happé, 1994; Minshew et al., 2002).

2.2.5. Multidimensionaler Empathietest (MET)

Der Multidimensionale Empathietest (MET) ist ein fotobasierter Test, der sowohl die kognitive als auch die emotionale/affektive Komponente der Empathie berücksichtigt und eine höhere ökologische Validität sowie Reliabilität ($\alpha = 0,71$ für die kognitive, $\alpha = 0,91$ bis $0,92$ für die emotionale Empathie) als der IRI aufweist (Dziobek et al., 2008). Der MET besteht aus einer Sequenz von Fotos, auf denen Menschen verschiedenen Alters und Geschlechts in unterschiedlichen emotionalen Zuständen und mit realistischem Alltagsbezug abgebildet sind (ebd.). Die Versuchspersonen werden im Testverlauf gebeten, jeweils drei Fragen zu den präsentierten Fotos zu beantworten:

1. Wie fühlt sich diese Person? (kognitive Empathie)
2. Wie sehr erregt Sie dieses Foto? (indirekte/implizite emotionale Empathie)
3. Wie sehr fühlen Sie mit der Person? (direkte/explicite emotionale Empathie)

Zur Beantwortung der ersten Frage („Wie fühlt sich diese Person?“) werden den Versuchspersonen jeweils vier Antwortoptionen vorgegeben, von denen sie die richtige Antwort (= Target-Emotion) aus drei falschen Antworten (= Distractor-Emotionen) auswählen sollen (siehe **Abbildung 3**).



Abbildung 3: Beispielitem der kognitiven Empathie im MET.

Die beiden anderen Fragen sollen von den Versuchspersonen auf einer Skala von „1“ (gar nicht erregt/mitfühlend) bis „9“ (sehr stark erregt/mitfühlend) bewertet werden (siehe **Abbildung 4**). Ein zeitliches Limit existiert für den Test nicht. Der MET ermöglicht die simultane und dennoch getrennte Erfassung sowohl der kognitiven als auch der emotionalen Empathie bei unterschiedlichen Stichproben bzw. Patientengruppen.



Abbildung 4: Beispielitem der emotionalen Empathie im MET.

Da das ursprüngliche Bildmaterial weitgehend aus erwachsenen Männern und Frauen bestand, wurde der Test zur Anwendung bei Jugendlichen ab zwölf Jahren (beispielsweise mit Autismus-Spektrum-Störung) neu überarbeitet (MET-J, Multidimensionaler Empathie Test - Jugend) (Poustka et al., 2010). Hierbei konnte gezeigt werden, dass Jugendliche mit Autismus-Spektrum-Störung (ASD) eine geringere kognitive Empathie aufwiesen als die Kontrollgruppe, wobei die Werte für die emotionale Empathie sich nicht signifikant unterschieden (ebd.). Weitere bisherige Anwendungsbereiche des MET war die Untersuchung von Patienten mit Persönlichkeitsstörungen (Dziobek et al., 2011; Ritter et al., 2011). Es zeigte sich, dass Patienten mit narzisstischer Persönlichkeitsstörung eine geringere emotionale Empathie, jedoch keine Defizite in der kognitiven Empathie aufwiesen (Ritter et al., 2011). Patienten mit einer emotional-instabilen Persönlichkeitsstörung vom Borderline-Typ zeigten in der emotionalen sowie auch in der kognitiven Empathie geringere Werte (Dziobek et al., 2011).

Eine weitere Überarbeitung des MET fand 2012 zur Untersuchung von vermuteten Empathiedefiziten im Rahmen der Pädophilieforschung statt (Lemme, 2012; Schuler et al., 2019). In dieser neu überarbeiteten Version, dem Multidimensionalen Empathie Test – Kinder/Erwachsene (MET-KE), variieren die abgebildeten Personen systematisch nach

Alter (Kind/Erwachsener), Geschlecht (weiblich/männlich) und Emotionsvalenz (positiv/negativ). Dadurch können Fragestellungen beantwortet werden, bei denen es notwendig ist, die Zielpersonen nach Alter und Geschlecht zu unterscheiden (Lemme, 2012). Diese Adaptation des MET besteht aus insgesamt 80 Bildern und ist im Blocksystem aufgebaut. Das heißt, die Versuchspersonen müssen jeweils für eine Gruppe von zehn Bildern eine der drei Fragen beantworten bevor die Fragestellung für den folgenden Bilderblock (aus wiederum zehn Bildern) wechselt (ebd.). Dabei ergab sich im MET-KE für den kognitiven Empathieteil eine sehr gute Gesamtreliabilität von $\alpha = 0,92$ (Lemme, 2012). Der MET-KE wurde schließlich für eine Studie im Rahmen der Pädophilieforschung verwendet (Schuler et al., 2019). Hier zeigte sich, dass pädophile Männer, welche noch nicht übergriffig geworden waren, eine signifikant höhere kognitive Empathie gegenüber Kindern zeigten als bereits übergriffig gewordene pädophile Männer (ebd.). Dabei wurden in beiden Gruppen die Bilder mit Kindern im Vergleich zu denen, welche Erwachsene zeigten, zuverlässiger erkannt (ebd.). Die emotionale Empathie gegenüber Kindern war in der Gruppe der Pädophilen sogar signifikant größer als in der gesunden Kontrollgruppe (ebd.).

Tabelle 1: Übersicht der Testverfahren zur Messung von Empathie.

Testverfahren	Art des Tests	Mehrdimensio- nale Messung	Anzahl Items	Reliabilität
Reading the Mind in the Eyes Test (RMET)	Fotobasiert	nein	36	$r = 0,66$
Movie for the Assessment of Social Cognition (MASC)	Kurzfilm	nein	50	$\alpha = 0,84$
Empathy Quotient (EQ)	Fragebogen	teilweise	60	$\alpha = 0,92$
Interpersonal Reactivity Index (IRI)	Fragebogen	ja	28	$\alpha = 0,70$ bis $0,78$
Multidimensionaler Empathietest – Kinder/Erwachsene (MET-KE)	Fotobasiert	ja	80	$\alpha = 0,92$ (kognitiv)

r = Korrelationskoeffizient in der Retest-Reliabilität, α = Cronbachs Alpha.

2.3. Erstellung von Parallelversionen

2.3.1. Notwendigkeit der Erstellung zweier Parallelformen des MET

Der multidimensionale Empathietest (MET) scheint also unter den bisher etablierten Testverfahren aufgrund seiner Mehrdimensionalität sowie der guten Reliabilität ein geeignetes Instrument zur Messung des Empathievermögens zu sein. Um jedoch, wie eingangs gefordert, mögliche Veränderungen des Empathievermögens, beispielsweise nach Remission oder nach einer therapeutischen Intervention, erfassen zu können, benötigt man ein Testverfahren, welches Empathie im longitudinalen Verlauf beurteilt. Hierfür müsste man eine Verlaufsmessung durchführen. Die einfachste Möglichkeit einer solchen Verlaufsmessung wäre es, den Probanden denselben Test (in unserem Fall den MET) ein zweites Mal zu präsentieren. Ein Störfaktor hierbei sind jedoch Erinnerungseffekte (Schermelleh-Engel & Werner, 2012). Diese beinhalten, dass sich Probanden bei einer

Testwiederholung an ihre Antworten aus dem ersten Durchgang erinnern und daher beim zweiten Durchgang dieselben Antworten geben, wodurch die wahren Messwerte verzerrt würden. Diese Erinnerungseffekte treten verstärkt bei geringen Zeitabständen zwischen den Messungen auf (ebd.). Eine Möglichkeit, um die beschriebenen Erinnerungseffekte zu vermeiden, besteht in der Verwendung von Parallelversionen, wobei Parallelversionen definiert sind als „voneinander verschiedene Varianten des gleichen Testverfahrens, die das gleiche Merkmal mit der gleichen Genauigkeit erfassen“ (Schermelleh-Engel & Werner, 2012, S. 125-126). Man benötigt folglich entsprechende Parallelversionen des MET, um das Empathievermögen wie gewünscht im longitudinalen Verlauf messen zu können.

2.3.2. Verfahren zur Erstellung von Parallelformen eines Tests

Da sich Parallelformen eines Tests nur unter großem Aufwand erstellen lassen, existieren selbige aktuell nur für wenige renommierte Tests (Schermelleh-Engel & Werner, 2012). Wie kann die Konstruktion von Parallelformen konkret aussehen, wenn bereits eine ursprüngliche Version des Tests existiert? Hierzu betrachten wir beispielhaft die Dysfunctional Attitude Scale (DAS). Die DAS ist ein Fragebogen zur Erfassung dysfunktionaler Grundeinstellungen und wurde zunächst als Test mit 100 Items konstruiert (Weissman & Beck, 1978). In der Validierungsstudie wurde durch die Teilnehmer die Länge des Tests kritisiert, woraufhin Weissmann & Beck (1978) aus den ursprünglichen 100 Items zwei Parallelversionen mit je 40 Items erstellten. Hierfür bearbeiteten 275 Studenten der Pennsylvania State University die DAS. Anschließend wurden mithilfe einer Faktorenanalyse Itempaare gebildet und diese auf die beiden Testhälften aufgeteilt (ebd.). Um sicher zu gehen, dass die so entstandenen Versionen parallel waren, wurden sie an 70 Studenten erprobt. In der statistischen Auswertung wurden die Ergebnisse der Parallelversionen miteinander korreliert; zusätzlich wurden Mittelwerte und Reliabilität der Versionen verglichen (ebd.). 2014 fand aufgrund schlechter psychometrischer Werte (Hautzinger et al., 2005; Power et al., 1994) eine erneute Überarbeitung der DAS als 18-Item-Parallelversionen statt (Rojas et al., 2014). Hierfür wurde aus den Daten einer früheren Studie (Hautzinger et al., 2005) erneut eine Faktorenanalyse durchgeführt und die Items mit ähnlichen Werten auf die beiden neuen Parallelversionen verteilt (Rojas et al., 2014). Im

Anschluss fand auch hier eine Überprüfung der Parallelität anhand zweier Stichproben statt (ebd.). In der anschließenden Auswertung wurden Trennschärfe und interne Konsistenz (Cronbachs Alpha) überprüft. T-Tests wurden eingesetzt, um Unterschiede zwischen den Skalen der Parallelversionen DAS-18A und DAS-18B zu detektieren. Zusätzlich wurde eine Faktorenanalyse durchgeführt, um die Faktorenladungen zu überprüfen (ebd.).

Zusammenfassend erfolgte die Konstruktion der Parallelversionen in drei Schritten: Erstellung der Testhälften, Erprobung der so entstandenen Parallelversionen, Überprüfung der Parallelität durch statistische Auswertung. Hierbei stellt die Faktorenanalyse, wie auch für die DAS angewendet, eine geeignete Methode für die Erstellung und auch Überprüfung der Parallelversionen dar (Schermelleh-Engel & Werner, 2012). Jedoch müssen für eine solche Faktorenanalyse die zugrunde liegenden Items ordinal-, besser noch intervallskaliert sein (Jöreskog & Moustaki, 2001). Bei Items, welche wie im MET nominalskaliert sind (richtige/falsche Antwort), kann dieses Verfahren nicht angewendet werden. Gulliksen (2013) schlug vor, Items anhand ihrer Kennwerte aus der Itemanalyse auf die beiden parallelen Testhälften aufzuteilen und beschrieb hierbei als Kriterien Schwierigkeit und Trennschärfe. Auch Schermelleh-Engel & Werner (2012) nannten die Aufteilung nach Schwierigkeit und Trennschärfe der Items als eine Möglichkeit, parallele Testversionen zu erhalten. Der Itemschwierigkeitsindex drückt aus, welcher Anteil der Probanden das Item richtig gelöst hat und kann Werte zwischen 0 und 100 annehmen (Kelava & Moosbrugger, 2012). Er dient dazu, Probanden mit hoher und niedriger Merkmalsausprägung zu unterscheiden. Laut Kelava & Moosbrugger (2012) können mittlere Schwierigkeiten (um $P = 50$) am ehesten Probanden mit hoher und niedriger Merkmalsausprägung differenzieren. Items mit einem Schwierigkeitsindex um 100 (sehr einfach) oder um 0 (sehr schwierig) können dagegen nicht gut differenzieren. Die Trennschärfe eines Items beschreibt die Korrelation zwischen Item- und Testwert der Probanden (ebd.). Dabei werden Trennschärfen von 0,4 bis 0,7 als „gut“ gewertet; Trennschärfen um und unter 0 sind ungeeignet, um Probanden mit hoher und niedriger Merkmalsausprägung zu unterscheiden (ebd.). Die Parallelität zweier Testverfahren kann schließlich neben der Faktorenanalyse auch indirekt über die Paralleltest-Reliabilität geprüft werden (Schermelleh-Engel & Werner, 2012). Die Paralleltest-Reliabilität ist hoch, wenn die beiden parallelen Testvarianten dieselben Mittelwerte und Streuungen aufweisen und sie ausreichend hoch

korrelieren und kann über die Korrelation der Testwerte der beiden Versionen geschätzt werden (ebd.).

2.4. Zielsetzung

Das Ziel der vorliegenden Arbeit war es, zwei Parallelversionen der jüngsten Adaptation des Multidimensionalen Empathietests (MET-KE) zu erstellen, welche dann für zukünftige Studien, beispielsweise im Rahmen der Depressionsforschung, genutzt werden können. Aus dem ursprünglichen Test mit 80 Items sollten zwei parallele Versionen MET-A und MET-B mit je 40 Items erstellt und auf ihre Vergleichbarkeit geprüft werden. Hierfür wurden Mittelwerte, Korrelationen, Itemschwierigkeitsindizes und Trennschärfen der beiden Versionen berechnet und verglichen.

3. Methodik

Nach dem positiven Ethikvotum der Ethikkommission der Charité (Antragsnummer EA1/324/15) begann am 01.10.2016 die Arbeit zur Erstellung und Erprobung der Parallelversionen des Multidimensionalen Empathietests.

3.1. Erstellung der Parallelversionen des Multidimensionalen Empathietests

Zur Erstellung der Parallelförmigen wurde das ursprüngliche Bildmaterial des MET-KE mit den zugehörigen Target- und Distractor-Emotionen (Lemme, 2012; siehe **Anhang 1**) sowie die damit erhobenen Daten von 263 Probanden aus dem NeMUP-Forschungsverbund (Neurobiologische Grundlagen von Pädophilie und sexuellem Missbrauchsverhalten gegenüber Kindern, siehe <http://www.nemup.de>) durch Prof. Dr. rer. nat. Isabel Dzio-bek vom Institut für Psychologie der Humboldt Universität Berlin zur Verfügung gestellt. Die Probanden setzten sich zusammen aus pädophilen, hebephilen und gesunden Kontrollprobanden, welche den MET-KE in Berlin, Hannover, Kiel, Magdeburg und Essen im Rahmen des Forschungsprojektes bearbeitet hatten. Zusätzlich wurden die Daten der abschließenden Validierung des MET-KE (n = 52) bereitgestellt, welche durch Benjamin Lemme im Rahmen seiner Bachelorarbeit erhoben wurden (Lemme, 2012). Diese Stichprobe bestand aus Probanden mit niedrigem Bildungsstand (Hauptschulabschluss) und aus Probanden mit Autismus-Spektrum-Störung (ebd.).

3.1.1. Erstellung von Itempaaren

Im ersten Schritt wurden für alle 80 Bilditems des MET-KE im Rahmen der Itemanalyse Itemschwierigkeit und Trennschärfe für beide bestehende Datensätze (NeMUP, Lemme) berechnet. Hierbei wurde die Itemschwierigkeit anhand folgender Formel als korrigierter Schwierigkeitsindex P (nach Kelava & Moosbrugger, 2012) berechnet:

$$P = \frac{NR - \left[\frac{NF}{(m-1)} \right]}{N} \times 100$$

NR = Anzahl der Testpersonen, die das Item richtig gelöst haben

NF = Anzahl der Testpersonen, die das Item falsch gelöst haben

m = Anzahl der Auswahlmöglichkeiten

N = Anzahl der Versuchspersonen

Anhand der ermittelten Werte wurden Itempaare gebildet, welche einerseits eine ähnliche Schwierigkeit und Trennschärfe aufwiesen und sich andererseits in den übrigen Itemmerkmalen (Geschlecht, Alter, Emotionsvalenz positiv/negativ) möglichst nicht unterschieden. Von diesen Itempaaren wurde in Absprache mit Prof. Dr. rer. nat. Dziobek jeweils eines der einen, das zweite Item der anderen Testversion (A, B) zugeteilt, sodass sich zwei Versionen mit je 40 Items ergaben. Zum Vergleich der auf diese Weise entstandenen MET-Versionen wurden die Mittelwerte von Trennschärfe und Schwierigkeit in den bereits erhobenen Stichproben des NeMUP-Projektes und der Validierung durch Lemme (2012) verglichen. Die Daten wurden mithilfe des Shapiro-Wilk-Tests, sowie eines Q-Q-Plots auf Normalverteilung getestet und anschließend mittels t-Test auf statistisch signifikante Unterschiede geprüft.

3.1.2. Erstellung der Frageblöcke

Wie in der Ursprungsversion des MET-KE sollten auch in den Parallelversionen die Bilditems in Frageblöcken präsentiert werden, wobei aus den ursprünglichen drei Fragestellungen für jedes der Items nur zwei beantwortet werden sollten. Diese waren:

1. Wie fühlt sich diese Person? (kognitive Empathie)
2. Wie sehr fühlen Sie mit dieser Person? (direkte emotionale Empathie)

Blöcke der beiden Fragestellungen sollten sich regelmäßig abwechseln, wobei für jedes Item Frage 1 („Wie fühlt sich diese Person?“) zuerst beantwortet werden sollte. Es wurden schließlich für jede Version acht Frageblöcke à zehn Items (vier Blöcke zum Abfragen der kognitiven und vier Blöcke mit jeweils identischen Items zum Abfragen der emotionalen Empathie) erstellt. Hierbei wurde darauf geachtet, dass der Mittelwert der Itemschwierigkeiten in allen Blöcken ähnlich und die Blöcke nach Emotionsvalenz (positiv/negativ) geordnet waren. Darüber hinaus sollten in allen Blöcken Items verschiedener Kategorien (Erwachsen/Kind, weiblich/männlich) vorhanden sein. Schließlich ergab sich für beide Testhälften das in **Abbildung 5** dargestellte Frageblockdesign.

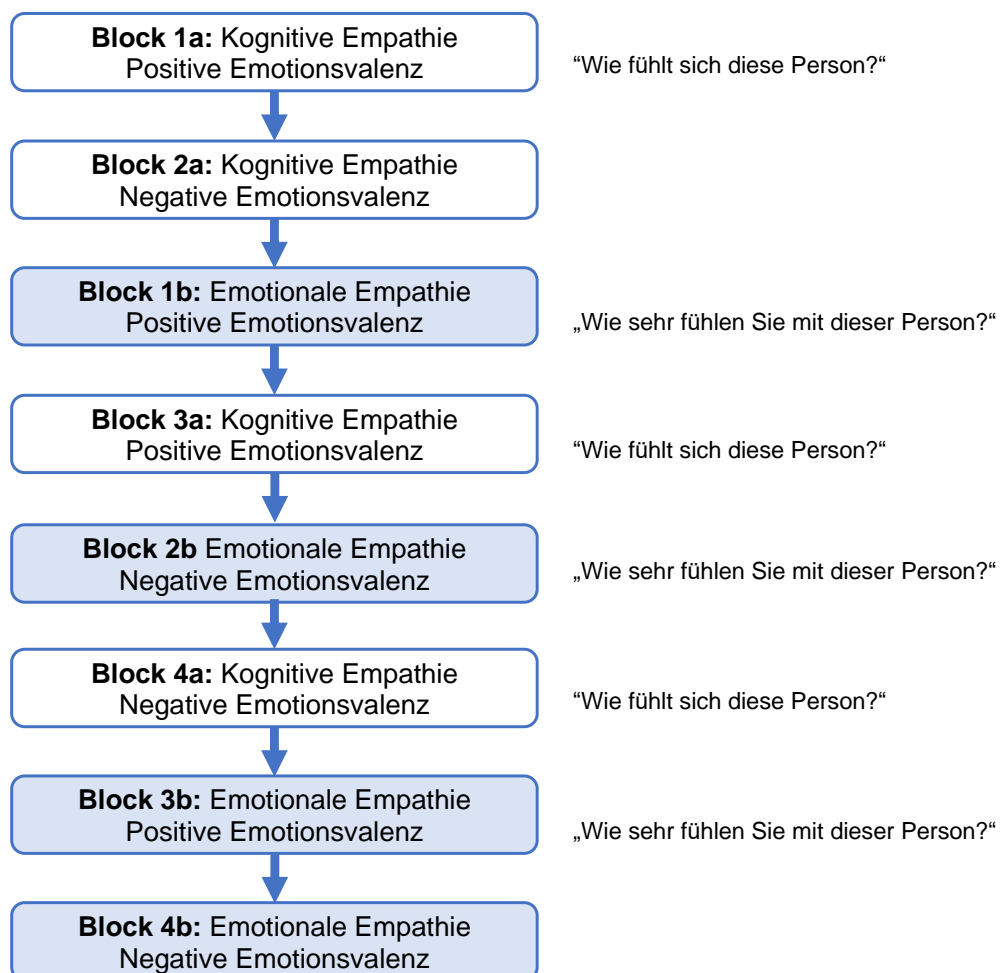


Abbildung 5: Abfolge der Frageblöcke in den Parallelversionen A und B.

3.1.4. Programmierung der Parallelversionen

Die Programmierung der MET-Versionen A und B (MET-A, MET-B) erfolgte durch Herrn Matthias Kupferschmidt (Cupfer Intelligent Systems UG, Berlin) in Java 8. Die Programmierung fand in äquivalentem Design der ursprünglichen MET-KE Version statt (siehe **Abbildung 6 und 7**). Geantwortet wird jeweils durch Betätigen der Tasten A, B, C, D bzw. 1 bis 9 auf der Tastatur.

