

Aus der Rheinischen Landes- und Hochschulklinik
- Psychiatrische Klinik der Heinrich-Heine-Universität -
Düsseldorf
Direktor: Prof. Dr. W. Gaebel

Wirkungen experimenteller Simulation auf das MMPI

Dissertation
zur Erlangung des Grades eines Doktors der Medizin
Der Medizinischen Fakultät der Heinrich-Heine-Universität
Düsseldorf

vorgelegt von
Luise Heitmann - von Franqué

2006

Als Inauguraldissertation gedruckt mit Genehmigung
der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

gez.: Univ.-Prof. Dr. med. Dr. rer. nat. Bernd Nürnberg
Dekan

Referent: Prof. Dr. Erlo Lehmann
Korreferent: Prof. Dr. Thomas Daldrup

Inhaltsverzeichnis

1.	Einleitung	1
1.1.	Begriffsbestimmung Simulation	1
1.2.	Die Bedeutung von Simulation auf medizinischem, sozialem und rechtlichem Gebiet	2
1.3.	Simulationsdiagnostik in Deutschland	3
1.4	Das Minnesota Multiphasic Personality Inventory (MMPI) als Instrument zur Simulationsdiagnostik	5
1.5	Validitätsskalen und klinische Skalen des MMPI	5
1.6	Spezielle Skalen des MMPI zur Simulationsdiagnostik	7
1.7	Studien zum Simulationsdiagnostik mit dem MMPI	8
1.8	Fragestellung	10
2.	Methode	11
2.1	Probanden	11
2.2	Versuchsplan	11
2.3	Messverfahren	11
2.3.1	MWT-B	11
2.3.2	MMPI–Die Deutsche Kurzfassung für Handauswertung	11
2.4.	Versuchsdurchführung	12
2.4.1	Versuchsablauf	12
2.4.2	Gültige Selbstbeschreibung	12
2.4.3	Absichtliche Verfälschung	12
2.5.	Auswertung	13
3.	Ergebnisse	14
3.1	Probandenstichprobe	14
3.2	Die Validitätsskalen L, F, K sowie die F-K-Differenz und ihre Beeinflussung durch Simulation	14
3.2.1	Die L-Skala	14
3.2.2.	Die F-Skala	15
3.2.3	Die K-Skala	17
3.2.4	Profilvergleich L, F, K	19
3.3	Die F-K-Differenz und ihre Beeinflussung durch Simulation	20
3.4	Geschlechterspezifische Beeinflussung der Validitätsskalen L, F, K und der F-K-Differenz	21
3.5	Die klinischen Skalen und ihre Beeinflussung durch Simulation	27

4.	Diskussion	29
4.1	Zur Methode	29
4.2	Zum Einfluss von Simulation auf die Validitätsskalen L, F, K und die F-K-Differenz	31
4.2.1	L-Skala	31
4.2.2	F-Skala	31
4.2.3	K-Skala	33
4.2.4	Profilvergleich L, F, K	33
4.2.5	F-K-Differenz	34
4.2.6	Cut-off-Werte F-K	35
4.3.	Zu den geschlechterspezifischen Unterschieden bei Simulation bezogen auf die Validitätsskalen L, F, K und die F-K-Differenz	36
4.4.	Zum Einfluss von Simulation auf die klinischen Skalen	38
5.	Zusammenfassung	41
6.	Literatur	44

1. Einleitung

1.1 Begriffsbestimmung Simulation

Der Begriff Simulation leitet sich etymologisch vom lateinischen *simulare* „ähnlich machen, nachahmen, zum Schein äußern oder vorgeben“ ab. In diesem Sinne ist Simulation weder ein spezifisches psychiatrisches Krankheitsbild noch lässt sie sich als Symptom einer bestimmten psychischen Störung zuordnen (53).

Der entsprechende englischsprachige Terminus „*malingering*“ ist im DSM-IV der American Psychiatric Association (APA) definiert als „the intentional production of false or greatly exaggerated physical or psychological symptoms motivated by external incentives“. Analog hierzu formuliert Merten, dass „der Begriff der Simulation von körperlichen und psychischen Störungen meint, dass gesteuert und absichtlich durch Symptomschilderung oder Produktion eine externe Zielerreichung (sekundärer Krankheitsgewinn) angestrebt wird“ (43). Dabei finden sich fließende Übergänge von der reinen, zielgerichteten Simulation zu Aggravationen, oder dem, was häufig als „psychisch überlagerte“ Störung bezeichnet wird. Wechselbeziehungen mit tatsächlichen körperlichen oder psychischen Symptomen sind im Einzelfall oft nur schwer zu erkennen (68). Im ICD-10 ist der Begriff Simulation nicht codiert.