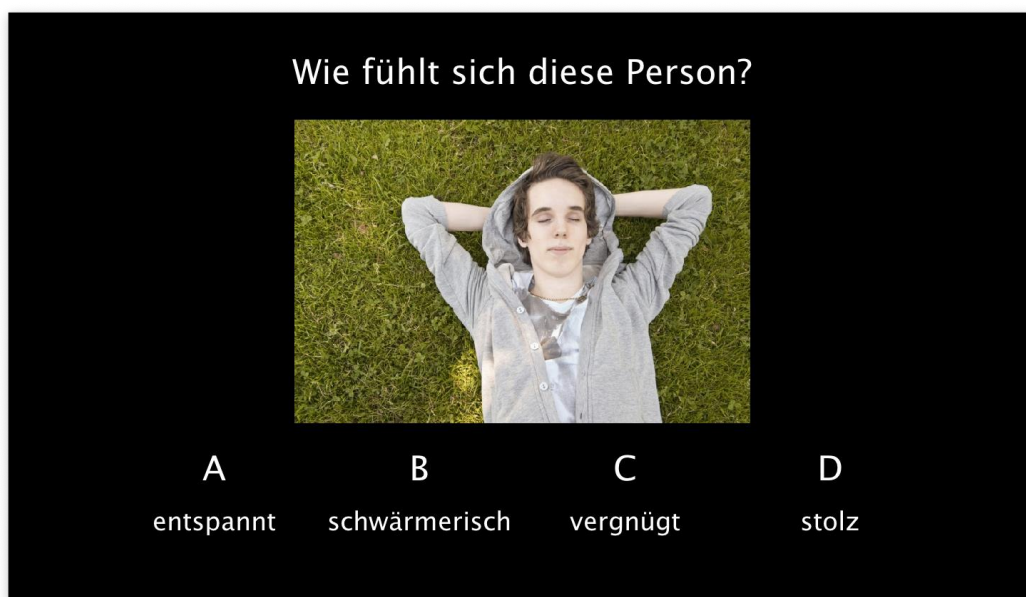


Abbildung 6: Beispiel für die Messung der kognitiven Empathie aus MET-B.

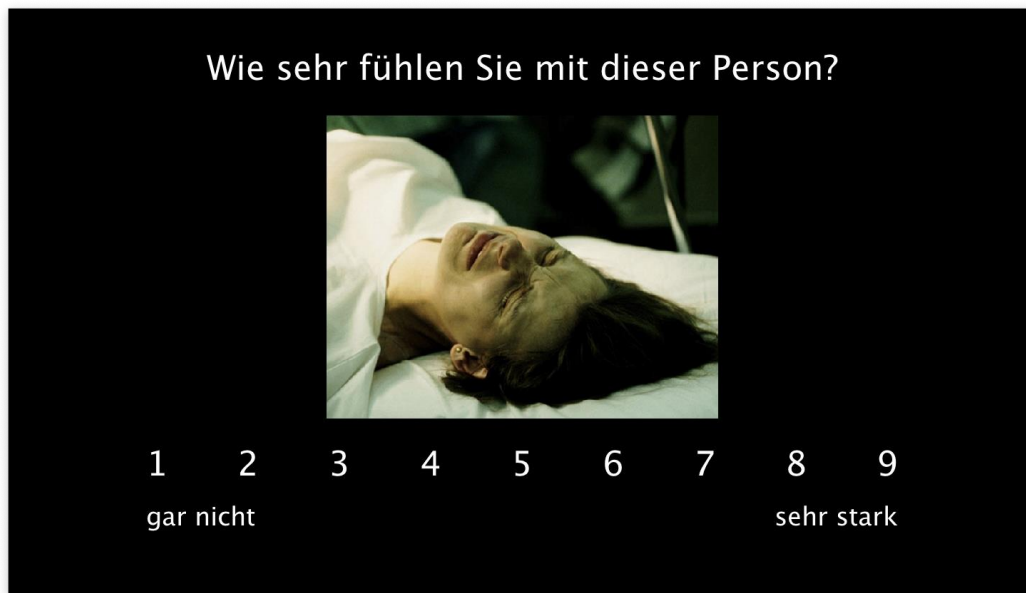


Abbildung 7: Beispiel für die Messung der emotionalen Empathie aus MET-B.

Vor Beginn einer neuen Aufgabenstellung wird die zugrunde liegende Frage („Wie fühlt sich diese Person?“ bzw. „Wie sehr fühlen Sie mit dieser Person?“) erneut einzeln präsentiert. Die Reihenfolge der zehn Items innerhalb der jeweiligen Frageblöcke sowie die Anordnung der Antworten (A-D) innerhalb der Abfrage der kognitiven Empathie ergibt sich bei jedem Start des Programmes zufällig, so dass mögliche Einflüsse der Präsentationsreihenfolge ausgeschlossen werden können. Die Antwortlatenzen in Millisekunden werden durch das Programm gemessen. Nach Beendigung des Programmes (durch Beantworten des letzten Items) wird automatisch eine aufsteigend nummerierte CSV-Datei mit den Rohdaten innerhalb des Programmordners erstellt. Diese enthält für die kognitiven Empathie die Information welche Emotion als Antwort ausgewählt wurde, ob die Antwort richtig war (richtig = 1, falsch = 0) und wie schnell geantwortet wurde (in ms). Für die emotionalen Empathie wird die ausgewählte Antwort (1-9) sowie die Antwortlatenz (in ms) gespeichert.

3.2. Erprobung der Parallelversionen

Die beiden Versionen MET-A bzw. MET-B wurden schließlich an einer Stichprobe gesunder Probanden erprobt, um sie hinsichtlich ihrer Parallelität überprüfen zu können.

3.2.1. Studienteilnehmer

Die Probanden wurden durch die Doktorandin persönlich in deren Umfeld sowie unter den Mitarbeitern der Klinik für Psychiatrie und Psychotherapie der Charité Mitte im Zeitraum von Oktober 2016 bis Dezember 2016 bzw. von März 2017 bis November 2017 rekrutiert. Folgende Kriterien mussten erfüllt sein:

- Keine aktuelle psychische Störung
- Anamnestisch keine depressive Episode in der Vorgeschichte
- 24-Item Hamilton Depression Rating Scale (Hamilton, 1960) < 5 Punkte

Zur Überprüfung, ob die Teilnehmer die Kriterien zum Einschluss in die Studie erfüllten, wurden vorab das Mini-International Neuropsychiatric Interview (M.I.N.I.) (Sheehan et al., 1998) sowie die 24-Item Hamilton Depression Rating Scale (Hamilton, 1960) als Fremdbeurteilungsverfahren durchgeführt. Die erste Stichprobe setzte sich aus 20 Probanden unterschiedlichen Alters zusammen, welche aus verschiedenen Bildungsschichten stammten. Ziel bei der Rekrutierung der ersten Stichprobe war, eine möglichst heterogene Gruppe von Teilnehmern zu erhalten, welche als Kontrollgruppe für weiterführende Studien mit episodisch oder chronisch depressiven Patienten genutzt werden könnte. Da aufgrund der Ergebnisse der ersten Stichprobe die Rekrutierung einer zweiten Stichprobe notwendig war, wurden weitere 15 Probanden rekrutiert. Hierbei wurden lediglich die oben genannten Ein- und Ausschlusskriterien beachtet.

3.2.2. Ablauf der Testung

Die Testung erfolgte an zwei Testtagen. Alle Teilnehmer wurden am ersten Tag mündlich und schriftlich über Ziel, Ablauf und Inhalt der Arbeit aufgeklärt und gaben ihr Einverständnis zur Erhebung, Auswertung und anonymisierten Veröffentlichung ihrer Daten. Das Einverständnis konnte zu jeder Zeit widerrufen werden. Sofern die Teilnehmer die Einschlusskriterien erfüllten, wurden sie in die Studie eingeschlossen. Andernfalls

wurden sie ausgeschlossen. Bei Einschluss in die Studie erfolgte eine ausführlichere klinische und neuropsychologische Testung der Probanden, welche insgesamt ca. 45-60 Minuten in Anspruch nahm. Es folgten die Instruktionen mit Beispielfragen zum Multidimensionalen Empathietest, welche den Studienteilnehmern anhand einer Microsoft Office Power-Point-Präsentation dargeboten wurden. Um systematische Übertragungseffekte auszuschließen (Schermelleh-Engel & Werner, 2012), fand im Anschluss eine randomisierte Bearbeitung der MET-Versionen statt. Das heißt, die Hälfte der Teilnehmer bearbeitete MET-A, die andere Hälfte MET-B. Am zweiten Testtag durften die Teilnehmer erneut die Instruktionen ansehen und bearbeiteten daraufhin die verbliebene MET-Version. Die Zeitspanne zwischen Tag 1 und 2 wurde so kurz wie möglich gewählt, da „bei längeren Intervallen zwischen der Durchführung beider Parallelformen [...] unsystematische Veränderungen der wahren Merkmalsausprägung der Testpersonen auftreten“ können (Schermelleh-Engel & Werner, 2012, S. 128). **Abbildung 8** zeigt zusammenfassend den Ablauf der Testung.

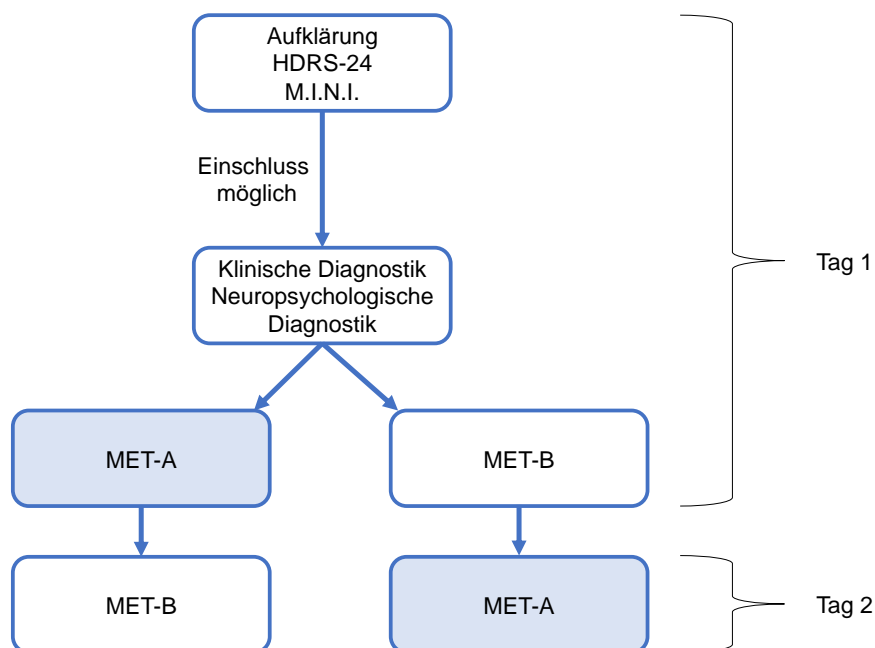


Abbildung 8: Ablauf der Testung an Tag 1 und Tag 2.

Die Testungen fanden in der Regel bei den Teilnehmern zu Hause oder am Arbeitsplatz statt, wobei darauf geachtet wurde, dass ein möglichst reizarmer Raum gewählt wurde,

in dem die Teilnehmer während der Testung am Computer keinen Blick auf Fotos, Bilder oder ähnliches hatten, welche sie von den emotionalen Zuständen des Testmaterials ablenken konnte.

3.2.3. Klinische und Neuropsychologische Diagnostik

Die klinische und neuropsychologische Diagnostik im Vorfeld der eigentlichen Testungen mittels MET bestand aus folgenden Komponenten:

- Hamilton Depression Rating Scale (HDRS-24)
- Mini-International Neuropsychiatric Interview (M.I.N.I.)
- Psychotherapie Basisdokumentation (Psy-BaDo)
- Beck-Depressions-Inventar (BDI-II)
- Childhood Trauma Questionnaire (CTQ)
- Toronto Alexithymie Skala (TAS)
- Zahlennachsprechen
- Regensburger Wortflüssigkeitstest
- Trail-Making-Test (TMT)
- Mehrfach-Wortschatz-Intelligenztest (MWT-B)
- Verbaler Lern- und Merkfähigkeitstest (VLMT)

3.2.3.1. Mini-International Neuropsychiatric Interview (M.I.N.I.)

Das Mini-International Neuropsychiatric Interview, kurz M.I.N.I., ist ein kurzes strukturiertes diagnostisches Interview, welches dem Untersucher ermöglicht, Diagnosen nach DSM-V und ICD-10 zu stellen (Sheehan et al., 1998). Hierbei werden die 17 häufigsten psychiatrischen Krankheitsbilder nach dem Ja-Nein-Prinzip abgefragt. Im Fall der depressiven Störung lautet die erste Frage beispielsweise: „Fühlten Sie sich in den letzten 2 Wochen beinahe jeden Tag und fast während des ganzen Tages traurig, niedergeschlagen oder deprimiert?“. Wird diese erste Frage mit „Ja“ beantwortet, folgen weitere Fragen

zur Sicherung der Diagnose. Die Bearbeitungszeit beträgt insgesamt etwa 15 Minuten (ebd.).

3.2.3.2. Hamilton Depression Rating Scale (HDRS-24)

Die Hamilton Depression Rating Scale ist eine Fremdbeurteilungsskala, mit welcher der Schweregrad einer Depression eingeschätzt werden kann. Sie wurde 1960 von Max Hamilton entwickelt (Hamilton, 1960). Für diese Studie wurde eine Skala mit 24 Items verwendet. Die Items beinhalten neben depressiver Verstimmung unter anderem Schuldgefühle, Suizidalität, Ein- und Durchschlafstörungen, sowie körperliche Symptomatik. Der Untersucher muss anhand eines Patienteninterviews beurteilen, wie stark das jeweilige Symptom in der vergangenen Woche auf einer Punkteskala ausgeprägt war, wobei für jeden Punktwert eine Kurzbeschreibung vorgegeben ist (Beispielitem Schuldgefühle: 0 Punkte – keine Schuldgefühle; 1 Punkt – Selbstvorwürfe; glaubt, Mitmenschen oder sich selbst enttäuscht zu haben; 2 Punkte – Schuldgefühle oder Grübeln über frühere Fehler und „Sünden“; 3 Punkte – Die jetzige Krankheit wird als Strafe gewertet, Versündigungswahn; 4 Punkte – Anklagende oder bedrohende akustische oder optische Halluzinationen). Insgesamt können 75 Punkte erreicht werden. In diese Studie wurden nur Teilnehmer mit einem Punktwert unter 5 Punkten eingeschlossen.

3.2.3.3. Psychotherapie Basisdokumentation (Psy-BaDo)

Mithilfe der Psychotherapie Basisdokumentation (Psy-BaDo) wurden soziodemographische Daten der Teilnehmer erhoben, darunter Alter, Bildungsstand, Wohnsituation und Migrationshintergrund (Heuft et al., 1998).

3.2.3.4. Beck-Depressions-Inventar (BDI-II)

Der BDI-II (Beck et al., 1996; Kühner et al., 2007) ist ein Selbstbeurteilungs-Fragebogen mit 21 vierstufigen Items, welcher zur Einschätzung der Schwere einer Depression

herangezogen wird. Die Bearbeitungsdauer beträgt etwa 5-10 Minuten (Kühner et al., 2007). Jedes der Items beschreibt ein mögliches Symptom einer Depression (u.a. Traurigkeit, Interessensverlust und Veränderung der Schlafgewohnheiten), für welches vier Aussagen mit ansteigendem Schweregrad und Punktwerten vorgegeben sind. Es muss aus den vorgegebenen Aussagen jeweils diejenige ausgewählt werden, die am ehesten dem eigenen Zustand in den vergangenen zwei Wochen entspricht (Beispielitem Traurigkeit: 0 Punkte – Ich bin nicht traurig; 1 Punkt – Ich bin oft traurig; 2 Punkte – Ich bin ständig traurig; 3 Punkte – Ich bin so traurig oder unglücklich, dass ich es nicht aushalte). Zur Auswertung werden die erreichten Punkte addiert und mit Cut-Off Werten abgeglichen. Dabei gilt laut Kühner et al. (2007):

- 0-13 Punkte: Keine oder minimale Depression
- 14-19 Punkte: Leichte Depression
- 20-28 Punkte: Mittelschwere Depression
- 29-63 Punkte: Schwere Depression

3.2.3.5. Childhood Trauma Questionnaire (CTQ)

Der CTQ ist ein Fragebogen zur Erfassung von Missbrauch und Vernachlässigung in der Kindheit und besteht aus fünf Unterskalen mit je fünf Items (emotionale Misshandlung, körperliche Misshandlung, emotionale Vernachlässigung, körperliche Vernachlässigung, sexueller Missbrauch) und einer Bagatellisierungsskala mit drei Items (Bernstein et al., 1994). Auf jeder Skala können 5 (keine Misshandlung/Vernachlässigung/sexueller Missbrauch) bis 25 Punkte (extreme Misshandlung/Vernachlässigung/sexueller Missbrauch) erreicht werden. Für diese Studie wurde der CTQ-Fragebogen mit 28 Items verwendet (Bernstein et al., 2003). Von einem Trauma im Sinne des Childhood Trauma Questionnaires kann ab einem Score von „mäßig bis schwer“ auf einer der Skalen ausgegangen werden (Häuser et al., 2011).

3.2.3.6. Toronto Alexithymie Skala (TAS-20)

Zur Überprüfung einer Alexithymie bei den Probanden wurde die Toronto Alexithymie Skala mit 20 Items (TAS-20) eingesetzt (Bagby et al., 1994). Als Alexithymie bezeichnet man die Unfähigkeit, eigene Gefühle wahrzunehmen und diese in Worte zu fassen (Nemiah & Sifneos, 1970). Studien zeigen, dass es Menschen mit einer Alexithymie im Vergleich zu Kontrollgruppen schwerer fällt, Gesichtsausdrücke zu deuten und in Worten zu beschreiben (Grynberg et al., 2012), weshalb sie auch für die Fragestellungen des MET von Bedeutung sein könnte. Die Toronto Alexithymie Skala ist ein Fragebogen, welcher drei Subskalen enthält: Schwierigkeiten bei der Identifikation von Gefühlen (Beispiel: „Mir ist oft unklar, welche Gefühle ich gerade habe.“), Schwierigkeiten Gefühle zu beschreiben (Beispiel: „Es fällt mir schwer, die richtigen Worte für meine Gefühle zu finden.“) und extern orientiertes Denken (Beispiel: „Ich unterhalte mich mit anderen nicht so gern über ihre Gefühle, sondern lieber darüber, womit sie sich täglich beschäftigen.“) (Haviland, 1996; Popp et al., 2008). Antworten mussten durch die Teilnehmer jeweils auf einer fünfstufigen Skala gegeben werden (1 = trifft gar nicht zu bis 5 = trifft völlig zu). Der Cut-Off-Wert für die Feststellung von Alexithymie betrug ≥ 61 (Bagby et al., 2006; Parker et al., 1993).

3.2.3.7. Zahlennachsprechen

Weiterhin wurden die Probanden gebeten, einen Untertest des Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE) (Wechsler, 1956), das Zahlennachsprechen, durchzuführen. Die Teilnehmer wurden gebeten, Zahlenfolgen mit ansteigendem Schwierigkeitsgrad (Zunahme der Zahlenanzahl pro Zahlenfolge) nachzusprechen (Zahlenspannen vorwärts). In einem zweiten Durchgang sollten Zahlenfolgen in umgekehrter Reihenfolge nachgesprochen werden (Zahlenspannen rückwärts). Pro Schwierigkeitsstufe wurden zwei Versuche mit unterschiedlichen Zahlenfolgen durchgeführt. Traten bei einer Schwierigkeitsstufe bei beiden Versuchen Fehler auf, wurde der Durchgang abgebrochen. Insgesamt konnten in beiden Durchgängen (Zahlenspannen vorwärts/rückwärts) jeweils 12 Punkte erreicht werden. Die Auswertung erfolgte unter Berücksichtigung des Teilnehmeralters.

3.2.3.8. Regensburger Wortflüssigkeitstest

Zur Erfassung der verbalen Flüssigkeit dienten Teile des Regensburger Wortflüssigkeitstests (Aschenbrenner et al., 2000). Hierbei wurden die Probanden gebeten, in einer Minute zunächst so viele Tiere wie möglich und im Anschluss so viele Wörter mit dem Anfangsbuchstaben S wie möglich zu nennen. In die Auswertung gingen die Anzahl der genannten Begriffe sowie das Alter der Teilnehmer ein.

3.2.3.9. Trail-Making-Test (TMT)

Der Trail-Making-Test ist ein Test zur allgemeinen Überprüfung von Hirnfunktionsleistungen wie Arbeitsgedächtnis, Aufmerksamkeit, visuomotorischer Verarbeitungsgeschwindigkeit, kognitive Flexibilität und Exekutivfunktionen (Reitan, 1992; Tischler & Petermann, 2010). Die Bearbeitungszeit beim Trail-Making-Test beträgt je nach Funktionsniveau fünf bis zehn Minuten. Die Probanden mussten hierbei jeweils 25 Kreise verbinden, welche Zahlen (Test A) bzw. Zahlen und Buchstaben (Test B) enthalten. Ziel war es, die Zahlen in aufsteigender Reihenfolge von 1 bis 25 (Test A) sowie die Zahlen in aufsteigender Reihenfolge von 1 bis 13 abwechselnd mit den Buchstaben in alphabetischer Reihenfolge von A bis L (Test B) in möglichst kurzer Zeit mit Bleistift zu verbinden, ohne dabei den Stift vom Papier abzusetzen. Die Auswertung erfolgte anschließend anhand der vom Probanden benötigten Bearbeitungszeit in Sekunden. Wurden während der Durchführung Fehler gemacht, wurden die Probanden umgehend auf diese hingewiesen. Fehler flossen nicht direkt in die Auswertung ein, führten jedoch zu einer verlängerten Bearbeitungszeit. Vor jedem der beiden Testteile A und B erfolgte ein Übungsbeispiel. Die Auswertung erfolgte über CERAD-Plus Online auf der Website der Memory Clinic Basel (siehe <http://www.memoryclinic.ch>).

3.2.3.10. Mehrfach-Wortschatz-Intelligenztest (MWT-B)

Die Parallelversion B des Mehrfachwahl-Wortschatz-Intelligenztests (MWT-B) diente zur Messung des allgemeinen Intelligenzniveaus und besteht aus 37 Items mit aufsteigendem Schwierigkeitsgrad (Lehrl, 1999). Ein Item besteht hierbei aus einer Reihe mit jeweils fünf Wörtern, wobei nur eines der präsentierten Wörter tatsächlich existiert, während die anderen fiktive Neukonstruktionen darstellen (Beispiel: Tinxur – Kukukur – Fraktan – Tinkur – Rimsuhr). Die Probanden wurden aufgefordert, nach dem Mehrfachwahl-Prinzip das ihnen bekannte Wort in jeder Reihe durchzustreichen. Für jede Reihe, in der das „richtige Wort“ durchgestrichen wurde, erhielten die Probanden einen Punkt, wobei maximal 37 Punkte erreicht werden konnten. Die Gesamtpunktzahl wurde mit der Leistung einer repräsentativen Stichprobe deutschsprachiger Erwachsener (n = 1952) verglichen, woraus sich unter anderem der IQ der Studienteilnehmer ermitteln ließ (ebd.). Der Mehrfach-Wortschatz-Intelligenztest ist hierfür ein gut geeignetes Messinstrument, da er zuverlässige Werte in kurzer Zeit (Testdauer etwa fünf Minuten) erbringt (Lehrl et al., 1995).

3.2.3.11. Verbaler Lern- und Merkfähigkeitstest (VLMT)

Mit dem Verbalen Lern- und Merkfähigkeitstest (VLMT) wurden unterschiedliche Parameter des episodischen Gedächtnisses der Probanden erfasst (Helmstaedter & Durwen, 1990). Der Verbale Lern- und Merkfähigkeitstest ist ein Listenlernetest mit mehreren Lerndurchgängen einer Lernliste sowie einer Interferenzliste mit jeweils 15 Wörtern. **Abbildung 9** zeigt zusammengefasst den Ablauf des VLMT.

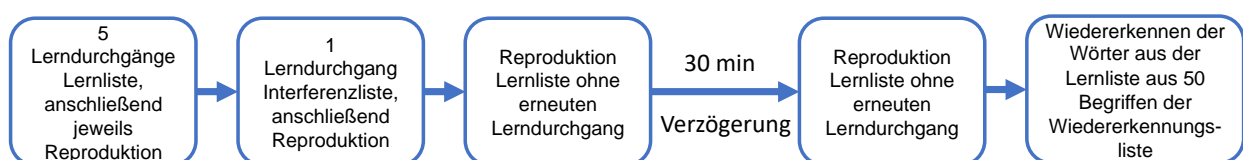


Abbildung 9: Ablauf des Verbalen Lern- und Merkfähigkeitstest (VLMT; Helmstaedter & Durwen, 1990).

In den ersten fünf Lerndurchgängen wurde durch die Doktorandin die 15 Begriffe der Lernliste vorgelesen, woraufhin diese durch die Studienteilnehmer nach jedem Durchgang reproduziert werden sollten. Im sechsten Lerndurchgang wurde die Interferenzliste mit 15 neuen Begriffen vorgelesen. Auch diese sollte im Anschluss reproduziert werden. Daraufhin wurden die Probanden aufgefordert, erneut die Lernliste zu reproduzieren. Es folgten 30 Minuten Verzögerung, in denen beispielsweise Fragebögen wie die Psychotherapie Basisdokumentation (siehe **3.2.3.3.**) ausgefüllt werden konnten. Anschließend wurden die Teilnehmer ein weiteres Mal nach den Begriffen der Lernliste gefragt. Zuletzt wurden 50 Begriffe einer Wiedererkennungsliste vorgelesen, welche sowohl die Begriffe der Lern- als auch der Interferenzliste enthielt. Die Teilnehmer sollten bei jedem Begriff angeben, ob dieser in der Lernliste vorhanden war („Ja“) oder nicht („Nein“) und somit nur die Begriffe der Lernliste wiedererkennen. Die Anzahl der reproduzierten und wiedererkannten Begriffe wurde in ein Protokoll eingetragen und schließlich getrennt nach Alter der Probanden bezüglich Supraspanne, Lernleistung, Wiedererkennungsleistung und anderen Parametern ausgewertet.

3.3. Datenauswertung

Die Berechnung der Itemschwierigkeit erfolgte anhand oben genannter Formel mit Microsoft Office Excel 2019 für macOS. Für die Erstellung der Abbildungen wurde Microsoft Office Power Point 2019 für macOS verwendet. Für die graphische Darstellung der Ergebnisse wurde GraphPad Prism Version 8.2.1 für MacOS verwendet.

Die Datenauswertung erfolgte mit IBM SPSS Statistics Version 25 für MacOS. Die Trennschärfe wurde hierbei mittels Reliabilitätsanalyse als Korrigierte-Item-Skala-Korrelation berechnet. Dabei wurden Trennschärfen $\geq 0,30$ als akzeptabel gewertet. Es wurden Summenwerte für die richtigen Antworten der kognitiven Empathie, Mittelwerte für die emotionale Empathie, sowie Mittelwerte der Antwortlatenzen in den jeweiligen Versionen und für jeden Probanden gebildet. Alle Daten wurden mithilfe des Shapiro-Wilk-Tests, sowie eines Q-Q-Plots auf Normalverteilung getestet. Aus den Summen- und Mittelwerten wurde zum Vergleich der Mittelwerte bei normalverteilten Daten ein t-Test bei gepaarten Stichproben durchgeführt. Waren die Daten nicht normalverteilt wurde der Wilcoxon-Test

für abhängige Stichproben verwendet. Zusätzlich erfolgte bei normalverteilten Daten die Berechnung des Korrelationskoeffizienten r nach Pearson, wobei $r \geq 0,3$ als kleiner, $r = 0,5$ bis $0,79$ als mittlerer und $r \geq 0,8$ als großer Effekt betrachtet wurde. Bei nicht normal verteilten Daten wurde eine Rangkorrelationsanalyse nach Spearman durchgeführt. Zur Berechnung der Effektstärke wurde Cohens d verwendet (Cohen, 1988). Hierbei wurde $d \geq 0,2$ als kleiner Effekt, $d = 0,5$ bis $0,79$ als mittlerer Effekt und $d \geq 0,8$ als großer Effekt betrachtet. Die Überprüfung der Reliabilität erfolgte mit Cronbachs Alpha als interne Konsistenz, wobei interne Konsistenzen $\geq 0,70$ als akzeptabel betrachtet wurden. Das Signifikanzniveau wurde festgelegt auf $p < 0,05$.

3.4. Anpassung der Items nach erster Erprobung

Da die Ergebnisse nach der Erprobung im Rahmen der ersten Stichprobe darauf hindeuteten, dass es sich nicht um parallele Testversionen handelt, war es notwendig, die beiden Versionen des MET im Hinblick auf die gewünschte Vergleichbarkeit noch einmal anzupassen. Hierfür wurden die Schwierigkeiten aller Items in der ersten Stichprobe verglichen und anhand dieser insgesamt sechs Items zwischen den MET-Versionen A und B ausgetauscht, so dass die mittlere Schwierigkeit und Trennschärfe innerhalb der Blöcke möglichst ähnlich waren.

Es musste folglich auch eine Anpassung der MET-Programme durch Matthias Kupferschmidt statt, sodass die erneute Erprobung und Auswertung der Daten an einer zweiten Stichprobe stattfinden konnte. Die übrigen Schritte waren dieselben wie bereits unter Punkt **3.2.** bis **3.3.** ausführlich beschrieben.

4. Ergebnisse

Das folgende Kapitel stellt die Ergebnisse der erhobenen Daten getrennt nach den beiden Stichproben, sowie im zeitlichen Ablauf der Erstellung der Parallelversionen dar.

4.1. Ergebnisse der ersten Stichprobe

4.1.1. Erstellung der Parallelversionen

Die Bildung der ersten Parallelversionen erfolgte anhand von Schwierigkeit und Trennschärfe aus den Daten der Stichprobe des NeMUP-Projektes sowie der Validierung des MET-KE durch Benjamin Lemme (Lemme, 2012). Dabei ergaben sich für die Stichprobe des NeMUP-Projektes Schwierigkeiten von 22,86 bis 97,06 (MW 62,04 \pm 19,24) und Trennschärfen für den kognitiven Empathieteil von -0,001 bis 0,390 (MW 0,191 \pm 0,102) und für den emotionalen Empathieteil von 0,540 bis 0,751 (MW 0,639 \pm 0,046). Für die „Lemme-Stichprobe“ ergaben sich Schwierigkeiten zwischen 10,26 und 94,87 (MW 66,06 \pm 17,32) und Trennschärfen für die kognitive Empathie zwischen -0,118 und 0,813 (MW 0,357 \pm 0,160). Für die einzelnen Werte der Items siehe **Anhang 2**. Den Aufbau der finalen Itemblöcke der MET-Versionen A und B sowie die Mittelwerte der Schwierigkeiten in den einzelnen Blöcken zeigt **Tabelle 2**.

Tabelle 2: Itemblöcke von MET-A und MET-B.

MET-Version	A	B
Block 1 a/b		
Item Nr.	1, 9, 13, 15, 26, 31, 50, 63, 72, 78	2, 3, 22, 27, 33, 42, 43, 58, 66, 74
P _{NeMUP} , MW ± SD	65,95 ± 20,81	66,28 ± 16,67
P _{Lemme} , MW ± SD	66,41 ± 24,24	68,97 ± 16,01
Block 2 a/b		
Item Nr.	5, 11, 29, 39, 46, 51, 52, 64, 71, 76	17, 20, 24, 25, 35, 45, 59, 60, 68, 80
P _{NeMUP} , MW ± SD	60,51 ± 23,70	56,88 ± 20,42
P _{Lemme} , MW ± SD	68,21 ± 14,81	60,26 ± 17,69
Block 3 a/b		
Item Nr.	7, 10, 21, 36, 37, 41, 49, 61, 62, 77	8, 14, 23, 32, 47, 48, 56, 57, 67, 73
P _{NeMUP} , MW ± SD	66,46 ± 17,55	65,25 ± 21,61
P _{Lemme} , MW ± SD	69,74 ± 14,69	70,26 ± 19,68
Block 4 a/b		
Item Nr.	4, 12, 28, 40, 44, 54, 65, 70, 75, 79	6, 16, 18, 19, 30, 34, 38, 53, 55, 69
P _{NeMUP} , MW ± SD	56,80 ± 13,80	55,26 ± 19,39
P _{Lemme} , MW ± SD	60,51 ± 15,63	63,08 ± 16,67

MW = Mittelwert, SD = Standardabweichung, Block a = kognitive Empathie, Block b = emotionale Empathie, P_{NeMUP} = Schwierigkeit in der NeMUP-Stichprobe, P_{Lemme} = Schwierigkeit in der Stichprobe von Lemme (2012), MET = Multidimensionaler Empathietest, MET-A = Multidimensionaler Empathietest Parallelversion A, MET-B = Multidimensionaler Empathietest Parallelversion B.

Tabelle 3 zeigt eine Übersicht der Mittelwerte von Schwierigkeit und Trennschärfe der jeweiligen MET-Version (A bzw. B). Bei der Validierung des MET-KE wurden die Items lediglich für den Bereich der kognitiven Empathie analysiert, weshalb für die emotionale Komponente keine Daten vorhanden sind.

Tabelle 3: Übersicht von Schwierigkeit und Trennschärfe von MET-A und MET-B.

MET-Version	A	B
Schwierigkeit		
P_{NeMUP} , MW \pm SD	63,17 \pm 19,18	60,92 \pm 19,48
P_{Lemme} , MW \pm SD	66,47 \pm 17,47	65,64 \pm 17,39
Trennschärfe		
Kognitive Empathie		
T_{KN} , MW \pm SD	0,212 \pm 0,104	0,170 \pm 0,096
T_{Lemme} , MW \pm SD	0,357 \pm 0,146	0,357 \pm 0,173
Trennschärfe		
Emotionale Empathie		
T_{EN} , MW \pm SD	0,635 \pm 0,051	0,643 \pm 0,042
T_{EL} , MW \pm SD	Keine Daten vorhanden	Keine Daten vorhanden

MW = Mittelwert, SD = Standardabweichung, P_{NeMUP} = Schwierigkeit in der NeMUP-Stichprobe, P_{Lemme} = Schwierigkeit in der Stichprobe von Lemme (2012), T_{KN} = Trennschärfe kognitive Empathie in der NeMUP-Stichprobe, T_{Lemme} = Trennschärfe der kognitiven Empathie in der Stichprobe von Lemme (2012), T_{EN} = Trennschärfe der emotionalen Empathie in der NeMUP-Stichprobe, T_{EL} = Trennschärfe der emotionalen Empathie in der Stichprobe von Lemme (2012), MET = Multidimensionaler Empathietest, MET-A = Multidimensionaler Empathietest Parallelversion A, MET-B = Multidimensionaler Empathietest Parallelversion B.

In der Tabelle ist zu sehen, dass die beiden Versionen sich bezüglich der Gesamtschwierigkeit kaum unterscheiden. Es wurden t-Tests zum Vergleich der Mittelwerte durchgeführt. Dabei zeigten sich für die NeMUP-Stichprobe ($t(40) = 2,25$, $p = 0,475$, $d = 0,11$) und für die Lemme-Stichprobe ($t(40) = 0,83$, $p = 0,594$, $d = 0,08$) keine signifikanten Unterschiede in der Schwierigkeit der Items. Bezüglich der Trennschärfe zeigte sich in der Lemme-Stichprobe kein signifikanter Unterschied zwischen den Versionen ($t(40) = -0,001$, $p = 0,887$, $d = 0,02$). Ebenfalls zeigte sich kein signifikanter Unterschied im Hinblick auf die Trennschärfe der emotionalen Empathie ($t(40) = -0,01$, $p = 0,474$, $d = 0,11$). Lediglich bei der Trennschärfe der kognitiven Empathie in der NeMUP-Stichprobe war ein signifikanter Unterschied festzustellen ($t(40) = 0,04$, $p = 0,036$, $d = 0,34$).

Obwohl die Merkmale der Items bezüglich Geschlecht, Alter und Emotionsvalenz innerhalb der beiden Testversionen möglichst gleich verteilt sein sollten, konnte dies aufgrund

der unterschiedlichen Schwierigkeiten und Trennschärfen der Items und daraus resultierender Itempaar-Bildung nicht ausnahmslos erfolgen. Da ähnliche Kennwerte der Items gegenüber der völligen Übereinstimmung bezüglich der Merkmale bevorzugt wurden, ergab sich letztlich die in **Tabelle 4** dargestellte Verteilung der Items bezüglich ihrer Merkmale.

Tabelle 4: Itemmerkmale von MET-A und MET-B.

MET-Version	A	B
Geschlecht, Anzahl der Items (%)		
Weiblich	21 (52,50 %)	19 (47,50 %)
Männlich	19 (47,50 %)	21 (52,50 %)
Alter, Anzahl der Items (%)		
Erwachsener	18 (45 %)	22 (55 %)
Kind	22 (55 %)	18 (45 %)
Emotionsvalenz (%)		
Positiv	20 (50 %)	20 (50 %)
Negativ	20 (50 %)	20 (50 %)

MET = Multidimensionaler Empathietest, MET-A = Multidimensionaler Empathietest Parallelversion A, MET-B = Multidimensionaler Empathietest Parallelversion B.

4.1.2. Studienteilnehmer

4.1.2.1. Demographische Daten

Für die erste Stichprobe wurden $n = 21$ Probanden rekrutiert, wobei ein Proband aufgrund eines HDRS-24 Wertes von > 5 aus der Studie ausgeschlossen wurde. Von den 20 eingeschlossenen Studienteilnehmern waren 15 weiblich (75 %) und 5 männlich (25%). Die Teilnehmer hatten ein mittleres Alter von 41,5 Jahren, wobei der jüngste Teilnehmer 24 und der älteste 59 Jahre alt war. Der Großteil der Studienteilnehmer lebte in einer festen Partnerschaft, wobei etwa die Hälfte mit diesem Partner verheiratet war. Die Mehrheit der Teilnehmer hatte einen hohen Bildungsgrad und war voll berufstätig (siehe **Tabelle 5**).

Tabelle 5: Demographische Daten von Stichprobe 1.

Frauen/Männer, Anzahl (%)	10 (75 %) / 5 (25 %)
Alter (Jahre), MW ± SD	41,5 ± 12,7
Familienstand, Anzahl (%)	
Ledig	10 (50 %)
Verheiratet	9 (45 %)
Geschieden	1 (5 %)
Partnersituation, Anzahl (%)	
Kein Partner	4 (20 %)
Wechselnde Partner	4 (20 %)
Fester Partner (kein Ehepartner)	5 (25 %)
Fester Partner (Ehepartner)	7 (35 %)
Schulabschluss, Anzahl (%)	
Mittlere Reife/Realschulabschluss	2 (10 %)
Abitur/Fachabitur	18 (90 %)
Berufsabschluss, Anzahl (%)	
Lehre/Fachschule	4 (20 %)
Meister	1 (5 %)
Fachhochschule/Universität	14 (70 %)
Ohne Berufsabschluss	1 (5 %)
Aktuelle berufliche Situation, Anzahl (%)	
Berufstätig, Vollzeit	13 (65 %)
Berufstätig, Teilzeit	4 (20 %)
Hausfrau/-mann, nicht berufstätig	1 (5 %)
Ausbildung	1 (5 %)
Anderweitig ohne Beschäftigung	1 (5 %)

MW = Mittelwert, SD = Standardabweichung.

Ein Studienteilnehmer war muttersprachlich Russisch, lebte jedoch seit 30 Jahren in Deutschland. Die restlichen Probanden waren deutsche Muttersprachler.

4.1.2.2. Klinische und neuropsychologische Daten

Das Mini-International Neuropsychiatric Interview (M.I.N.I.) ergab bei zwei Probanden eine Panikstörung in der Vorgeschichte, welche zum Untersuchungszeitpunkt nicht mehr bestand. Ein Teilnehmer gab an, in der Vergangenheit unter einer einmaligen psychotischen Episode gelitten zu haben. Im Hinblick auf eine mögliche depressive Symptomatik ergaben sich in der HDRS-24 entsprechend der Einschlusskriterien bei keinem Teilnehmer Hinweise auf eine klinisch relevante depressive Symptomatik. Auch laut BDI-II bestand bei keinem der Teilnehmer eine depressive Symptomatik. Im CTQ ergab sich für einen Teilnehmer definitionsgemäß eine traumatische Kindheitserfahrung durch sexuellen Missbrauch. In der Toronto Alexithymie Skala lag ein Teilnehmer mit einem Wert von 65 Punkten über dem Cut-Off Wert (≥ 61) für die Feststellung einer Alexithymie. Da die Feststellung einer Alexithymie grundsätzlich mit den Ein- und Ausschlusskriterien vereinbar war und für diese Arbeit die genannte Stichprobe nicht als Kontrollgruppe fungierte, wurde der Teilnehmer nicht aus der Studie ausgeschlossen.

In den neuropsychologischen Tests schnitten die Probanden im Vergleich zu Normstichproben insgesamt überdurchschnittlich gut ab. Eine Übersicht der Ergebnisse der klinischen und neuropsychologischen Diagnostik gibt **Tabelle 6**.

Tabelle 6: Übersicht klinische und neuropsychologische Diagnostik in Stichprobe 1.

HDRS-24, MW ± SD	0,85 ± 1,42
BDI-II, MW ± SD	2,55 ± 2,69
CTQ, MW ± SD	
Emotionale Misshandlung	6,15 ± 1,39
Körperliche Misshandlung	5,25 ± 0,64
Emotionale Vernachlässigung	7,30 ± 2,47
Körperliche Vernachlässigung	5,85 ± 1,50
Sexueller Missbrauch	5,15 ± 0,67
TAS, MW ± SD	36,05 ± 10,39
Zahlennachsprechen, MW ± SD	
Prozentrang vorwärts	73,16 ± 30,04
Prozentrang rückwärts	60,05 ± 34,99
Regensburger Wortflüssigkeitstest, MW ± SD	
Prozentrang Tiere	62,55 ± 30,49
Prozentrang S-Wörter	51,50 ± 31,09
Trail Making Test, MW ± SD	
Zeit in s, Teil A	29,90 ± 12,31
Zeit in s, Teil B	55,80 ± 19,60
z-Wert, Teil A	0,49 ± 1,34
z-Wert, Teil B	0,83 ± 1,12
MWT-B (IQ), MW ± SD	114,00 ± 11,64
VLMT (Rohwert), MW ± SD	
Supraspanne	7,90 ± 1,74
Lernleistung	13,55 ± 1,63
Abruf nach Interferenz	12,45 ± 1,93
Abruf nach Verzögerung	12,85 ± 2,16
Wiedererkennungslleistung	14,25 ± 1,02

MW = Mittelwert, SD = Standardabweichung, HDRS-24 = Hamilton Depression Rating Scale 24-Item Version, BDI II = Beck-Depressions-Inventar II, CTQ = Childhood Trauma Questionnaire, TAS = Toronto Alexithymie Skala, MWT-B = Mehrfach-Wortschatz-Intelligenztest Version B, VLMT = Verbaler Lern- und Merkfähigkeitstest.

4.1.3. Statistische Auswertung

11 Teilnehmer (55 %) bearbeiteten zuerst MET-A, 9 Teilnehmer (45 %) zuerst MET-B. Zwischen der Bearbeitung der beiden Versionen lagen 1 bis 14 Tage (MW $3,90 \pm 3,75$), wobei 16 Teilnehmer (80 %) die beiden Testversionen innerhalb einer Woche bearbeiteten.

4.1.3.1. Korrelation

MET-A und MET-B korrelierten bezüglich der Bewertungen der emotionalen Empathie (Pearson-Korrelation, $r = 0,96$, $p < 0,001$) sowie der Antwortlatenzen der emotionalen (Spearman Korrelation, $r = 0,59$, $p = 0,006$) und kognitiven (Spearman Korrelation, $r = 0,82$, $p < 0,001$) Empathie im mittleren bis hohen Bereich. Bezüglich der Summe der richtigen Antworten (= kognitive Empathie) gab es keine signifikante Korrelation zwischen MET-A und MET-B (Pearson Korrelation, $r = 0,36$, $p = 0,120$).

4.1.3.2. Vergleich der Mittelwerte

Beim Vergleich der Mittelwerte zeigten sich im t-Test sowohl für die kognitive ($t(20) = 1,75$, $p = 0,048$, $d = 0,47$) als auch für die emotionale Empathie ($t(20) = -0,46$, $p = 0,002$, $d = 0,78$) signifikante Unterschiede zwischen den Versionen MET-A und MET-B. Bei den Mittelwerten der Antwortlatenzen der kognitiven und der emotionalen Empathie zeigte sich hingegen kein signifikanter Unterschied zwischen MET-A und MET-B im Wilcoxon-Test (kognitiv: $p = 0,411$, $r = 0,18$; emotional: $p = 0,455$, $r = 0,17$). Siehe hierzu auch **Abbildung 10**.