Im Unterschied zur Simulation werden artifizielle und somatoforme Störungen als Krankheitsbilder im ICD-10 dargestellt. Die Abgrenzung von der Simulation erfolgt einerseits über Art und Bewusstheit der zugrundeliegenden Motivation, andererseits über die Bewusstheit der Beschwerdeschilderung bzw. –produktion. Von einer artifiziellen Störung (ICD 10: F68.1) wird dann gesprochen, wenn zwar wie bei der Simulation die Beschwerdeschilderung oder –erzeugung vorsätzlich erfolgt, das Motiv dafür jedoch nicht unbedingt in äußeren Vergünstigungen liegt, sondern im Kranksein selbst. Bei diesem Störungsbild hat das Erzeugen selbstmanipulierter Krankheiten am ehesten zwanghaften Charakter, und die Patienten fühlen sich ihren unterbewussten Zielen unfreiwillig unterworfen (12).

Auch bei der somatoformen Störung (ICD 10: F45) ist für den Patienten der sekundäre Krankheitsgewinn nicht vorrangig, selbst wenn er dem Untersucher nicht selten offensichtlich erscheint (46). Zusätzlich wird von der somatoformen Störung gefordert, dass die Symptomproduktion unbewusst erfolgen muss. Besonders bei den Konversionssyndromen scheint die Erzeugung von Beschwerden jedoch häufig zumindest bewusstseinsnah zu erfolgen „Konversionssymptome sind nicht einfach vorgetäuscht, auch wenn sie nicht immer ganz unbewusst ablaufen“ (67) oder, aus dem entgegengesetzten Betrachtungswinkel, „man wird also dann und wann nicht unterscheiden können, wie viel von der Simulation bewusster Betrug ist und wie viel aufgrund der Krankheit aus dem Unterbewussten heraus gespielt wird“ (6). Um diese fließenden Übergänge und Mischbilder zu berücksichtigen, schlägt Merten vor, die

o.g. Störungen sowie die Simulation vor dem Hintergrund dreier sich überlappender Kontinua zwischen „bewusst“ und „unbewusst“, zwischen offensichtlich sekundärem und ausschließlich primärem Krankheitsgewinn und zwischen „Normalität“ und „Persönlichkeitsstörung“ zu sehen (46).

1.2 Die Bedeutung von Simulation auf medizinischem, sozialem und rechtlichem Gebiet

Per definitionem liegt Simulation das bewusste Streben nach äußeren Vergünstigungen zugrunde, d.h., Simulationsverdacht wird beim Untersucher nur in den Fällen entstehen, in denen ein sekundärer Krankheitsgewinn deutlich wird. Ob das im Verlauf einer Exploration geschieht, hängt sowohl von den Fähigkeiten des Untersuchers wie auch von denen des Simulanten ab. Über die Häufigkeit von Simulation lässt sich daher nur schwer eine Aussage treffen. Ein großer Anteil der Simulationen oder groben Aggravationen bleibt allein schon deshalb im Dunkeln, weil sie undramatisch und nur für kurze Zeit arrangiert werden. Sie führen in der Regel nicht zu einem Klinikaufenthalt, wohl aber zu einem kurzzeitigen Fehlen am Arbeitsplatz (45). In Kliniken oder bei Begutachtungen scheitern empirische Studien über Simulationen naturgemäß daran, dass „erfolgreiche“, d.h. nicht erkannte Simulationen nicht zu erfassen sind (69).

Fest steht, dass Simulation und Aggravation bei all jenen gutachterlichen Fragen Probleme aufwerfen können, bei denen den Eigenangaben des Probanden eine erhebliche Bedeutung zukommt. Dies trifft vor allem für psychiatrische und neurologische Fragestellungen zu, bei denen apparative Untersuchungen mit „harten“ Daten eine untergeordnete Rolle spielen (40, 60, 69).

So wurde die Simulation von Epilepsien, Krämpfen, Geistesstörungen und Lähmungen bereits Ende des 19. Jahrhunderts im deutschen Sprachraum ausführlich beschrieben, zum Beispiel: „Die simulierten Krankheiten der Wehrpflichtigen“ 1880 oder „Die Simulation von Geistesstörungen und Epilepsie“ 1907 (42). Das erste englischsprachige Werk zu diesem Thema erschien bereits 1843 (20). Wurde Simulation in den frühen Schriften zunächst nur im Bereich des Militärdienstes beschrieben, führte die Erweiterung des Versicherungswesens im Laufe des 19. Jahrhunderts und die Einführung der Schuldfähigkeit im deutschen Strafgesetzbuch 1871 dazu, dass auch in diesen Zusammenhängen zunehmend Simulation beobachtet wurden (42). Auch heute noch stellt sich die Frage nach Simulation überwiegend in diesen Bereichen: Einberufung zum Militär, Einsatz in Kampfgebieten, Beurteilung der Schuld-, Haft- oder Verhandlungsfähigkeit, Berentung aufgrund von Arbeits- oder Berufsunfähigkeit, Entschädigungszahlungen nach Unfällen. In neuerer

Zeit hinzugekommen sind ausländerrechtliche Fragen, wie die nach Reisefähigkeit bei der Ausweisung von abgelehnten Asylbewerbern.