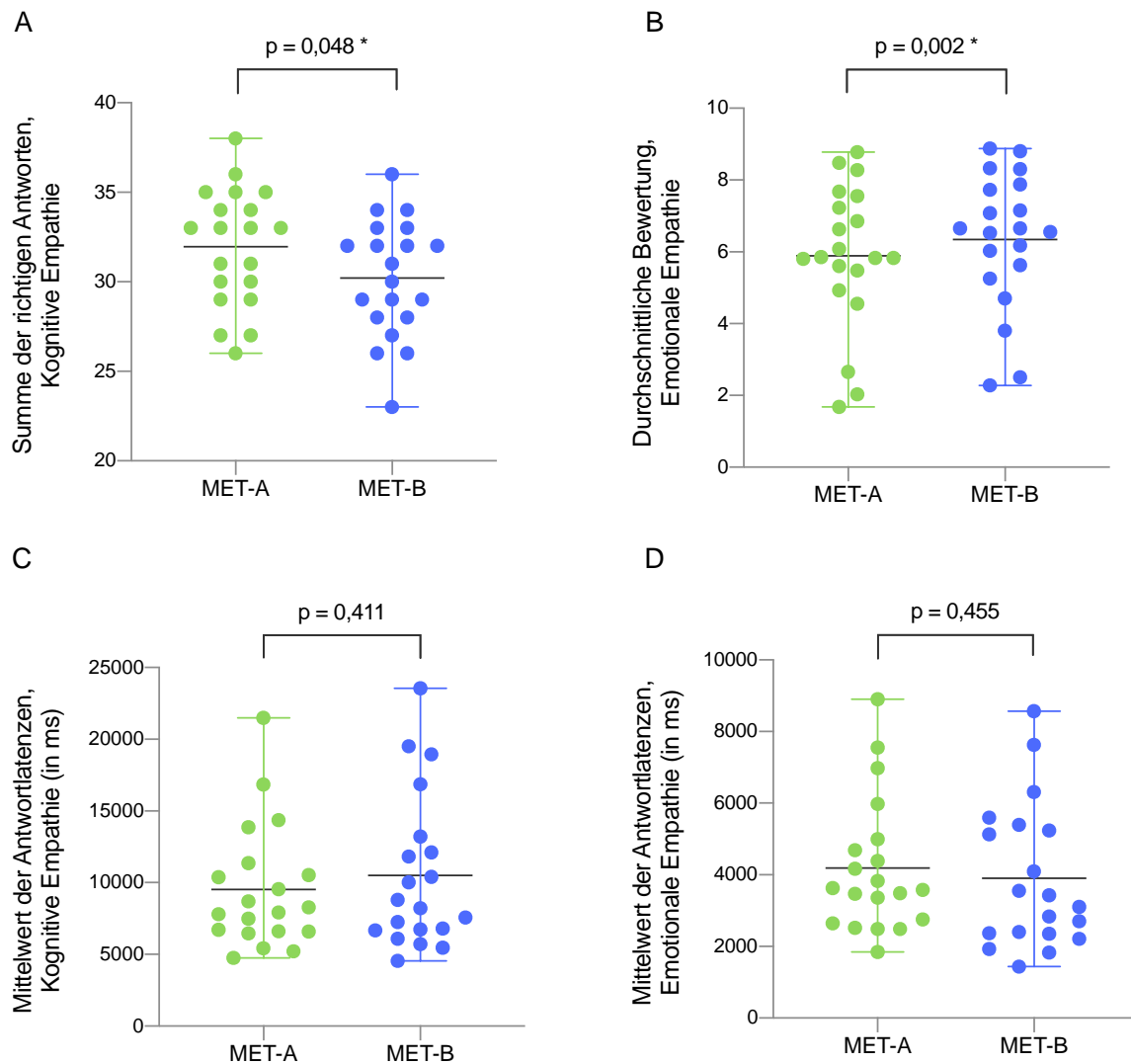


Abbildung 10: Vergleich der Mittelwerte in Stichprobe 1.

- A:** Summe der richtigen Antworten der kognitiven Empathie in MET-A und MET-B.
 - B:** Durchschnittliche Bewertung der emotionalen Empathie in MET-A und MET-B.
 - C:** Mittelwert der Antwortlatenzen der kognitiven Empathie in MET-A und MET-B.
 - D:** Mittelwert der Antwortlatenzen der emotionalen Empathie in MET-A und MET-B.
- MET-A = Multidimensionaler Empathietest Parallelversion A, MET-B = Multidimensionaler Empathietest Parallelversion B. p = Signifikanzwert.

Eine Übersicht über die Mittelwerte mit Standardabweichungen und p -Werten gibt **Tabelle 7**.

Tabelle 7: Vergleich der Mittelwerte in Stichprobe 1.

MET-Version	A	B	p-Wert
Kognitive Empathie			
Summe der richtigen Antworten, MW ± SD	31,95 ± 3,28	30,20 ± 3,27	0,048*
Antwortlatenz (in ms), MW ± SD	9520,93 ± 4279,16	10518,24 ± 5368,26	0,411
Emotionale Empathie			
Durchschnittliche Bewertung, MW ± SD	5,89 ± 2,00	6,34 ± 1,89	0,002*
Antwortlatenz (in ms), MW ± SD	4188,01 ± 1871,69	3905,72 ± 2019,65	0,455

MW = Mittelwert, SD = Standardabweichung, p-Wert = Signifikanzwert, MET = Multidimensionaler Empathietest.

4.1.3.3. Schwierigkeit

Die Itemschwierigkeit lag sowohl im MET-A als auch im MET-B bei $P = 33,33$ bis $P = 100$. In der ersten Stichprobe gab es in beiden Versionen jeweils fünf Items, welche alle Teilnehmer richtig lösten ($P = 100$). Im Mittel lag die Schwierigkeit in MET-A bei $73,17 (\pm 18,46)$, in MET-B bei $67,33 (\pm 23,18)$. **Abbildung 11** zeigt die Häufigkeitsverteilung der Schwierigkeiten in den beiden MET-Versionen. Hierbei ist zu erkennen, dass einige Items wie gewünscht mittlere Schwierigkeiten aufweisen, insbesondere in Version MET-B. Jedoch gibt es auch Items, die Schwierigkeiten um $P = 100$ aufweisen und damit nicht gut zwischen hoher und niedriger Merkmalsausprägung differenzieren können.

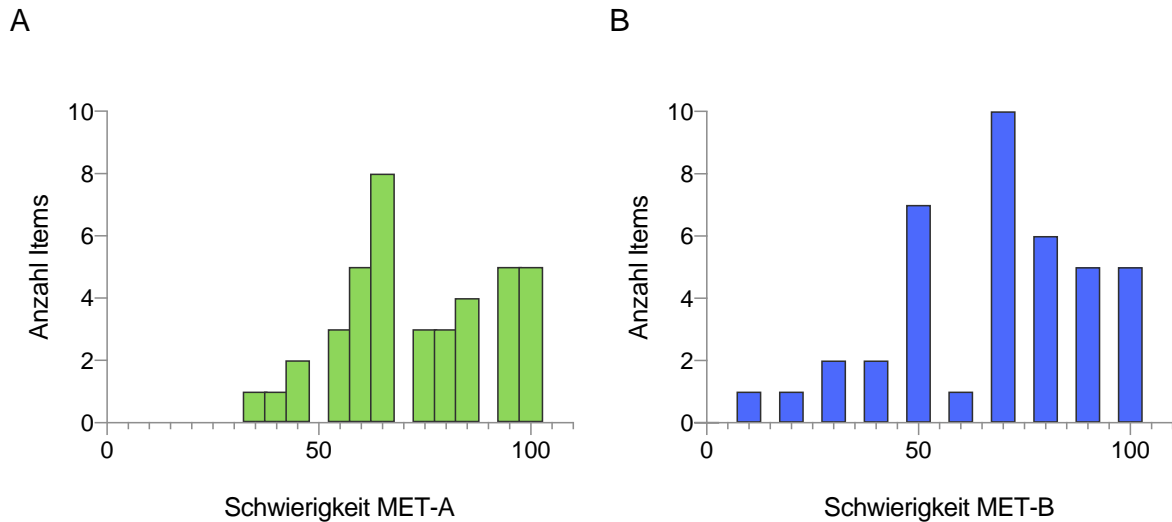


Abbildung 11: Häufigkeitsverteilung der Itemschwierigkeiten in Stichprobe 1.

A: Itemschwierigkeit in MET-A.

B: Itemschwierigkeit in MET-B.

MET-A = Multidimensionaler Empathietest Parallelversion A, MET-B = Multidimensionaler Empathietest Parallelversion B.

4.1.3.4. Trennschärfe

Die Trennschärfekoeffizienten für die kognitive Empathie von MET-A lagen bei $-0,380$ bis $0,540$ (MW $0,099 \pm 0,197$). Dabei lag die Trennschärfe von 16 Items bei ≤ 0 . Nur 7 Items (17,50 %) lagen über $0,30$, was als zufriedenstellender Wert angesehen wurde. Die Trennschärfekoeffizienten für die kognitive Empathie von MET-B lagen bei $-0,391$ bis $0,478$ (MW $0,078 \pm 0,216$). Dabei lag die Trennschärfe von 19 Items bei ≤ 0 . Auch hier lagen nur 7 Items (17,50 %) über $0,30$. Die Trennschärfekoeffizienten für die emotionale Empathie von MET-A lagen bei $0,617$ bis $0,946$ (MW $0,844 \pm 0,060$), die für MET-B bei $0,711$ bis $0,931$ (MW $0,837 \pm 0,057$). Damit wiesen beide Testhälften sehr gute Trennschärfen im Bereich der emotionalen Empathie auf (Kelava & Moosbrugger, 2012).

Abbildung 12 zeigt die Häufigkeitsverteilung der Trennschärfekoeffizienten in den beiden MET-Versionen. Bezüglich der Trennschärfe der kognitiven Empathie weisen in beiden Versionen viele Items Trennschärfen um und unter 0 auf und sind damit nicht gut geeignet, um Probanden mit hoher und niedriger Merkmalsausprägung zu differenzieren (Kelava & Moosbrugger, 2012). Die Trennschärfen der emotionalen Empathie weisen durchweg hohe und damit gute Werte auf.

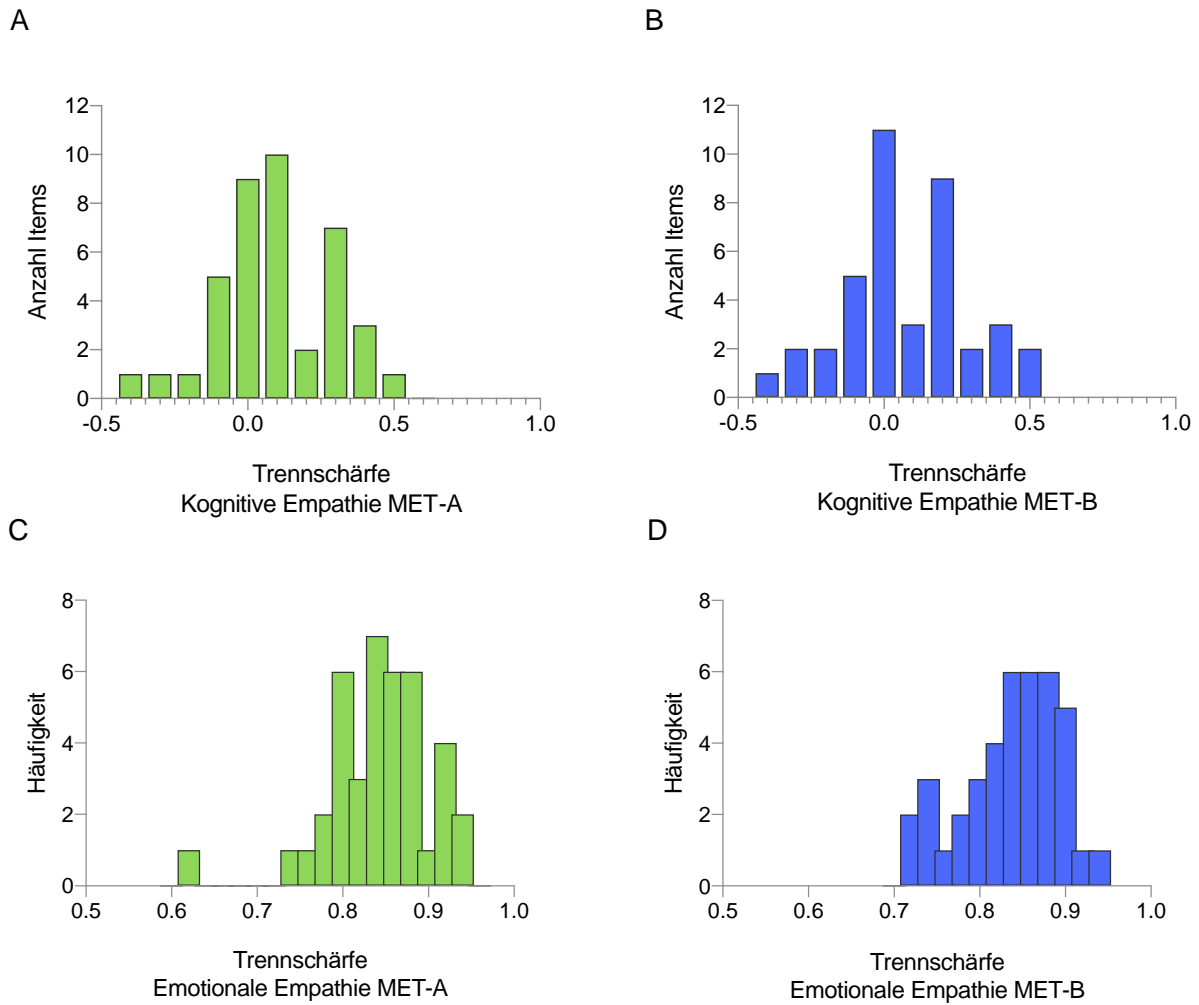


Abbildung 12: Häufigkeitsverteilung der Trennschärfekoeffizienten in Stichprobe 1.

- A:** Trennschärfe der kognitiven Empathie in MET-A.
 - B:** Trennschärfe der kognitiven Empathie in MET-B.
 - C:** Trennschärfe der emotionalen Empathie in MET-A.
 - D:** Trennschärfe der emotionalen Empathie in MET-B.
- MET-A = Multidimensionaler Empathietest Parallelversion A, MET-B = Multidimensionaler Empathietest Parallelversion B.

Zu den einzelnen Werten der Schwierigkeitsindizes und Trennschärfekoeffizienten in Stichprobe 1 siehe **Tabelle 8A** und **8B**.

ERGEBNISSE

Tabelle 8A: Itemkennwerte von MET-A.

Item	P	T _K	T _E	Item	P	T _K	T _E
1	66,67	0,264	0,806	44	53,33	0,427	0,824
4	60,00	0,382	0,749	46	46,67	0,193	0,805
5	66,67	-0,108	0,946	49	80,00	0,381	0,879
7	86,67	-0,098	0,806	50	100	0,000	0,844
9	93,33	-0,142	0,807	51	66,67	0,149	0,889
10	100	0,000	0,863	52	66,67	0,187	0,848
11	66,67	-0,380	0,881	54	60,00	0,017	0,831
12	93,33	0,145	0,905	61	93,33	0,293	0,848
13	46,67	0,092	0,795	62	60,00	0,540	0,857
15	93,33	0,145	0,863	63	33,33	-0,047	0,911
21	100	0,000	0,823	64	86,67	0,111	0,850
26	86,67	0,329	0,784	65	60,00	0,123	0,880
28	66,67	0,264	0,851	70	73,33	0,063	0,916
29	80,00	0,102	0,862	71	100	0,000	0,784
31	46,67	-0,038	0,840	72	73,33	0,144	0,806
36	53,33	-0,062	0,920	75	80,00	0,058	0,847
37	73,33	-0,055	0,817	76	93,33	0,293	0,881
39	40,00	-0,256	0,617	77	66,67	0,000	0,933
40	60,00	0,344	0,923	78	86,67	-0,200	0,760
41	66,67	0,304	0,848	79	100	0,000	0,880

P = Schwierigkeit in Stichprobe 1, T_K= Trennschärfe der kognitiven Empathie in Stichprobe 1, T_E = Trennschärfe der emotionalen Empathie in Stichprobe 1, MET-A = Multidimensionaler Empathietest Parallelversion A.

ERGEBNISSE

Tabelle 8B: Itemkennwerte von MET-B.

Item	P	T _K	T _E	Item	P	T _K	T _E
2	100	0,000	0,787	38	66,67	0,478	0,884
3	86,67	0,085	0,742	42	66,67	-0,135	0,773
6	40,00	0,227	0,833	43	33,33	0,000	0,886
8	73,33	0,024	0,890	45	46,67	0,159	0,849
14	66,67	0,315	0,897	47	60,00	0,177	0,733
16	73,33	0,443	0,818	48	80,00	0,419	0,803
17	80,00	0,229	0,712	53	6,67	0,091	0,736
18	80,00	0,229	0,859	55	66,67	0,046	0,796
19	80,00	-0,042	0,884	56	33,33	-0,155	0,915
20	66,67	0,121	0,854	57	100	0,000	0,890
22	86,67	0,193	0,839	58	86,67	-0,125	0,882
23	100	0,000	0,866	59	53,33	0,458	0,848
24	46,67	-0,285	0,855	60	86,67	-0,226	0,865
25	73,33	0,443	0,846	66	80,00	0,229	0,797
27	46,67	-0,133	0,895	67	80,00	-0,085	0,931
30	53,33	0,308	0,826	68	20,00	-0,013	0,821
32	40,00	-0,068	0,830	69	100	0,000	0,891
33	73,33	-0,391	0,840	73	73,33	0,288	0,878
34	46,67	-0,006	0,868	74	93,33	-0,266	0,766
35	46,67	0,159	0,711	80	100	0,000	0,876

P = Schwierigkeit in Stichprobe 1, T_K = Trennschärfe der kognitiven Empathie in Stichprobe 1, T_E = Trennschärfe der emotionalen Empathie in Stichprobe 1, MET-B = Multidimensionaler Empathietest Parallelversion B.

4.1.3.5. Reliabilität

Die Version MET-A zeigte für die kognitiven Empathie eine interne Konsistenz von $\alpha = 0,46$. Die Version MET-B zeigte für die kognitive Empathie eine interne Konsistenz von $\alpha = 0,40$. Wurden alle 80 Items zusammengefasst ergab sich eine Gesamtreliabilität von $\alpha = 0,58$ für die kognitive Empathie. Die interne Konsistenz der emotionalen Empathie lag in MET-A und MET-B bei $\alpha = 0,99$.

4.2. Ergebnisse der zweiten Stichprobe

4.2.1. Erstellung der Parallelversionen

Aus den erhobenen Daten der ersten Stichprobe wurden wie in **4.1.3.3.** beschrieben erneut Schwierigkeit und Trennschärfe aller Items bestimmt sowie nach Testhälften aufgeteilt. Eine Übersicht der Mittelwerte ist in **Tabelle 9** zu sehen.

Tabelle 9: Mittelwerte der Schwierigkeiten und Trennschärfen in Stichprobe 1.

MET-Version	A	B
P, MW ± SD	73,17 ± 18,46	67,33 ± 23,18
T_K, MW ± SD	0,099 ± 0,197	0,078 ± 0,216
T_E, MW ± SD	0,844 ± 0,060	0,837 ± 0,057

MW = Mittelwert, SD = Standardabweichung, P = Schwierigkeit in Stichprobe 1, T_K = Trennschärfe der kognitiven Empathie in Stichprobe 1, T_E = Trennschärfe der emotionalen Empathie in Stichprobe 1, MET = Multidimensionaler Empathietest.

Hieraus ist erkennbar, dass in den aktuellen Parallelversionen in dieser Stichprobe die Schwierigkeit in Version A geringer war als in Version B (da der Schwierigkeitsindex höher ausfällt, je leichter ein Item ist). Die subjektive Rückmeldung der Probanden war ebenfalls, dass Testhälfte B als schwerer empfunden wurde. Der Unterschied ist jedoch nicht signifikant ($t(20) = 6,08$, $p = 0,175$, $d = 0,22$). Gleiches gilt für die Trennschärfen der kognitiven Empathie ($t(20) = 0,02$, $p = 0,655$, $d = 0,07$) und die Trennschärfen der emotionalen Empathie ($p = 0,554$, $r = 0,13$, Wilcoxon-Test). Es wurden deshalb die Itemblöcke wie in **Tabelle 10** dargestellt angepasst, so dass die Schwierigkeiten der einzelnen Blöcke, welche anhand der Daten aus der ersten Stichprobe errechnet wurden, möglichst ähnlich waren. Da MET-B subjektiv und objektiv als schwerer bewertet wurde, wurden hierzu Items mit hohem Schwierigkeitsindex (= leichte Items) aus MET-A gegen Items mit niedrigem Schwierigkeitsindex (= schwere Items) aus MET-B getauscht. Die Trennschärfen sollten bei den getauschten Items nach Möglichkeit ähnlich ausfallen.

Tabelle 10: Itemblöcke von MET-A* und MET-B*.

MET-Version	A*	B*
Block 1 a/b		
Item Nr.	1, 3 , 9, 13, 26, 31, 50, 63, 74 , 78	2, 15 , 22, 27, 33, 42, 43, 58, 66, 72
P, MW ± SD	73,99 ± 23,81	73,99 ± 20,71
Block 2 a/b		
Item Nr.	5, 11, 25 , 29, 46, 51, 52, 64, 68 , 76	17, 20, 24, 35, 39 , 45, 59, 60, 71 , 80
P, MW ± SD	66,70 ± 20,84	66,67 ± 23,31
Block 3 a/b		
Item Nr.	7, 10, 21, 36, 37, 41, 56 , 61, 62, 77	8, 14, 23, 32, 47, 48, 49 , 57, 67, 73
P, MW ± SD	73,33 ± 21,78	73,67 ± 15,91
Block 4 a/b		
Item Nr.	4, 12, 28, 40, 53 , 54, 65, 70, 75, 79	6, 16, 18, 19, 30, 34, 38, 44 , 55, 69
P, MW ± SD	66,00 ± 25,42	66,00 ± 18,18

MW = Mittelwert, SD = Standardabweichung, P = Schwierigkeit in Stichprobe 1, Block a = kognitive Empathie, Block b = emotionale Empathie, MET = Multidimensionaler Empathietest, MET-A* = Multidimensionaler Empathietest Parallelversion A*, MET-B* = Multidimensionaler Empathietest Parallelversion MET-B*. Fett markiert sind jeweils die geänderten Items.

Tabelle 11 enthält die Mittelwerte der Schwierigkeit und Trennschärfe bezogen auf alle Items der so entstandenen neuen MET-Versionen A* und B*. Die Itemmerkmale (Alter, Geschlecht, Emotionsvalenz) innerhalb der Versionen blieben auch nach der Anpassung unverändert. Die MET-Versionen A* und B* konnten nun an einer neuen Stichprobe 2 erprobt werden.

Tabelle 11: Übersicht von Schwierigkeit und Trennschärfe von MET-A* und MET-B*.

MET-Version	A*	B*
P, MW ± SD	70,00 ± 22,44	70,10 ± 19,33
T_K, MW ± SD	0,115 ± 0,230	0,084 ± 0,224
T_E, MW ± SD	0,843 ± 0,056	0,830 ± 0,070

MW = Mittelwert, SD = Standardabweichung, P = Schwierigkeit in Stichprobe 1, T_K = Trennschärfe der kognitiven Empathie in Stichprobe 1, T_E = Trennschärfe der emotionalen Empathie in Stichprobe 1, MET = Multidimensionaler Empathietest, MET-A* = Multidimensionaler Empathietest Parallelversion A*, MET-B* = Multidimensionaler Empathietest Parallelversion MET-B*.

4.2.2. Studienteilnehmer

4.2.2.1. Demographische Daten

Für die zweite Stichprobe wurden n = 15 Probanden rekrutiert, welche alle in die Studie eingeschlossen werden konnten. Aufgrund des recht hohen Zeitaufwandes der Studie bei fehlender Aufwandsentschädigung erwies es sich im zeitlich anvisierten Rahmen als schwierig, erneut eine Stichprobe von 20 gesunden Probanden zu rekrutieren, weshalb schließlich eine kleinere Stichprobe gewählt wurde. Von den 15 eingeschlossenen Probanden waren 13 weiblich (86,7 %) und 2 männlich (13,3%). Die Teilnehmer hatten ein mittleres Alter von 30,2 Jahren, wobei der jüngste Teilnehmer 23 und der älteste 50 Jahre alt war. Der Großteil der Studienteilnehmer lebte in einer festen Partnerschaft. Die Mehrheit der Teilnehmer hatte einen hohen Bildungsgrad, dennoch waren weniger als die Hälfte der Teilnehmer berufstätig (siehe **Tabelle 12**). Alle Probanden waren deutsche Muttersprachler. Ein Teilnehmer war österreichischer Staatsbürger, ein anderer hatte zusätzlich die spanische Staatsbürgerschaft.

Tabelle 12: Demographische Daten von Stichprobe 2.

Frauen/Männer, Anzahl (%)	13 (86,7 %) / 2 (13,3 %)
Alter (Jahre), MW ± SD	30,2 ± 7,4
Familienstand, Anzahl (%)	
Ledig	13 (86,7 %)
Verheiratet	2 (13,3 %)
Partnersituation, Anzahl (%)	
Kein Partner	3 (20 %)
Fester Partner (kein Ehepartner)	9 (60 %)
Fester Partner (Ehepartner)	3 (20 %)
Schulabschluss, Anzahl (%)	
Mittlere Reife/Realschulabschluss	1 (6,7 %)
Abitur/Fachabitur	14 (93,3 %)
Berufsabschluss, Anzahl (%)	
Noch in Berufsausbildung	4 (26,7 %)
Lehre/Fachschule	1 (6,7 %)
Fachhochschule/Universität	9 (60 %)
Sonstige	1 (6,7 %)
Aktuelle berufliche Situation, Anzahl (%)	
Berufstätig, Vollzeit	2 (13,3 %)
Berufstätig, gelegentlich	4 (26,7 %)
Hausfrau/-mann, nicht berufstätig	3 (20 %)
Ausbildung	2 (13,3 %)
Mithelfender Familienangehöriger	4 (26,7 %)

MW = Mittelwert, SD = Standardabweichung.

4.2.2.2. Klinische und neuropsychologische Daten

Das Mini-International Neuropsychiatric Interview (M.I.N.I.) ergab bei keinem Teilnehmer eine aktuelle oder in der Vergangenheit bestehende psychiatrische Diagnose. Im Hinblick auf eine mögliche depressive Symptomatik ergaben sich in der HDRS-24 entsprechend dem Einschlusskriterium bei keinem der Teilnehmer eine klinisch relevante depressive Symptomatik. Auch laut BDI-II bestand bei keinem der Teilnehmer eine depressive

Symptomatik. Der CTQ zeigte auf der Skala der emotionalen Misshandlung bei einem Teilnehmer nach Definition von Häuser und Kollegen (2011) eine traumatische Kindheits-erfahrung im Sinne einer emotionalen Misshandlung. In der Toronto Alexithymie Skala lag kein Proband über dem Cut-Off für die Feststellung einer Alexithymie.

Auch die Probanden der zweiten Stichprobe schnitten in den neuropsychologischen Tests im Vergleich zu Normstichproben überdurchschnittlich gut ab. Eine Übersicht der Ergebnisse der klinischen und neuropsychologischen Diagnostik gibt **Tabelle 13**.

Tabelle 13: Übersicht klinische und neuropsychologische Diagnostik in Stichprobe 2.

HDRS-24 (MW ± SD)	0,53 ± 0,99
BDI-II (MW ± SD)	1,47 ± 1,89
CTQ (MW ± SD)	
Emotionale Misshandlung	6,73 ± 2,81
Körperliche Misshandlung	5,27 ± 1,03
Emotionale Vernachlässigung	7,27 ± 2,79
Körperliche Vernachlässigung	5,73 ± 1,10
Sexueller Missbrauch	5,13 ± 0,52
TAS (MW ± SD)	34,73 ± 8,72
Zahlennachsprechen (MW ± SD)	
Prozentrang vorwärts	71,13 ± 30,09
Prozentrang rückwärts	67,47 ± 29,74
Regensburger Wortflüssigkeitstest (MW ± SD)	
Prozentrang Tiere	59,60 ± 29,81
Prozentrang S-Wörter	50,33 ± 23,52
Trail Making Test (MW ± SD)	
Zeit in s, Teil A	23,20 ± 6,16
Zeit in s, Teil B	49,87 ± 13,87
z-Wert, Teil A	0,85 ± 1,41
z-Wert, Teil B	0,27 ± 1,77
MWT-B (IQ MW ± SD)	116,20 ± 12,90
VLMT (Rohwert MW ± SD)	
Supraspanne	9,47 ± 1,77
Lernleistung	14,13 ± 1,30
Abruf nach Interferenz	13,60 ± 1,30
Abruf nach Verzögerung	13,40 ± 1,81
Wiedererkennungslleistung	14,33 ± 1,23

MW = Mittelwert, SD = Standardabweichung, HDRS-24 = Hamilton Depression Rating Scale 24-Item Version, BDI II = Beck-Depressions-Inventar II, CTQ = Childhood Trauma Questionnaire, TAS = Toronto Alexithymie Skala, MWT-B = Mehrfach-Wortschatz-Intelligenztest Version B, VLMT = Verbaler Lern- und Merkfähigkeitstest.

4.2.3. Statistische Auswertung

8 Teilnehmer (53,3 %) bearbeiteten zuerst MET-A, 7 Teilnehmer (46,7 %) zuerst MET-B. Zwischen der Bearbeitung der beiden Versionen lagen 1 bis 7 Tage (MW $2,07 \pm 1,90$), 12 Teilnehmer (80 %) bearbeiteten die beiden Testversionen innerhalb von 3 Tagen.

4.2.3.1. Korrelation

MET-A* und MET-B* zeigten in der Pearson Korrelation bezüglich der Summe der richtigen Antworten ($r = 0,720$, $p = 0,002$), ihrer Bewertungen der emotionalen Empathie ($r = 0,970$, $p < 0,001$) sowie der Antwortlatenzen der kognitiven ($r = 0,841$, $p < 0,001$) und emotionalen ($r = 0,615$, $p = 0,015$) Empathie Korrelationen im mittleren bis hohen Bereich.

4.2.3.2. Vergleich der Mittelwerte

Beim Vergleich der Mittelwerte zeigte sich im t-Test kein signifikanter Unterschied zwischen den Versionen MET-A* und MET-B* für die Summe der richtigen Antworten der kognitiven Empathie ($t(15) = -0,33$, $p = 0,709$, $d = 0,10$) sowie für die Bewertungen der emotionalen Empathie ($t(15) = 0,23$, $p = 0,053$, $d = 0,55$). Ebenfalls zeigte sich kein signifikanter Unterschied zwischen MET-A* und MET-B* bezüglich der Mittelwerte der Antwortlatenzen der kognitiven ($t(15) = 329,67$, $p = 0,487$, $d = 0,18$) und der emotionalen Empathie ($t(15) = -78,14$, $p = 0,760$, $d = 0,08$). Siehe hierzu auch **Abbildung 13**.

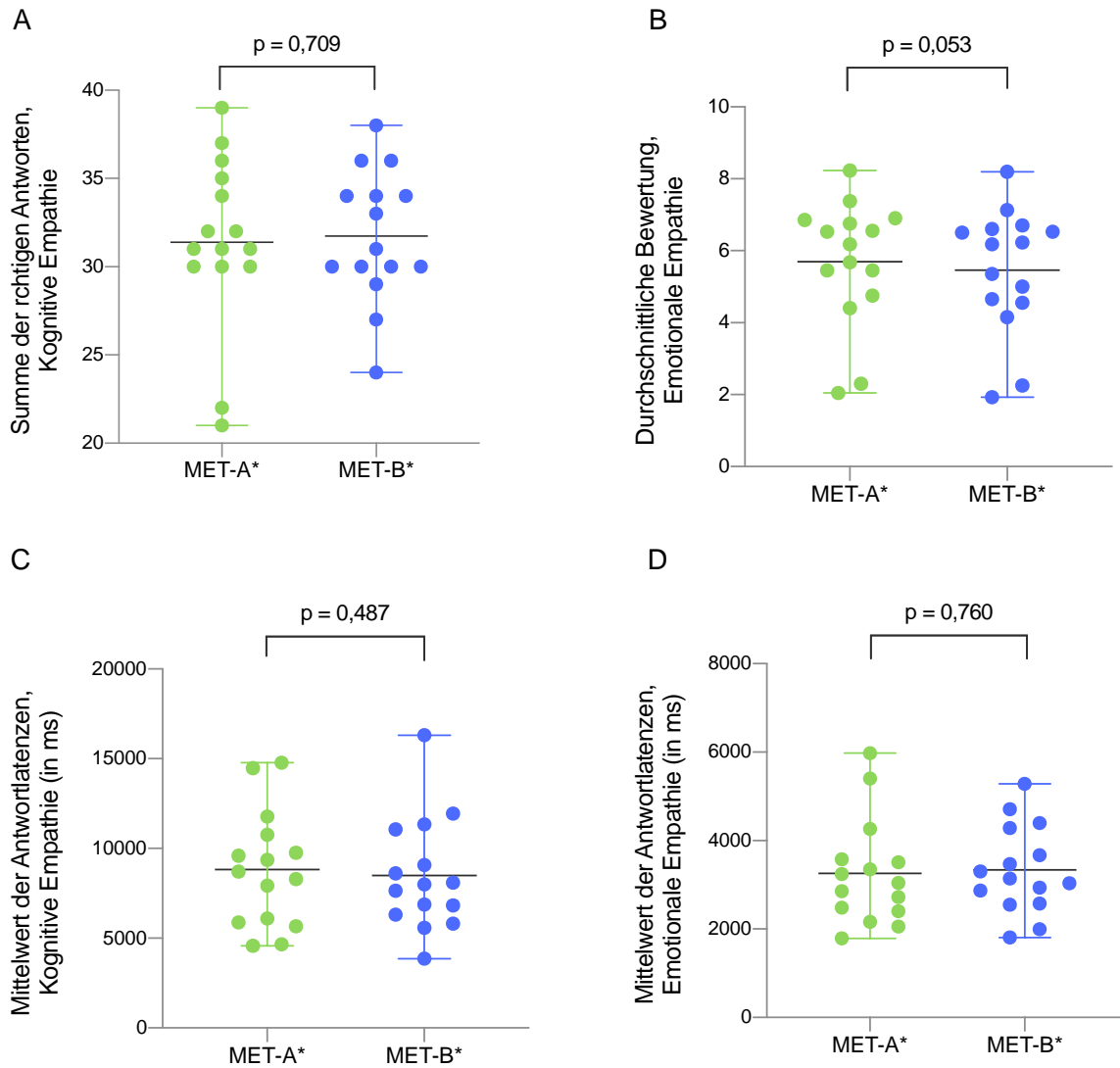


Abbildung 13: Vergleich der Mittelwerte in Stichprobe 2.

- A:** Summe der richtigen Antworten der kognitiven Empathie in MET-A* und MET-B*.
 - B:** Durchschnittliche Bewertung der emotionalen Empathie in MET-A* und MET-B*.
 - C:** Mittelwert der Antwortlatenzen der kognitiven Empathie in MET-A* und MET-B*.
 - D:** Mittelwert der Antwortlatenzen der emotionalen Empathie in MET-A* und MET-B*.
- MET-A* = Multidimensionaler Empathietest Parallelversion A*, MET-B* = Multidimensionaler Empathietest B*. p = Signifikanzwert.

Eine Übersicht über die Mittelwerte mit Standardabweichungen und p -Werten gibt **Tabelle 14**.

Tabelle 14: Vergleich der Mittelwerte in Stichprobe 2.

MET-Version	A*	B*	p-Wert
Kognitive Empathie			
Summe richtige Antworten, MW ± SD	31,40 ± 4,88	31,73 ± 3,73	0,709
Antwortlatenz (in ms), MW ± SD	8817,65 ± 3208,70	8487,98 ± 3121,42	0,487
Emotionale Empathie			
Durchschnittliche Bewertung, MW ± SD	5,70 ± 1,74	5,46 ± 1,75	0,053
Antwortlatenz (in ms), MW ± SD	3256,88 ± 1189,05	3335,02 ± 988,69	0,760

MW = Mittelwert, SD = Standardabweichung, MET = Multidimensionaler Empathietest. p-Wert = Signifikanzwert.

4.2.3.3. Schwierigkeit

Die Itemschwierigkeit lag sowohl im MET-A* als auch im MET-B* bei $P^* = 20,00$ bis $P^* = 100$. In der zweiten Stichprobe gab es in MET-A* drei Items, in MET-B* fünf Items, welche alle Teilnehmer richtig lösten ($P^* = 100$). Im Mittel lag die Schwierigkeit in MET-A* bei 71,33 ($\pm 21,06$), in MET-B* bei 72,22 ($\pm 21,28$). Dabei zeigte sich kein signifikanter Unterschied im Wilcoxon-Test ($p = 0,805$, $r = 0,06$). **Abbildung 14** zeigt die Häufigkeitsverteilung der Schwierigkeiten in MET-A* und MET B*. Es weisen immer noch einige Items Schwierigkeiten um $P^* = 50$ (= gut geeignet), jedoch auch um $P^* = 100$ (= nicht gut geeignet) auf.

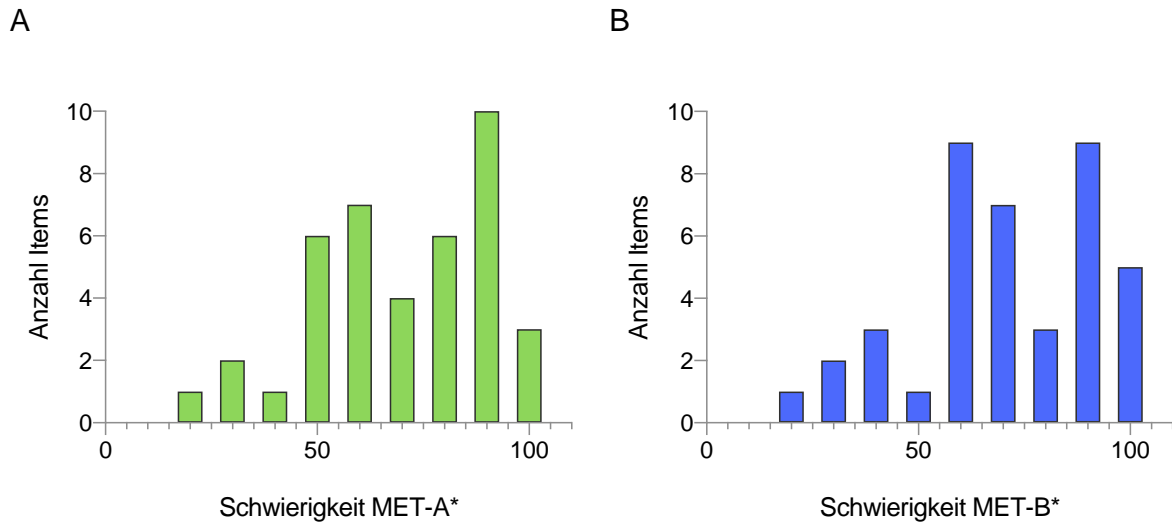


Abbildung 14: Häufigkeitsverteilung der Itemschwierigkeiten in Stichprobe 2.

A: Itemschwierigkeit in MET-A*.

B: Itemschwierigkeit in MET-B*.

MET-A* = Multidimensionaler Empathietest Parallelversion A*, MET-B* = Multidimensionaler Empathietest Parallelversion B*.

4.2.3.4. Trennschärfe

Die Trennschärfekoeffizienten für die kognitive Empathie von MET-A* lagen bei -0,254 bis 0,798 (MW $0,249 \pm 0,288$). Dabei lag die Trennschärfe von 12 Items bei ≤ 0 . 18 Items (45 %) lagen über 0,30, was als zufriedenstellender Wert angesehen wurde. Die Trennschärfekoeffizienten für die kognitive Empathie von MET-B* lagen bei -0,439 bis 0,586 (MW $0,143 \pm 0,294$). Die Trennschärfe von 14 Items lag bei ≤ 0 . Nur 10 Items (25 %) lagen über 0,30. Dabei zeigte sich bezüglich der Trennschärfen der kognitiven Empathie kein signifikanter Unterschied zwischen den Versionen ($t(15) = 0,106$, $p = 0,116$, $d = 0,25$). Die Trennschärfekoeffizienten für die emotionale Empathie von MET-A* lagen bei 0,097 (Item 64) bis 0,970 (MW $0,781 \pm 0,137$), die für MET-B* bei 0,474 bis 0,925 (MW $0,767 \pm 0,095$). Auch hier zeigte sich kein signifikanter Unterschied ($p = 0,532$, $r = 0,16$, Wilcoxon-Test). Damit wiesen beide Testhälften größtenteils gute Trennschärfen im Bereich der emotionalen Empathie auf (Kelava & Moosbrugger, 2012). **Abbildung 15** zeigt die Häufigkeitsverteilung der Trennschärfekoeffizienten in MET-A* und MET B*. Es ist zu erkennen, dass es bezüglich der kognitiven Empathie zwar immer noch einige Items mit Trennschärfen um und unter 0 gibt, jedoch im Vergleich zu den vorherigen Versionen

mehr Items mit guten Trennschärfekoeffizienten. Insbesondere MET-A* weist für die kognitive Empathie einige Items mit Trennschärfen $> 0,5$ auf. Für die emotionale Empathie weisen beide Versionen gute Trennschärfen auf. Lediglich in Version MET-A* gibt es für Item 64 einen Ausreißer.

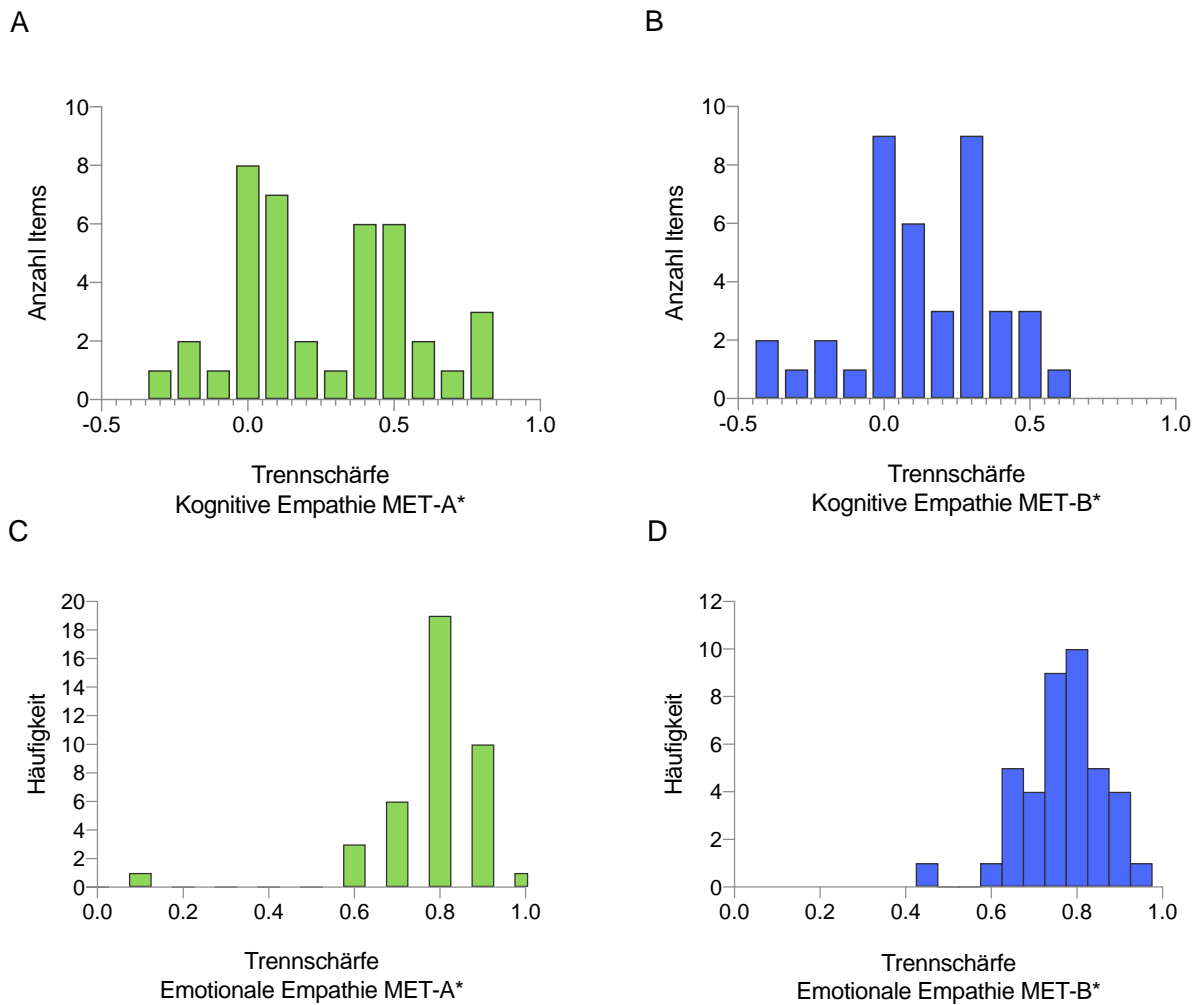


Abbildung 15: Häufigkeitsverteilung der Trennschärfekoeffizienten in Stichprobe 2.

- A:** Trennschärfe der kognitiven Empathie in MET-A*.
 - B:** Trennschärfe der kognitiven Empathie in MET-B*.
 - A:** Trennschärfe der emotionalen Empathie in MET-A*.
 - B:** Trennschärfe der emotionalen Empathie in MET-B*.
- MET-A* = Multidimensionaler Empathietest Parallelversion A*, MET-B* = Multidimensionaler Empathietest Parallelversion B*.

Zu den Werten des Schwierigkeitsindex und Trennschärfekoeffizienten in Stichprobe 1 siehe **Tabelle 15A** und **15B**.

ERGEBNISSE

Tabelle 15A: Itemkennwerte von MET-A*.

Item	P*	T _K *	T _E *	Item	P*	T _K *	T _E *
1	73,33	0,358	0,853	46	46,67	-0,006	0,820
3	82,22	-0,039	0,817	50	91,11	0,493	0,809
4	55,56	0,081	0,694	51	46,67	0,579	0,970
5	82,22	0,433	0,855	52	28,89	0,100	0,694
7	91,11	0,553	0,840	53	64,44	0,152	0,761
9	100	0,000	0,832	54	91,11	0,493	0,815
10	91,11	-0,087	0,839	56	20,00	0,456	0,882
11	73,33	-0,182	0,787	61	82,22	0,798	0,747
12	82,22	0,798	0,859	62	73,33	0,358	0,917
13	37,78	0,450	0,863	63	46,67	0,201	0,786
21	91,11	0,493	0,806	64	100	0,000	0,097
25	91,11	-0,030	0,593	65	82,22	0,258	0,820
26	82,22	0,798	0,797	68	46,67	-0,035	0,770
28	64,44	-0,042	0,710	70	64,44	0,086	0,792
29	28,89	0,389	0,881	74	64,44	0,119	0,841
31	91,11	-0,169	0,627	75	73,33	0,708	0,634
36	46,67	0,432	0,774	76	100	0,000	0,754
37	91,11	-0,254	0,746	77	55,56	0,122	0,856
40	46,67	0,100	0,862	78	55,56	0,122	0,696
41	91,11	0,389	0,932	79	91,11	0,493	0,795

P* = Schwierigkeit in Stichprobe 2, T_K*= Trennschärfe der kognitiven Empathie in Stichprobe 2, T_E* = Trennschärfe der emotionalen Empathie in Stichprobe 2, MET-A* = Multidimensionaler Empathietest Parallelversion A*.

ERGEBNISSE

Tabelle 15B: Itemkennwerte von MET-B*.

Item	P*	T _K *	T _E *	Item	P*	T _K *	T _E *
2	73,33	0,278	0,749	39	28,89	-0,042	0,670
6	73,33	0,481	0,773	42	37,78	0,091	0,849
8	82,22	0,441	0,773	43	37,78	-0,022	0,795
14	91,11	0,288	0,918	44	20,00	0,156	0,796
15	100	0,000	0,864	45	73,33	0,586	0,891
16	73,33	0,085	0,665	47	46,67	0,310	0,791
17	64,44	-0,205	0,582	48	82,22	0,265	0,925
18	91,11	-0,015	0,899	49	91,11	0,060	0,871
19	91,11	0,060	0,818	55	55,56	0,478	0,660
20	73,33	-0,009	0,657	57	100	0,000	0,675
22	55,55	0,389	0,780	58	64,44	0,306	0,474
23	100	0,000	0,806	59	73,33	0,229	0,797
24	91,11	0,288	0,698	60	91,11	0,135	0,850
27	28,89	-0,142	0,796	66	64,44	-0,439	0,772
30	37,78	0,541	0,709	67	55,56	0,175	0,846
32	64,44	0,445	0,779	69	100	0,000	0,685
33	82,22	-0,332	0,649	71	100	0,000	0,894
34	64,44	0,261	0,736	72	91,11	-0,376	0,772
35	55,56	0,302	0,730	73	73,33	0,278	0,728
38	91,11	0,060	0,765	80	91,11	-0,162	0,778

P* = Schwierigkeit in Stichprobe 2, T_K* = Trennschärfe der kognitiven Empathie in Stichprobe 2, T_E* = Trennschärfe der emotionalen Empathie in Stichprobe 2, MET-B* = Multidimensionaler Empathietest Parallelversion MET-B*.

4.2.3.5. Reliabilität

Die Version MET-A* zeigte für die kognitiven Empathie eine interne Konsistenz von $\alpha = 0,76$. Die Version MET-B* zeigte für die kognitive Empathie eine interne Konsistenz von $\alpha = 0,59$. Wurden alle 80 Items zusammengefasst ergab sich eine Gesamtreliabilität von $\alpha = 0,82$ für die kognitive Empathie. Die interne Konsistenz der emotionalen Empathie lag in MET-A* und MET-B* bei $\alpha = 0,98$.

5. Diskussion

Bislang fehlt ein Messinstrument, um Empathie im longitudinalen Verlauf, beispielsweise in verschiedenen Stadien der Depression oder im Rahmen einer Empathie fördernden Intervention, erfassen zu können. Das Ziel dieser Arbeit war es, zwei Parallelversionen des Multidimensionalen Empathietests (MET) zu erstellen und sie zur Überprüfung der Parallelität an einer Stichprobe psychisch gesunder Probanden zu erproben. Die erzielten Ergebnisse sollen im Folgenden diskutiert werden. Da die initialen MET-Versionen keine ausreichende Vergleichbarkeit bezüglich Korrelation und Mittelwertvergleich aufwiesen, beziehen sich die zunächst diskutierten Ergebnisse ausschließlich auf die finalen Parallelversionen.

5.1. Korrelation zwischen den Parallelversionen

Die Parallelversionen MET-A* und MET-B* korrelieren in allen Kategorien im mittleren bis hohen Bereich ($r = 0,62$ bis $0,97$). Bei der Erstellung der DAS-18-Parallelformen ergab sich ebenfalls eine Korrelation im mittleren bis hohen Bereich von $r = 0,50$ bis $1,0$ (Rojas et al., 2014). Bei den DAS-40 Parallelformen berichteten Weissman & Beck (1978) eine Korrelation von $r = 0,79$. Lehl (1995) gab für die Parallelformen MWT-A und MWT-B eine Korrelation von $r = 0,84$ an. Im Vergleich zu bisher erstellten Parallelformen zeigen die Parallelversionen MET-A* und MET-B* also eine zufriedenstellende Korrelation.

5.2. Vergleich der Mittelwerte

Im Mittelwertvergleich zeigten sich für MET-A* und MET-B* keine signifikanten Unterschiede in den Antwortsummen, den emotionalen Empathiewerten und den Antwortlatenzen ($p = 0,053$ bis $0,760$). Auch für die Parallelversionen DAS-18A und DAS-18B von Rojas et al. zeigten sich keine signifikanten Unterschiede in den Mittelwertvergleichen ($p = 0,07$ bis $0,65$) (Rojas et al., 2014). In der zuvor erfolgten Paralleltesterstellung der DAS-40A und DAS-40B wurden die Mittelwerte zwar genannt (113,01 für die DAS-40A; 113,73 für die DAS-40B), jedoch nicht miteinander verglichen, so dass keine Aussage zu statistisch signifikanten Unterschieden gemacht werden kann (Weissman & Beck, 1978). Auch

für die Mittelwertvergleiche zeigen die in dieser Arbeit erstellten Parallelversionen zufriedenstellende Ergebnisse.

5.3. Itemschwierigkeit

In der Validierung des MET-KE (Lemme, 2012) ergaben sich Schwierigkeiten zwischen 10,26 und 94,87 (MW 66,02 \pm 17,36). Damit liegen die in dieser Arbeit berechneten Schwierigkeitsindizes für den MET-A* (MW 71,33 \pm 21,06) und MET-B* (MW 72,22 \pm 21,28) höher als in der ursprünglichen Validierung. Insgesamt scheinen die Items für die hier gewählte Stichprobe also zu leicht zu sein, da es mehrere Items gibt, welche von allen Teilnehmern richtig gelöst wurden und die somit keine weitere Differenzierung der offenbar bestehenden hohen Merkmalsausprägung ermöglichen. Dies könnte mit der gewählten Stichprobe zusammenhängen, welche weitgehend aus Personen mit hohem Bildungsgrad und hoher Intelligenz (MW des IQ im MWT-B 116,20 \pm 12,90) besteht. Die Stichprobe von Lemme bestand im Gegensatz hierzu aus Personen mit niedrigem Bildungsabschluss, wobei keine genaueren demographischen Angaben gemacht wurden (Lemme, 2012). Auch in der NeMUP-Stichprobe liegen die Schwierigkeitsindizes mit 22,86 bis 97,06 (MW 62,04 \pm 19,24) unterhalb der in dieser Arbeit berechneten Werte. Soziodemographische Daten liegen nicht vor. In vorherigen Studien, in denen der MET eingesetzt wurde, wird nicht explizit auf die Itemschwierigkeit sowie auf den Zusammenhang von soziodemographischen Aspekten und erreichten Empathiewerten eingegangen (Dziobek et al. 2008; Poustka et al., 2010; Dziobek et al., 2011, Ritter et al., 2011; Schuler et al., 2019), so dass keine Aussage darüber gemacht werden kann, ob hoher Bildungsgrad und hohe Intelligenz tatsächlich auch mit höheren kognitiven Empathiewerten im MET einhergehen. In der Studie von Schuler et al. (2019) wird für die übergriffig gewordenen pädoophilen Probanden ein durchschnittlicher IQ Wert im Wechsler-Intelligenztest (WAIS) von 95,72 (\pm 14,70) angegeben, wohingegen die nicht übergriffig gewordenen pädoophilen Probanden einen durchschnittlichen IQ Wert von 103,74 (\pm 13,66) erreichten. Im Bereich der kognitiven Empathie, aus welcher sich die Schwierigkeit errechnen ließe, schnitten letztere besser ab (Schuler et al., 2019). Das höhere Intelligenzniveau könnte ebenfalls einen Faktor hierfür darstellen. Jedoch wurden keine Angaben gemacht, ob es signifikante Gruppenunterschiede bezüglich des

Intelligenzniveaus gab (ebd.). Darüber hinaus konnten statistisch signifikante Unterschiede im Bereich der kognitiven Empathie nur für die Items mit abgebildeten Kindern gezeigt werden (ebd.). In den Arbeiten von Rojas et al. (2014) und Weissman & Beck (1978) im Rahmen der Erstellung der DAS-Parallelversionen wurde nicht auf die Itemschwierigkeit eingegangen, was jedoch durch die Konstruktion der DAS zu erklären ist, bei deren Skalen es kein Richtig oder Falsch und damit keine Schwierigkeit als solche gibt. Für den MWT-A und MWT-B konnten keine expliziten Itemschwierigkeiten gefunden werden, jedoch ist der MWT so konstruiert, dass die Schwierigkeit vom ersten zum letzten Item kontinuierlich ansteigt (Lehrl, 1999) und somit mit dem MET nur bedingt vergleichbar.

5.4. Trennschärfe

Bei den finalen Parallelversionen des MET können bezüglich der Trennschärfe für die kognitive Empathie im MET-A* 18 von 40 Items (45 %) und im MET-B* 10 von 40 Items (25 %) als zufriedenstellend angesehen werden. Für die emotionale Empathie weisen beide Versionen gute Trennschärfekoeffizienten auf. Ausnahme bildet Item 64 im MET-A*, welches für die emotionale Empathie eine nicht zufriedenstellende Trennschärfe aufweist. In der initialen Parallelversion sowie in der Validierungsstudie weist dieses Item jedoch eine gute Trennschärfe auf, so dass der schlechte Wert möglicherweise zufällig bedingt ist. Bei der Erstellung der DAS-18A und DAS-18B wiesen in der Stichprobe der gesunden Probanden 15 von 23 Items (83,33 %, DAS-18A) bzw. 16 von 23 Items (88,89 %, DAS-18B) akzeptable Trennschärfen $\geq 0,30$ auf (Rojas et al., 2014). Damit erreichen die DAS-18-Parallelversionen insgesamt bessere Trennschärfen als die im Rahmen dieser Arbeit erstellten Parallelversionen. In der Validierungsstudie des MET-KE (Lemme, 2012) wiesen 47 von 80 Items (58,75 %) nach den Maßstäben dieser Arbeit als akzeptabel angesehene Trennschärfen $\geq 0,30$ auf. 33 Items (41,25 %) wiesen bereits in der Validierung nicht akzeptable Trennschärfen auf. In der Stichprobe des NeMUP-Projektes weisen ebenfalls nur 15 von 80 Items (18,75%) akzeptable Trennschärfen $\geq 0,30$ auf. Daraus lässt sich schlussfolgern, dass sich die zum Teil ungenügenden Trennschärfekoeffizienten im Bereich der kognitiven Empathie nicht aus der Erstellung der Parallelversionen und damit Verkürzung des MET auf 40 Bilder ergeben, sondern bereits in der ursprünglichen 80-Bilder Version zu finden sind. Die geringen Trennschärfen könnten zum

Teil durch die Konstruktion des MET bedingt sein, da es sich bei der Messung der kognitiven Empathie um ein dichotomes Merkmal (richtig = 1/falsch = 0) mit vier vorgegebenen Antwortmöglichkeiten handelt und somit keine Abstufung vorgesehen ist. Die Trennschärfe ist jedoch als Korrelation definiert. Hierfür sprechen die sehr guten Trennschärfen im Bereich der emotionalen Empathie. Zum anderen scheinen die Trennschärfen auch durch die Wahl der Stichprobe beeinflusst zu werden. So konnte in der Konstruktion der DAS-18-Parallelversionen bei den gesunden Probanden geringere Trennschärfekoeffizienten festgestellt werden als bei den depressiven Probanden (Rojas et al., 2014). Grundsätzlich sollten bei der Konstruktion eines Tests nur Items mit akzeptablen Trennschärfekoeffizienten verwendet werden. Bei der Erstellung der Parallelversionen stand jedoch kein weiteres Bildmaterial zur Verfügung. Zudem sollten die Versionen nicht kürzer als 40 Bilder werden.

5.5. Interne Konsistenz

MET-A* und MET-B* weisen für die kognitive Empathie mit $\alpha = 0,76$ und $\alpha = 0,59$ eine geringere interne Konsistenz und damit Reliabilität auf als in der Validierung des MET-KE (Lemme, 2012). Werden jedoch alle 80 Items der Parallelversionen zusammen betrachtet, verbessert sich die interne Konsistenz auf $\alpha = 0,82$. Das spricht dafür, dass die geringere Reliabilität auch Folge des verkürzten Testdesigns ist. Die interne Konsistenz für die emotionale Empathie liegt in beiden Versionen mit $\alpha = 0,98$ bis $0,99$ sehr hoch. Nimmt α sehr hohe Werte an ($\alpha > 0,90$), kann dies jedoch auch darauf hindeuten, dass die Items redundant bzw. zu homogen sind (Streiner, 2013). In der Studie von Schuler et al. wurden interne Konsistenzen von $\alpha = 0,72$ für die kognitive und $\alpha = 0,95$ für die emotionale Empathie beschrieben (Schuler et al., 2019), was mit den in dieser Arbeit erhobenen Werten sowie denen des ursprünglichen MET vergleichbar ist (Dziobek et al., 2008). Grundsätzlich gilt, dass die interne Konsistenz nicht als Eigenschaft eines Messinstruments verstanden werden sollte, sondern als eine Eigenschaft der Testwerte des Messinstruments (Streiner, 2013). Somit ist α auch immer von der gewählten Stichprobe abhängig (ebd.). Das zeigt sich auch bei der Erstellung der DAS-18-Parallelversionen: Für die gesunden Probanden wurde eine interne Konsistenz von $\alpha = 0,80$ angegeben, wogegen bei den depressiven Probanden mit $\alpha = 0,89$ eine höhere Reliabilität erreicht wurde

(Rojas et al, 2014). Hohe oder niedrige Werte der internen Konsistenz lassen demzufolge noch nicht auf das Messinstrument als solches schließen. Zudem sollte ein Testverfahren als Ganzes betrachtet werden und nicht allein aufgrund einer geringeren internen Konsistenz verworfen werden (Schmitt, 1996).

5.6. Methodenkritische Überlegungen

Im folgenden Abschnitt soll auf die Limitationen dieser Arbeit eingegangen werden, welche sich aus den gewählten Methoden ergeben.

5.6.1. Erstellung der Parallelversionen

Wie bereits unter **2.3.2.** beschrieben, stellt die Faktorenanalyse mit Bestimmung der Faktorenladungen der Items eine geeignete Methode zur Aufteilung der Items auf die Parallelversionen dar. Bei der Erstellung der MET-Versionen konnte dieses Verfahren aufgrund der Skalierung der Items nicht zur Anwendung kommen, so dass eine Aufteilung nach Schwierigkeit und Trennschärfe als Alternative gewählt wurde. Hierbei wurde Schwierigkeit und Trennschärfe für jedes Item aus den Datensätzen der existierenden Stichprobe bzw. im zweiten Schritt erneut aus der initial untersuchten Stichprobe berechnet und anhand dieser Werte die Parallelversionen erstellt. Für folgende Arbeiten wäre auch die vorherige Erprobung des Ursprungstests an einer mit der Zielgruppe vergleichbaren Stichprobe, in diesem Fall eine Stichprobe depressiver Patienten, in Erwägung zu ziehen, um auf diese Weise eine noch bessere Übertragbarkeit der Ergebnisse zu gewährleisten. So wurden für die Erstellung der DAS-18-Parallelversionen die Testdaten von depressiven und gesunden Probanden aus einer Vorgängerstudie verwendet, was für das genannte Messverfahren ebenfalls der Zielgruppe entsprach (Rojas et al., 2014). Für die Erstellung der ursprünglichen 40-Item-Parallelversionen wurden jedoch wie in dieser Studie ausschließlich Daten von gesunden Probanden verwendet (Weissmann & Beck, 1978). Zusätzlich kann darüber diskutiert werden, ob eine objektivere und damit präzisere Möglichkeit zur Bildung der Itempaare existiert. Bei der Erstellung der hier vorliegenden Parallelversionen gab es keine konkreten, objektiven Kriterien, wann Items als

ähnlich genug anzusehen waren. Außerdem wurden mangels Alternativen teilweise Itempaare gebildet, die sich in ihrer Schwierigkeit stark ähnelten, jedoch unterschiedliche Trennschärfen aufwiesen und umgekehrt. Möglicherweise muss der Itempool sogar erweitert werden, um bei Beibehaltung der Testlänge tatsächlich gleichwertige Itempaare bilden zu können. So wurden bei der Erstellung der ursprünglichen DAS-40-Parallelversionen aus den 100 zur Verfügung stehenden Items 20 Items ausgeschlossen (Weissmann & Beck, 1978). Hierbei ist jedoch zu erwähnen, dass die 100-Item-DAS im Gegensatz zum MET kein etabliertes Testverfahren darstellt, sondern im Rahmen derselben Studie erst konstruiert wurde (ebd.). Bei der Erstellung der DAS-18-Parallelversionen wurden aufgrund von unzureichender Trennschärfe- und Reliabilitätswerte in vorherigen Studien 4 Items aus dem ursprünglichen Itempool ausgeschlossen (Rojas et al., 2014). Die Items wurden dann anhand ihrer Faktor-Item-Ladung aus der Faktorenanalyse den Kurzversionen zugeteilt (ebd.). Objektive Kriterien für die Ähnlichkeit der Itempaare wurden durch die Autoren nicht genannt (ebd.).

5.6.2. Studienpopulation

Eine weitere Limitation ergibt sich aus der geringen Stichprobengröße, welche aus $n = 15$ Teilnehmern besteht. Bei der Erstellung der DAS-18 Parallelformen wurden 204 Probanden (davon 110 gesunde und 94 depressive Probanden) für die Erprobung der Versionen rekrutiert (Rojas et al., 2014). Weissman & Beck (1978) erprobten ihre DAS-40 Parallelformen an 70 Probanden. In der aktuellen Arbeit sollten ausschließlich gesunde Teilnehmer rekrutiert werden, was sich jedoch aufgrund einer fehlenden Aufwandsentschädigung bei doch nicht unerheblichem Zeitaufwand der Untersuchung als schwierig erwies und daher einen längeren Zeitraum als geplant in Anspruch nahm. Die geringe Stichprobengröße bedingt ein größeres Konfidenzintervall und damit eine geringere Präzision der Ergebnisse. Somit können statistische Tests, beispielweise Mittelwertvergleiche, zunächst nur eingeschränkt beurteilt werden. Auch die aus den Daten hervorgehenden Itemschwierigkeiten und Trennschärfen sind bei einer kleinen Stichprobengröße weniger aussagekräftig als dies bei einer großen Stichprobe der Fall wäre. Darüber hinaus wurden die Parallelversionen an einer Stichprobe gesunder Teilnehmer erprobt, sollen aber zukünftig für Studien mit depressiven Probanden genutzt werden. Die

Vergleichbarkeit der hier erstellten Versionen kann somit streng genommen zunächst nur für gesunde Probanden beurteilt werden. Letztlich wäre auch die Erprobung an einer Stichprobe depressiver Probanden denkbar, die man direkt in der Klinik rekrutieren und untersuchen könnte. Zusätzlich ist kritisch anzumerken, dass die finale Stichprobe fast ausschließlich aus Personen mit hohem Bildungsniveau bestand und auch dadurch für zukünftige Studien möglicherweise nicht als repräsentativ betrachtet werden kann. In der Studie von Rojas et al. (2014) wurden sowohl depressive Probanden als auch gesunde Probanden mit beiden Parallelversionen der DAS getestet, wobei die Stichprobe der gesunden Probanden aus Personen verschiedener sozialer Schichten, Berufen und Altersstufen bestand (Rojas et al., 2014). Konkrete Angaben zur Repräsentativität ihrer Stichproben im Hinblick auf das erwartete Patientenkollektiv machen die Autoren jedoch nicht (ebd.). Für die Erprobung der ursprünglichen 40-Item-Parallelversionen fand die Testung an einer Stichprobe aus gesunden Studenten statt und damit ähnlich wie bei dieser Studie nicht an einer repräsentativen Personengruppe (Weissmann & Beck, 1978).

5.6.3. Erprobung der Parallelversionen

Bei der Erprobung der Parallelversionen wurde versucht, möglichst ähnliche Bedingungen für jeden Teilnehmer bzw. an beiden Testtagen zu gestalten. Dies war jedoch nur eingeschränkt möglich, da die Testungen in verschiedenen Räumen, zumeist bei den Teilnehmern zu Hause oder am Arbeitsplatz, stattfanden. Zusätzlich war es aufgrund von veränderten Arbeitszeiten, bereits geplanten Terminen oder ähnlichem nicht immer möglich, dass beide Testungen zur gleichen Uhrzeit stattfanden. Darüber hinaus war der Zeitraum zwischen den Untersuchungen nicht bei allen Teilnehmern gleich lang. Auch dadurch könnten letztlich Unterschiede in den Ergebnissen der beiden Versionen entstanden sein. Je länger der Zeitraum zwischen den Testungen, desto größer wird die Wahrscheinlichkeit, dass sich unsystematische Veränderungen der Merkmalsausprägung der Teilnehmer ergeben und gemessene Unterschiede auf diesen tatsächlichen Veränderungen beruhen (Schermelleh-Engel & Werner, 2012). Dennoch kann auch bei kurzen Zeiträumen zwischen den Testungen nicht ausgeschlossen werden, dass unterschiedliche Ergebnisse in den Versionen auf Faktoren wie Tagesform, persönlicher

Motivation oder unsystematischen Veränderungen in der Merkmalsausprägung beruhen. Schließlich muss auch der geplante Ablauf der Testungen kritisch hinterfragt werden, da am ersten Untersuchungstag der Bearbeitung des eigentlichen Testverfahrens eine etwa einstündige neuropsychologische Testung vorausging, am zweiten Tag jedoch nur der MET bearbeitet werden musste. Die neuropsychologische Testung hätte demnach an einem gesonderten Tag stattfinden können, was sich aber wahrscheinlich negativ auf die Bereitschaft der Teilnahme an der Studie ausgewirkt hätte. Positiv hervorzuheben ist, dass alle Probanden durch denselben Untersucher instruiert und getestet wurden und somit Effekte, die durch verschiedene Untersucher entstehen könnten, vermieden wurden. Darüber hinaus fand die Präsentation der Versionen in randomisierter Reihenfolge und die Präsentation der einzelnen Items innerhalb der Blöcke zufällig statt. Vergleicht man die Testbedingungen bei der Erstellung der MET-Parallelversionen mit denen anderer Studien, so zeigt sich, dass beispielsweise bei der Erstellung der DAS-18-Parallelversionen die Teilnehmer alle Fragebögen per Internetplattform von zu Hause bearbeiteten (Rojas et al., 2014). Dabei gab es lediglich für jede Parallelversion zwei verschiedene Fragebogenversionen, um einen möglichen Einfluss der Präsentationsreihenfolge der Items zu verhindern (ebd.). Ein solches Vorgehen hat den Vorteil eines geringeren Aufwandes für die Probanden, da die Testungen nebenbei im Alltag erfolgen können und somit vermutlich entsprechend mehr Probanden in kürzerer Zeit rekrutiert werden können. Andererseits gibt es keine Kontrolle darüber, in welchem situativen Kontext, zu welcher Tageszeit und durch wen die Testungen tatsächlich erfolgen. Für unsere Studie, in welcher sich die Probanden auf die präsentierten emotionalen Inhalte einlassen sollten, wäre ein solches Vorgehen eher nicht geeignet. Hinzu kommt, dass die neuropsychologische Testung einen zusätzlichen Untersucher erfordert und nicht als Selbstbeurteilungsinstrument von den Teilnehmern bearbeitet werden kann. Bei der Erprobung der DAS-40-Parallelversionen bearbeiteten die Probanden im Abstand von einer Woche Version A und anschließend Version B (Weissmann & Beck, 1978). Eine randomisierte Präsentation oder Variation der Items fand im Gegensatz zu dieser Arbeit nicht statt (ebd.).

5.7. Vergleich der initialen und finalen Parallelversionen

Abschließend sollen die beiden im Rahmen dieser Arbeit entstandenen Parallelversionen vergleichend diskutiert werden, um auf diese Weise mögliche Ursachen für die erhaltenen Ergebnisse aufzuzeigen und somit Schlussfolgerungen für die Erstellung zukünftiger Parallelversionen zu ziehen. Zwei Versionen des Multidimensionalen Empathietests wurden für diese Arbeit als parallel angesehen, wenn die Mittelwerte keine statistisch signifikanten Unterschiede aufwiesen und die Versionen gleichzeitig signifikant miteinander korrelierten. Diese Kriterien werden nicht von den initialen, sondern ausschließlich von den finalen MET-Versionen erfüllt. Hierbei könnten die Ergebnisse sowohl von der Testerstellung bzw. Aufteilung der Items auf die Versionen selbst als auch von den Testbedingungen und der für die Erprobung gewählten Stichprobe abhängen.

Die Erstellung der initialen Parallelversionen (MET-A, MET-B) erfolgte anhand von Schwierigkeit und Trennschärfe bereits existierender Datensätzen, von denen keine konkreten klinischen oder soziodemographischen Daten vorliegen. Jedoch ist zu beachten, dass die Stichprobe von Lemme (2012) aus Personen mit niedrigem Bildungsabschluss sowie aus Personen mit einer Erkrankung aus dem Autismus-Spektrum bestand. Die Stichprobe aus dem NeMUP-Projekt setzte sich aus pädophilen Probanden zusammen. Die Parallelversionen wurden anhand von Schwierigkeit und Trennschärfe gebildet, wobei davon auszugehen ist, dass sich die Schwierigkeiten in einer gesunden Stichprobe von denen in Stichproben mit pädophilen oder autistischen Personen unterscheiden. In der Studie von Schuler et al. (2019) zeigte sich beispielsweise, dass sich pädophile Männer je nach Alter der dargestellten Person in ihrer empathischen Reaktion im Vergleich zu gesunden Probanden unterschieden. So zeigten bereits übergriffig gewordene pädophile Männer im Vergleich zur gesunden Kontrollgruppe eine signifikant schlechtere kognitive Empathie bei abgebildeten Kindern (Schuler et al., 2019). Die Notwendigkeit der Anpassung nach der ersten Erprobung erscheint unter diesen Gesichtspunkten nachvollziehbar. Für die Erstellung der finalen Parallelversionen lagen hingegen Daten einer homogenen Gruppe gesunder Probanden vor, welche zuvor klinisch und neuropsychologische Testverfahren durchlaufen mussten und somit gesichert keine gröberen Defizite aufwiesen.

Damit zwei Parallelversionen tatsächlich gleiche Werte messen, müssen außerdem auch die Testbedingungen möglichst gleich sein. Bei der Erprobung der initialen Parallelversionen war der Zeitraum zwischen der Bearbeitung der beiden Versionen im Schnitt länger als bei den finalen Versionen (MW $3,90 \pm 3,75$ Tage in den initialen vs. MW $2,07 \pm 1,90$ Tage in den finalen Versionen). Tatsächliche Änderungen in der Merkmalsausprägung wurden hierdurch wahrscheinlicher und könnten sich ebenfalls negativ auf die Bestimmung der Parallelität in den initialen Versionen ausgewirkt haben. Eine kurze, einheitliche Zeitspanne wäre für zukünftige Paralleltesterstellung empfehlenswert.

Schließlich unterscheiden sich auch die Stichproben, anhand derer die Parallelversionen erprobt wurden. Die initialen Parallelversionen wurden an $n = 20$ Probanden erprobt, die finalen Versionen an $n = 15$ Probanden. Durch die größere Stichprobe sind die Ergebnisse der Erprobung der initialen Versionen prinzipiell aussagekräftiger als die der finalen Versionen. So wäre es möglich, dass auch die Unterschiede der finalen Parallelversionen signifikant sind, die gewählte Stichprobe jedoch zu klein ist, um diese nachzuweisen. Grundsätzlich sind die Ergebnisse beider Erprobungen durch die geringen Probandenzahlen nur unter Vorbehalt zu werten.

Hinsichtlich der berechneten Schwierigkeiten und Trennschärfen zeigten sich trotz identischer Items ebenfalls Unterschiede in den beiden Versionen. So sind die berechneten Schwierigkeiten aus der Erprobung der initialen Versionen höher und die Items damit durchschnittlich leichter als in der Erprobung der finalen Parallelversionen. Die Trennschärfe der Items in den finalen Parallelversionen ist höher und somit besser als in den initialen Versionen. Auch die interne Konsistenz ist in den finalen MET-Versionen höher als in den initialen Versionen. Vergleicht man die demographischen Daten der betreffenden Stichproben, so zeigt sich, dass in der ersten Stichprobe der prozentuale Anteil der männlichen Probanden größer ist als in der zweiten Stichprobe (25 % vs. 13,3 %), außerdem sind die Teilnehmer der ersten Stichprobe durchschnittlich älter als die in der zweiten Stichprobe (MW $41,5 \pm 12,7$ Jahre vs. MW $30,2 \pm 7,4$ Jahre). In beiden Stichproben hat die Mehrheit der Teilnehmer einen hohen Bildungsgrad (Anteil der Probanden mit Abitur/Fachabitur 90 % vs. 93,3 % und Anteil der Probanden mit Abschluss an einer Fachhochschule/Universität 70 % vs. 60 %). Auch die Ergebnisse der klinischen und neuropsychologischen Tests fielen ähnlich aus. Daraus kann geschlussfolgert werden, dass sich bereits geringe Unterschiede zwischen den Probanden auf Itemkennwerte wie

Schwierigkeit und Trennschärfe auswirken können. Diese Werte sind demnach nicht als absolut anzusehen, sondern lediglich als orientierende Größen zu interpretieren. Dies sollte auch bei der Aufteilung von Items auf künftige Parallelversionen berücksichtigt werden.

5.8. Fazit und Ausblick

Die bisherigen Ergebnisse, insbesondere die Korrelationen und Mittelwertvergleiche, weisen darauf hin, dass es sich in den zunächst im Rahmen dieser Arbeit erstellten Versionen MET-A und MET-B nicht um parallele Testverfahren handelt. In den daraufhin angepassten Parallelversionen MET-A* und MET-B* konnte eine bessere Vergleichbarkeit anhand von Korrelationen und Mittelwerten gezeigt werden. Die in dieser Arbeit zum Teil geringen Trennschärfekoeffizienten und Reliabilitäten sollten in zukünftigen Messungen erneut untersucht werden. Bei Bestätigung der hier erhobenen Ergebnisse muss über eine Anpassung bzw. Erweiterung des Bildmaterials diskutiert werden. Hierzu wird die erneute Erprobung der Parallelversionen MET-A* und MET-B* anhand einer größeren Stichprobe empfohlen, insbesondere auch um die bisherigen Ergebnisse im Hinblick auf das Vorliegen der Parallelität der Versionen sowie der Reliabilität zu überprüfen. Zusammenfassend kann jedoch nach bisherigem Kenntnisstand von einer Parallelität der MET-Versionen A* und B* ausgegangen werden.

Die Parallelversionen MET-A* und MET-B* werden aktuell im Rahmen der Studie „Empathiedefizit chronisch depressiver Patienten nach emotionaler Aktivierung“ unter Leitung von PD Dr. med. Stephan Köhler und Dr. rer. nat. Anne Guhn in der Klinik für Psychiatrie und Psychotherapie der Charité Universitätsmedizin Berlin verwendet. Hierbei wird untersucht, ob Defizite des Empathievermögens ein spezifisches Merkmal der chronischen Depression sind und ob diese im Zusammenhang mit traumatischen Kindheitserfahrungen stehen. Das ist insofern von Bedeutung, da wie eingangs beschrieben eine spezifische Therapieform (CBASP) existiert, die gezielt eine Verbesserung dieser Empathiedefizite und damit einher gehender interpersoneller Schwierigkeiten bewirken soll (McCullough, 2000). Eine weitere Fragestellung der Studie ist es demnach, ob eine stationäre CBASP-Behandlung tatsächlich eine Verbesserung der Empathiefähigkeit

bewirkt und ob diese in Verbindung mit einer Verbesserung der depressiven Symptomatik steht. Hierzu werden 20 episodisch depressive, 20 chronisch depressive und 20 gesunde Probanden rekrutiert. Für alle Probanden sollen zwei auditive Skripte erstellt werden, die auf individuellen Prägungen wichtiger Bezugspersonen beruhen. Dabei soll ein Skript auf einem emotional belastenden interpersonellen Ereignis aufbauen und als Kontrollbedingung ein weiteres Skript eine neutrale Erinnerung beinhalten, welche keine emotionale Aktivierung bewirkt. Nach der Präsentation der Skripte wird jeweils eine Parallelversion des MET (MET-A*, MET-B*) bearbeitet, um zu untersuchen, ob sich eine emotionale Aktivierung auf das Empathievermögen auswirkt und ob dies wie vermutet nur auf die Gruppe der chronisch depressiven Probanden zutrifft. Zusätzlich wird das kognitive Empathievermögen ohne emotionale Aktivierung mit dem „Reading the Mind in the Eyes Test“ gemessen (Baron-Cohen et al., 2001). Zur Beantwortung der zweiten Fragestellung durchlaufen die rekrutierten chronisch depressiven Probanden nach der Empathietestung mit und ohne vorherige emotionale Aktivierung ein 12-wöchiges stationäres CBASP-Programm (Brakemeier & Normann, 2012) und werden im Anschluss an dieses erneut mittels derselben Skripte getestet.

Die beschriebene Studie ist nur ein konkretes Beispiel dafür, wie die im Rahmen dieser Arbeit erstellten Parallelversionen in Zukunft dabei helfen können, die Psychopathologie der Depression besser zu verstehen, differentielle Therapieindikationen zu stellen und die Einschränkungen des Funktionsniveaus der betroffenen Patienten zu reduzieren.

6. Literaturverzeichnis

- Altmann, T. (2014). Empathie. In M. A. Wirtz (Hrsg), *Dorsch - Lexikon der Psychologie* (18. Aufl., S. 447). Bern, Schweiz: Hogrefe Verlag.
- Aschenbrenner, S., Lange, K. W., & Tucha, O. (2000). *RWT : Regensburger Wortflüssigkeits-Test*. Göttingen, Deutschland: Hogrefe Verlag für Psychologie.
- Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. *Journal of psychosomatic research, 38*(1), 23-32.
- Bagby, R. M., Taylor, G. J., Parker, J. D., & Dickens, S. E. (2006). The development of the Toronto Structured Interview for Alexithymia: item selection, factor structure, reliability and concurrent validity. *Psychotherapy and psychosomatics, 75*(1), 25-39. <https://doi.org/10.1159/000089224>
- Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J Autism Dev Disord, 34*(2), 163-175. <https://doi.org/10.1023/B:JADD.0000022607.19833.00>
- 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. *J Child Psychol Psychiatry, 42*(2), 241-251. <https://doi.org/10.1017/S0021963001006643>
- 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*(3), 413. <https://doi.org/10.1037/0022-3514.61.3.413>
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *San Antonio, 78*(2), 490-498.
- Bernstein, D. P., Fink, L., Handelsman, L., Foote, J., Lovejoy, M., Wenzel, K., Sapareto, E., & Ruggiero, J. (1994). Initial reliability and validity of a new retrospective measure of child abuse and neglect. *The American journal of psychiatry, 151*(8), 1132-1136. <https://doi.org/10.1176/ajp.151.8.1132>
- Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., Stokes, J., Handelsman, L., Medrano, M., & Desmond, D. (2003). Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child abuse & neglect, 27*(2), 169-190. [https://doi.org/10.1016/S0145-2134\(02\)00541-0](https://doi.org/10.1016/S0145-2134(02)00541-0)
- 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*(2), 33-41.

- Blanco, C., Okuda, M., Markowitz, J. C., Liu, S.-M., Grant, B. F., & Hasin, D. S. (2010). The epidemiology of chronic major depressive disorder and dysthymic disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *The Journal of clinical psychiatry*, *71*(12), 1645.
<https://doi.org/10.4088/JPC.09m05663gry>
- Bora, E., & Berk, M. (2016). Theory of mind in major depressive disorder: A meta-analysis. *Journal of affective disorders*, *191*, 49-55.
<https://doi.org/10.1016/j.jad.2015.11.023>
- Brakemeier, E.-L., & Normann, C. (2012). *Praxisbuch CBASP: Behandlung chronischer Depression. Mit Online-Materialien*. Weinheim, Deutschland: Beltz.
- Brakemeier, E.-L., Schramm, E., & Hautzinger, M. (2012). *Chronische Depression* (1. Aufl.). Göttingen, Deutschland: Hogrefe Verlag.
- Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the national Academy of Sciences*, *100*(9), 5497-5502. <https://doi.org/10.1073/pnas.0935845100>
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology*, *76*(6), 893.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Curwen, T. (2003). The importance of offense characteristics, victimization history, hostility, and social desirability in assessing empathy of male adolescent sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, *15*(4), 347-364.
<https://doi.org/10.1023/A:1025056312935>
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85.
- De Gelder, B., Snyder, J., Greve, D., Gerard, G., & Hadjikhani, N. (2004). Fear fosters flight: a mechanism for fear contagion when perceiving emotion expressed by a whole body. *Proceedings of the National Academy of Sciences*, *101*(47), 16701-16706. <https://doi.org/10.1073/pnas.0407042101>
- Decety, J., & Lamm, C. (2006). Human empathy through the lens of social neuroscience. *The scientific World journal*, *6*, 1146-1163.
<https://doi.org/10.1100/tsw.2006.221>
- Derntl, B., Seidel, E.-M., Schneider, F., & Habel, U. (2012). How specific are emotional deficits? A comparison of empathic abilities in schizophrenia, bipolar and depressed patients. *Schizophrenia research*, *142*(1-3), 58-64.
<https://doi.org/10.1016/j.schres.2012.09.020>

- Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., Kessler, J., Woike, J. K., Wolf, O. T., & Convit, A. (2006). Introducing MASC: a movie for the assessment of social cognition. *J Autism Dev Disord*, *36*(5), 623-636. <https://doi.org/10.1007/s10803-006-0107-0>
- Dziobek, I., Preißler, S., Grozdanovic, Z., Heuser, I., Heekeren, H. R., & Roepke, S. (2011). Neuronal correlates of altered empathy and social cognition in borderline personality disorder. *Neuroimage*, *57*(2), 539-548. <https://doi.org/10.1016/j.neuroimage.2011.05.005>
- Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., & Convit, A. (2008). Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the Multifaceted Empathy Test (MET). *J Autism Dev Disord*, *38*(3), 464-473. <https://doi.org/10.1007/s10803-007-0486-x>
- Eisenberg, N., Fabes, R. A., Murphy, B., Karbon, M., Maszk, P., Smith, M., O'Boyle, C., & Suh, K. (1994). The relations of emotionality and regulation to dispositional and situational empathy-related responding. *Journal of personality and social psychology*, *66*(4), 776.
- Gilmer, W. S., Gollan, J. K., Wisniewski, S. R., Howland, R. H., Trivedi, M. H., Miyahara, S., Fleck, J., Thase, M. E., Alpert, J. E., & Nierenberg, A. A. (2008). Does the duration of index episode affect the treatment outcome of major depressive disorder? A STAR* D report. *The Journal of clinical psychiatry*, *69*(8), 1246-1256. <https://doi.org/10.4088/jcp.v69n0807>
- Grynberg, D., Chang, B., Corneille, O., Muraige, P., Vermeulen, N., Berthoz, S., & Luminet, O. (2012). Alexithymia and the processing of emotional facial expressions (EFEs): systematic review, unanswered questions and further perspectives. *PloS ONE*, *7*(8), Article e42429. <https://doi.org/10.1371/journal.pone.0042429>
- Gulliksen, H. (2013). *Theory of mental tests*. New York, NY: Routledge.
- Hamilton, M. (1960). A rating scale for depression. *J Neurol Neurosurg Psychiatry*, *23*, 56-62. <https://dx.doi.org/10.1136%2Fjnnp.23.1.56>
- Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of autism and Developmental disorders*, *24*(2), 129-154. <https://doi.org/10.1007/BF02172093>
- Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current directions in psychological science*, *2*(3), 96-100.
- Hautzinger, M., Joormann, J., & Keller, F. (2005). *Die Skala dysfunktionaler Einstellungen (DAS)*. Göttingen, Deutschland: Hogrefe Verlag.

- Haviland, M. G. (1996). Structure of the twenty-item Toronto Alexithymia Scale. *Journal of personality assessment*, 66(1), 116-125.
https://doi.org/10.1207/s15327752jpa6601_9
- Helmstaedter, C., & Durwen, H. (1990). VLMT: Verbaler Lern-und Merkfähigkeitstest: Ein praktikables und differenziertes Instrumentarium zur Prüfung der verbalen Gedächtnisleistungen. *Schweizer Archiv für Neurologie, Neurochirurgie und Psychiatrie*.
- Heuft, G., Senf, W., Bell, K., Cording, C., Geyer, M., Janssen, P. L., Lamprecht, F., Meermann, R., Strauß, B., & Wirsching, M. (1998). Psy-BaDo. *Psychotherapeut*, 43(1), 48-52. <https://doi.org/10.1007/s002780050099>
- Häuser, W., Schmutzer, G., Brähler, E., & Glaesmer, H. (2011). Maltreatment in childhood and adolescence: results from a survey of a representative sample of the German population. *Deutsches Ärzteblatt international*, 108(17), 287.
<https://dx.doi.org/10.3238%2Farztebl.2011.0287>
- Hölzel, L., Härter, M., Reese, C., & Kriston, L. (2011, Mar). Risk factors for chronic depression--a systematic review. *J Affect Disord*, 129(1-3), 1-13.
<https://doi.org/10.1016/j.jad.2010.03.025>
- Ireland, J. L. (1999). Provicictim attitudes and empathy in relation to bullying behaviour among prisoners. *Legal and Criminological Psychology*, 4(1), 51-66.
<https://doi.org/10.1348/135532599167789>
- Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, 36(3), 347-387. <https://doi.org/10.1207/S15327906347-387>
- Kelava, A., & Moosbrugger, H. (2012). Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilungen. In *Testtheorie und Fragebogenkonstruktion* (S. 75-102). Berlin, Deutschland: Springer Verlag.
- Köhler, S., Sterzer, P., Normann, C., Berger, M., & Brakemeier, E. L. (2016). Überwindung der Therapieresistenz bei chronischer Depression. *Der Nervenarzt*, 87(7), 701-707. <https://doi.org/10.1007/s00115-015-0034-4>
- Kühner, C., Bürger, C., Keller, F., & Hautzinger, M. (2007). Reliability and validity of the revised Beck Depression Inventory (BDI-II). Results from German samples. *Der Nervenarzt*, 78(6), 651-656. <https://doi.org/10.1007/s00115-006-2098-7>
- Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of nonverbal behavior*, 27(3), 145-162.
<https://doi.org/10.1023/A:1025389814290>

- Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: reliability and validity of the Empathy Quotient. *Psychological medicine*, 34(5), 911-920. <https://doi.org/10.1017/S0033291703001624>
- Lehrl, S. (1999). *Mehrfachwahl-Wortschatz-Intelligenztest: MWT-B*. (5. Aufl.). Balingen, Deutschland: Spitta.
- Lehrl, S., Triebig, G., & Fischer, B. (1995). Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. *Acta Neurologica Scandinavica*, 91(5), 335-345. <https://doi.org/10.1111/j.1600-0404.1995.tb07018.x>
- Lemme, B. (2012). *Entwicklung und Validierung einer neuen Adaption des Multidimensionalen Empathie Test (MET) zur Untersuchung kognitiv- und emotional-empathischer Reaktionen auf Kinder vs. Erwachsene*. (Unpublished bachelor's thesis). Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany.
- Mattern, M., Walter, H., Hentze, C., Schramm, E., Drost, S., Schoepf, D., Fangmeier, T., Normann, C., Zobel, I., & Schnell, K. (2015). Behavioral evidence for an impairment of affective theory of mind capabilities in chronic depression. *Psychopathology*, 48(4), 240-250. <https://doi.org/10.1159/000430450>
- McCullough, J. P. (2000). *Treatment for chronic depression: Cognitive behavioral analysis system of psychotherapy*. Guilford Press. <https://doi.org/10.1037/1053-0479.13.3-4.241>. <https://doi.org/10.1037/0894-4105.16.3.327>
- Minshew, N. J., Meyer, J., & Goldstein, G. (2002). Abstract reasoning in autism: A disassociation between concept formation and concept identification. *Neuropsychology*, 16(3), 327. <https://doi.org/10.1037/0894-4105.16.3.327>
- Nemiah, J. C., & Sifneos, P. E. (1970). Psychosomatic illness: a problem in communication. *Psychotherapy and psychosomatics*, 18(1-6), 154-160. <https://doi.org/10.1159/000286074>
- Parker, J., Taylor, G., & Bagby, M. (1993). Alexithymia and the processing of emotional stimuli: an experimental study. *New Trends in Experimental and Clinical Psychiatry*, 9, 9-9.
- Paulus, C. (2009). Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index.
- Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in cognitive sciences*, 3(9), 337-344. [https://doi.org/10.1016/S1364-6613\(99\)01362-5](https://doi.org/10.1016/S1364-6613(99)01362-5)

- Pfaltz, M. C., McAleese, S., Saladin, A., Meyer, A. H., Stoecklin, M., Opwis, K., & Martin-Soelch, C. (2013). The Reading the Mind in the Eyes Test: Test-retest reliability and preliminary psychometric properties of the German version. *International Journal of Advances in Psychology*, 2(1), 1-9.
- Popp, K., Schäfer, R., Schneider, C., Brähler, E., Decker, O., Hardt, J., & Franz, M. (2008). Faktorstruktur und Reliabilität der Toronto-Alexithymie-Skala (TAS-20) in der deutschen Bevölkerung. *PPmP-Psychotherapie· Psychosomatik· Medizinische Psychologie*, 58(05), 208-214. <https://doi.org/10.1055/s-2007-986196>
- Poustka, L., Rehm, A., Holtmann, M., Bock, M., Böhmert, C., & Dziobek, I. (2010). Dissoziation von kognitiver und affektiver Empathie bei Jugendlichen mit Autismus-Spektrum-Störungen. *Kindheit und Entwicklung*, 19, 177-183. <https://doi.org/10.1026/0942-5403/a000022>
- Power, M., Katz, R., McGuffin, P., Duggan, C., Lam, D., & Beck, A. (1994). The Dysfunctional Attitude Scale (DAS): A comparison of forms A and B and proposals for a new subscaled version. *Journal of research in personality*, 28(3), 263-276. <https://doi.org/10.1006/jrpe.1994.1019>
- Preston, S. D., & De Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and brain sciences*, 25(1), 1-20. <https://doi.org/10.1017/S0140525X02000018>
- Reitan, R. M. (1992). *Trail making test*. Tucson, AZ: Reitan Neuropsychology Laboratory.
- Ritter, K., Dziobek, I., Preissler, S., Ruter, A., Vater, A., Fydrich, T., Lammers, C. H., Heekeren, H. R., & Roepke, S. (2011). Lack of empathy in patients with narcissistic personality disorder. *Psychiatry Res*, 187(1-2), 241-247. <https://doi.org/10.1016/j.psychres.2010.09.013>
- Rojas, R., Geissner, E., & Hautzinger, M. (2014). DAS-18 Form A und Form B: Entwicklung und psychometrische Überprüfung von zwei vergleichbaren Kurzversionen der Skala Dysfunktionaler Einstellungen (DAS). *Diagnostica*, 61(4), 173-183. <https://doi.org/10.1026/0012-1924/a000123>
- Schermelleh-Engel, K., & Werner, C. (2012). Methoden der Reliabilitätsbestimmung. In *Testtheorie und Fragebogenkonstruktion* (S. 119-141). Berlin, Deutschland: Springer Verlag.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350-353. <https://psycnet.apa.org/doi/10.1037/1040-3590.8.4.350>
- Schreiter, S., Pijnenborg, G. H., & Aan Het Rot, M. (2013). Empathy in adults with clinical or subclinical depressive symptoms. *J Affect Disord*, 150(1), 1-16. <https://doi.org/10.1016/j.jad.2013.03.009>

- Schuler, M., Mohnke, S., Amelung, T., Dziobek, I., Lemme, B., Borchardt, V., Gerwin, H., Kärgel, C., Kneer, J., & Massau, C. (2019). Empathy in pedophilia and sexual offending against children: A multifaceted approach. *Journal of abnormal psychology, 128*(5), 453. <https://doi.org/10.1037/abn0000412>
- Seifritz, E., Esposito, F., Neuhoff, J. G., Lüthi, A., Mustovic, H., Dammann, G., Von Bardeleben, U., Radue, E. W., Cirillo, S., & Tedeschi, G. (2003). Differential sex-independent amygdala response to infant crying and laughing in parents versus nonparents. *Biological psychiatry, 54*(12), 1367-1375. [https://doi.org/10.1016/S0006-3223\(03\)00697-8](https://doi.org/10.1016/S0006-3223(03)00697-8)
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry, 59 Suppl 20*, 22-33;quiz 34-57.
- Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment, 80*(1), 99-103. https://doi.org/10.1207/S15327752JPA8001_18
- Thoma, P., Zalewski, I., von Reventlow, H. G., Norra, C., Juckel, G., & Daum, I. (2011). Cognitive and affective empathy in depression linked to executive control. *Psychiatry research, 189*(3), 373-378. <https://doi.org/10.1016/j.psychres.2011.07.030>
- Tischler, L., & Petermann, F. (2010). Trail Making Test (TMT). *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie, 58*(1), 79-81.
- van Randenborgh, A., Hüffmeier, J., Victor, D., Klocke, K., Borlinghaus, J., & Pawelzik, M. (2012). Contrasting chronic with episodic depression: an analysis of distorted socio-emotional information processing in chronic depression. *Journal of Affective Disorders, 141*(2-3), 177-184. <https://doi.org/10.1016/j.jad.2012.02.039>
- Wechsler, D. (1956). Die Messung der Intelligenz Erwachsener. Textband zum Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE); Deutsche Bearbeitung Anne von Hardesty, und Hans Lauber [The measurement of adult intelligence]. Bern, Schweiz: Hans Huber.
- Weissman, A. N., & Beck, A. T. (1978). *Development and validation of the Dysfunctional Attitude Scale: A preliminary investigation*. Annual Meeting of The American Educational Research Association, Toronto.
- Wiersma, J. E., Hovens, J. G., van Oppen, P., Giltay, E. J., van Schaik, D. J., Beekman, A. T., & Penninx, B. W. (2009). The importance of childhood trauma and childhood life events for chronicity of depression in adults. *J Clin Psychiatry, 70*(7), 983-989. <https://doi.org/10.4088/JPC.08m04521>
-






- Wilbertz, G., Brakemeier, E.-L., Zobel, I., Härter, M., & Schramm, E. (2010). Exploring preoperational features in chronic depression. *Journal of affective disorders*, 124(3), 262-269. <https://doi.org/10.1016/j.jad.2009.11.021>
- Volkenstein, L., Schönenberg, M., Schirm, E., & Hautzinger, M. (2011). I can see what you feel, but I can't deal with it: Impaired theory of mind in depression. *Journal of affective disorders*, 132(1-2), 104-111. <https://doi.org/10.1016/j.jad.2011.02.010>
- Yamada, K., Inoue, Y., & Kanba, S. (2015). Theory of mind ability predicts prognosis of outpatients with major depressive disorder. *Psychiatry research*, 230(2), 604-608. <https://doi.org/10.1016/j.psychres.2015.10.011>

7. Anhang

Anhang 1: Bildmaterial des MET-KE

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
1		ausgelassen	erfreut	amüsiert	liebvoll
2		stolz	ausgelassen	interessiert	neugierig
3		vergnügt	stolz	erleichtert	erfreut
4		verzweifelt	bockig	betrübt	enttäuscht

ANHANG

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
5		missmutig	schmerzerfüllt	verwirrt	eifersüchtig
6		besorgt	panisch	melancholisch	angeekelt
7		fasziniert	begeistert	überrascht	genießend
8		fröhlich	übermütig	entschlossen	verlegen
9		zufrieden	triumphierend	erleichtert	überrascht

ANHANG

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
10		beglückt	interessiert	zufrieden	amüsiert
11		enttäuscht	wütend	bockig	angstvoll
12		freudlos	entsetzt	durcheinander	schmerzerfüllt
13		genüsslich	freudestrahlend	zufrieden	entspannt
14		heiter	berauscht	stolz	interessiert

ANHANG

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
15		entspannt	stolz	vergnügt	schwärmerisch
16		genervt	verärgert	besorgt	missmutig
17		besorgt	verängstigt	bockig	neidisch
18		betrübt	verzweifelt	verlegen	angeekelt
19		bedrückt	untröstlich	beschämt	beleidigt

ANHANG

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
20		beunruhigt	verachtend	beleidigt	angstvoll
21		triumphierend	zufrieden	entzückt	erleichtert
22		belustigt	stolz	erstaunt	entspannt
23		entspannt	erfreut	dankbar	befreit
24		entgeistert	betreten	erstaunt	genervt

ANHANG

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
25		traurig	geknickt	grüblerisch	trotzig
26		triumphierend	ergriffen	glücklich	dankbar
27		zufrieden	ekstatisch	entspannt	erwartungsvoll
28		angstvoll	besorgt	panisch	traurig
29		frustriert	schmerzerfüllt	betrübt	traurig

ANHANG

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
30		gepeinigt	verstört	entmutigt	neidisch
31		euphorisch	stolz	erfreut	siegessicher
32		genießend	entspannt	schwärmend	amüsiert
33		hochzufrieden	euphorisch	gelassen	vergebend
34		gequält	verstört	teilnahmslos	unglücklich
35		gepeinigt	unwohl	bestürzt	angstvoll






ANHANG

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
36		amüsiert	überschäumend	berührt	genüsslich
37		heiter	enthusiastisch	neugierig	verliebt
38		hoffnungslos	verschreckt	durcheinander	ungeduldig
39		entsetzt	wütend	schuldig	argwöhnisch
40		niedergeschlagen	verlegen	besorgt	aufgebracht






ANHANG

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
41		innig	wohlwollend	unbekümmert	zufrieden
42		freudig	überwältigt	gerührt	friedlich
43		ausgelassen	euphorisch	gelassen	entspannt
44		betrübt	verwirrt	bockig	missmutig
45		bekümmert	erschreckt	panisch	verwirrt

ANHANG

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
46		verzweifelt	betrübt	angstvoll	verlegen
47		amüsiert	ausgelassen	interessiert	überrascht
48		ausgelassen	jubilierend	zufrieden	entspannt
49		glücklich	euphorisch	fasziniert	dankbar
50		genüsslich	überglücklich	innig	erleichtert

ANHANG

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
51		frustriert	verzweifelt	besorgt	eifersüchtig
52		melancholisch	bockig	verzweifelt	ängstlich
53		geknickt	hilflos	verwirrt	schockiert
54		eingeschüchtert	widerwillig	unzufrieden	hasserfüllt
55		angeekelt	verbittert	schuldig	einsam

ANHANG

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
56		verlegen	überwältigt	bezaubert	sorglos
57		vergnügt	zufrieden	berauscht	fasziniert
58		stolz	zufrieden	übermütig	selbstgefällig
59		traurig	schuldig	eingeschüchtert	änstlich
60		traurig	aufgebracht	beleidigt	erschrocken

ANHANG

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
61		schwärmerisch	ekstatisch	zufrieden	jubilend
62		liebend	beschwingt	erleichtert	aufmerksam
63		verlegen	verblüfft	euphorisch	träumerisch
64		überrascht	panisch	interessiert	verliebt
65		kummervoll	verlegen	enttäuscht	eifersüchtig

ANHANG

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
66		fasziniert	interessiert	stolz	belustigt
67		interessiert	überwältigt	gemütlich	gefasst
68		hoffnungslos	verzweifelt	gelangweilt	eifersüchtig
69		schmerzerfüllt	angstvoll	hoffnungslos	traurig
70		ängstlich	missmutig	angeekelt	enttäuscht

ANHANG

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
71		verängstigt	verzweifelt	überrascht	zögerlich
72		erheitert	entschlossen	erleichtert	triumphierend
73		ausgelassen	amüsiert	interessiert	fasziniert
74		verträumt	überschäumend	hingebungsvoll	beschwingt
75		flehend	verstimmt	melancholisch	traurig

ANHANG

Item Nr.	Bild	Target-Emotion	Distractor-Emotion 1	Distractor-Emotion 2	Distractor-Emotion 3
76		schmerzerfüllt	verärgert	freudlos	verzweifelt
77		wehmütig	vergnügt	erwartungsvoll	depressiv
78		erfreut	überwältigt	bewegt	wissbegierig
79		todunglücklich	betrübt	verbittert	mutlos
80		nachdenklich	todunglücklich	schuldbe-wusst	bedrängt

Quelle: von <http://www.met-ke.de/stimuli> (abgerufen am 01.04.2020)

Anhang 2: Itemkennwerte des MET-KE

Tabelle 16A: Itemkennwerte des MET-KE, Items 1-27.

Item	P_{NeMUP}	P_{Lemme}	T_{KN}	T_{Lemme}	T_{EN}
1	51,52	58,97	0,353	0,288	0,751
2	82,00	89,74	0,101	0,310	0,637
3	61,07	69,23	0,225	0,440	0,701
4	37,19	33,33	0,113	0,286	0,601
5	56,29	61,54	0,286	0,282	0,591
6	28,74	30,77	0,111	0,227	0,701
7	59,23	69,23	0,348	0,378	0,628
8	82,74	82,05	0,065	0,263	0,579
9	83,47	87,18	0,164	0,547	0,578
10	75,76	84,62	0,240	0,739	0,564
11	54,45	64,10	0,104	0,189	0,555
12	69,88	69,23	0,256	0,213	0,572
13	62,17	64,10	0,235	0,316	0,540
14	77,23	82,05	0,233	-0,118	0,630
15	79,80	76,92	0,294	0,450	0,586
16	66,21	69,23	0,026	0,286	0,658
17	66,94	51,28	0,173	0,182	0,630
18	73,92	79,49	0,328	0,526	0,575
19	78,32	69,23	0,158	0,343	0,681
20	67,31	71,79	0,168	0,196	0,644
21	92,65	89,74	0,389	0,521	0,622
22	62,53	69,23	0,001	0,137	0,679
23	89,35	84,62	0,177	0,382	0,704
24	47,84	64,10	0,187	0,294	0,623
25	53,35	64,10	0,168	0,178	0,713
26	88,98	89,74	0,362	0,600	0,618
27	45,64	74,36	0,133	0,293	0,639

P_{NeMUP} = Schwierigkeit in der NeMUP-Stichprobe, P_{Lemme} = Schwierigkeit in der Lemme-Stichprobe, T_{KN} = Trennschärfe der kognitiven Empathie in der NeMUP-Stichprobe, T_{Lemme} = Trennschärfe der kognitiven Empathie in der Lemme-Stichprobe, T_{EN} = Trennschärfe der emotionalen Empathie in der NeMUP-Stichprobe.

Tabelle 16B: Itemkennwerte des MET-KE, Items 28 – 54.

Item	P_{NeMUP}	P_{Lemme}	T_{KN}	T_{Lemme}	T_{EN}
28	43,80	38,46	0,158	0,109	0,664
29	69,88	82,05	0,093	0,500	0,693
30	29,11	46,15	0,346	0,409	0,646
31	58,49	56,41	0,327	0,364	0,591
32	52,98	69,23	0,098	0,267	0,615
33	58,13	71,79	0,302	0,513	0,677
34	51,52	56,41	0,228	0,395	0,614
35	27,27	41,03	0,308	0,361	0,687
36	39,76	46,15	0,119	0,293	0,660
37	73,19	64,10	0,287	0,393	0,693
38	49,31	66,67	0,196	0,492	0,662
39	51,15	48,72	0,185	0,274	0,563
40	57,76	76,92	0,029	0,382	0,653
41	38,29	58,97	0,364	0,492	0,666
42	74,66	56,41	0,121	0,263	0,672
43	40,86	48,72	0,109	0,445	0,661
44	41,60	58,97	0,053	0,197	0,622
45	40,50	46,15	0,136	0,293	0,603
46	24,70	51,28	0,010	0,380	0,591
47	37,19	56,41	0,230	0,253	0,665
48	73,92	89,74	0,390	0,509	0,650
49	78,70	76,92	0,257	0,352	0,652
50	90,82	94,87	0,222	0,350	0,556
51	57,76	69,23	0,317	0,274	0,723
52	28,74	53,85	0,208	0,143	0,671
53	38,29	56,41	0,052	0,346	0,666
54	64,37	69,23	0,212	0,455	0,590

P_{NeMUP} = Schwierigkeit in der NeMUP-Stichprobe, P_{Lemme} = Schwierigkeit in der Lemme-Stichprobe, T_{KN} = Trennschärfe der kognitiven Empathie in der NeMUP-Stichprobe, T_{Lemme} = Trennschärfe der kognitiven Empathie in der Lemme-Stichprobe, T_{EN} = Trennschärfe der emotionalen Empathie in der NeMUP-Stichprobe.

Tabelle 16C: Itemkennwerte des MET-KE, Items 55 – 80.

Item	P_{NeMUP}	P_{Lemme}	T_{KN}	T_{Lemme}	T_{EN}
55	56,29	66,67	0,126	0,406	0,639
56	22,86	25,64	0,048	0,120	0,644
57	81,63	79,49	0,341	0,676	0,645
58	83,84	84,62	0,210	0,652	0,691
59	47,47	48,72	-0,001	0,261	0,649
60	83,10	82,05	0,169	0,566	0,616
61	80,90	84,61	0,177	0,494	0,646
62	65,84	71,79	0,233	0,517	0,647
63	23,97	10,26	0,070	0,303	0,701
64	93,02	82,05	0,274	0,137	0,677
65	44,17	53,85	0,120	0,346	0,624
66	61,80	41,03	0,152	0,195	0,650
67	60,70	53,85	0,157	0,356	0,681
68	41,97	41,03	0,156	0,332	0,591
69	80,90	89,74	0,152	0,649	0,680
70	68,41	53,85	0,044	0,305	0,667
71	97,06	92,31	0,249	0,548	0,646
72	83,10	74,36	0,302	0,170	0,691
73	73,92	79,49	0,288	0,364	0,721
74	92,29	84,62	0,055	0,443	0,637
75	66,21	76,92	0,364	0,558	0,555
76	72,08	76,92	0,107	0,258	0,543
77	60,33	51,28	0,222	0,245	0,616
78	66,57	61,54	0,151	0,150	0,664
79	74,66	74,36	0,195	0,469	0,631
80	93,02	92,31	0,196	0,813	0,592

P_{NeMUP} = Schwierigkeit in der NeMUP-Stichprobe, P_{Lemme} = Schwierigkeit in der Lemme-Stichprobe, T_{KN} = Trennschärfe der kognitiven Empathie in der NeMUP-Stichprobe, T_{Lemme} = Trennschärfe der kognitiven Empathie in der Lemme-Stichprobe, T_{EN} = Trennschärfe der emotionalen Empathie in der NeMUP-Stichprobe.

8. Eidesstattliche Versicherung

„Ich, Karin Diane Müller, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Erstellung und Erprobung zweier Parallelversionen des Multidimensionalen Empathietests (MET) zur longitudinalen Erfassung des Empathievermögens“ selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

01.03.2021

9. Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

10. Danksagung

Mein Dank gilt zuerst den Probanden, die Zeit und Mühe geopfert haben, um an der Erstellung der Parallelversionen mitzuwirken.

Ein großer Dank richtet sich außerdem an meinen Doktorvater Prof. Dr. med. Stephan Köhler für seine ständige Unterstützung, Hilfe und Motivation bei der Erstellung dieser Arbeit.

Ebenfalls danke ich Dr. rer. nat. Anne Guhn für ihre Unterstützung, Hilfe und Motivation, insbesondere auch bei allen statistischen und methodischen Fragestellungen.

Mein großer Dank geht darüber hinaus an Prof. Dr. rer. nat. Isabel Dziobek für die Bereitstellung des Bildmaterials und der Daten des Multidimensionalen Empathietests, ohne die diese Arbeit nicht möglich gewesen wäre, sowie für die Hilfestellung bei der Zusammenstellung der ersten Testversionen.

Ich danke zudem Matthias Kupferschmidt für die Programmierung der Parallelversionen. Ohne seine Bereitschaft, kleine und große Änderungen in kürzester Zeit möglich zu machen, wäre diese Arbeit nicht entstanden.

Camille Bouchez danke ich für das Gegenlesen der Arbeit und die konstruktiven Anmerkungen.

Ein besonderer Dank richtet sich an meinen Freund Alexander Bremer für die ständige Motivation und liebevolle Unterstützung.

Zuletzt möchte ich meinen Eltern und meinem Bruder Andreas danken, die mich während meines gesamten Studiums und der Erstellung dieser Arbeit immerzu und in jeder erdenklichen Weise unterstützt haben.