Das letztgenannte Beispiel lässt bereits erkennen, dass Simulation in bestimmten Kontexten durchaus adaptiv, adäquat und sogar moralisch hochwertig, je nach ethischem und historischem Stand des Urteilenden, sein kann (45). Einzelfälle, wie der eines 14-jährigen Mädchens, das eine Psychose simulierte, um durch die Klinikeinweisung dem sexuellen Missbrauch durch ihren Stiefvater zu entgehen, verdeutlichen diesen Gesichtspunkt (52).

Die simulierten Symptome variieren abhängig von den Erfahrungen des Individuums (das o.g. Mädchen hatte beispielsweise zuvor die psychotische Episode eines Cousins miterlebt), sie sind aber auch abhängig von dem Kontext, in dem sie vorgebracht werden. So sind nach leichten Schädel-Hirn-Traumata, oder nach Lösungsmittelexposition vor allem kognitive Störungen abzuklären. Gutachten zur Reisefähigkeit erfordern hingegen häufig den Ausschluss bzw. die Bestätigung einer posttraumatischen Belastungsstörung.

Die Simulation von Psychosen wird vor allem bei Begutachtungen im forensisch-psychiatrischen Bereich beobachtet. Es wird angenommen, dass hier der Anteil „reiner“ Simulanten höher ist als z.B. bei Begutachtungen nach Unfällen am Arbeitsplatz (45). Angeklagte simulieren Psychosen, um für schuld-, haft- oder verhandlungsunfähig gehalten zu werden. Bereits Verurteilte simulieren, um aus der Haftanstalt in ein Krankenhaus verlegt zu werden oder (vor allem bei Drogenabhängigen) um sedierende Medikamente zu erhalten. Gelegentlich ist auch bereits die Aussicht auf ein Dach über dem Kopf – und sei es das in einer Psychiatrischen Klinik – Krankheitsgewinn genug für das Simulieren einer Psychose (52).

1.3 Simulationsdiagnostik in Deutschland

In aktuellen Lehrbüchern der klinischen Psychiatrie findet Simulation, und damit verbunden Simulationsdiagnostik, höchstens am Rande Berücksichtigung (4, 19, 36, 48, 62). Wie oben beschrieben, spielt Simulation im klassischen klinischen Kontext auch eine deutlich geringere Rolle als bei gutachterlichen Fragestellungen. Verwunderlich erscheint, dass auch in den deutschen Standardwerken zur Begutachtung die Simulationsdiagnostik einen vergleichsweise geringen Stellenwert einnimmt.

Venzlaff und Förster führen verschiedene klinische Auffälligkeiten an, wie appellatives Verhalten, Waghheit der Symptomschilderung oder Diskrepanz zwischen Beschwerdeschilderung und Verhalten, die als Hinweise auf Simulation gedeutet

werden können. Zur weiteren Abklärung wird geraten, dass diesen Hinweisen „detailliert nachgegangen werden sollte“ (69). Rasch empfiehlt dem Gutachter, Simulationsverdacht in jedem Fall im psychischen Befund zu erwähnen und ggf. durch Verhaltensbeobachtung im Rahmen einer stationären Begutachtung zu verifizieren bzw. auszuschließen (51).

Die Simulationsdiagnostik durch stationäre Überwachung ist bereits bei Bleuler beschrieben, „wichtig ist eine sehr genaue, wohlmöglich durch Tag und Nacht ununterbrochene Beobachtung, die einesteils den Simulanten ermüdet und andernteils gewöhnlich bald eine Menge von Beweisen für eine versuchte Täuschung bringt“ (6). Verschiedene Studien haben jedoch gezeigt, dass selbst Experten Schwierigkeiten haben, Simulation anhand von ausschließlich klinischer Beobachtung zu stellen (13, 17, 57).

Rasch diskutiert darüber hinaus die Anwendung von Tests bei der psychiatrischen Begutachtung im Strafverfahren. Ausdrücklich weist er auf die „Gefahr verengter Optik“ und die Antizipation der Resultate bei hypothesengeleiteten Testverfahren hin, kommt aber letztlich zu dem Schluss, dass Bedenken vor allem gegenüber solchen psychischen Befunden bestehen sollten, die sich lediglich auf individuell gesammelte Erfahrungen stützen, und die dem Risiko vorurteilsbedingter Verzerrung unterliegen (51).

Die Forschung zur neuropsychologischen Testdiagnostik im Zusammenhang mit Simulationsverdacht bei mnestischen und kognitiven Störungen hat in den 90er Jahren in Deutschland einen gewaltigen Aufschwung erfahren. Neue, spezifische Tests zur Diagnostik suboptimalen Leistungsverhaltens wurden entwickelt, von bereits existierenden Verfahren wurden Adaptationen vorgelegt und von Heubrock und Petersen in einer Testbatterie zur Forensischen Neuropsychologie vereinigt. Sie finden vor allem im Zivil- und Sozialrecht, zum Beispiel bei der Begutachtung zur Berentung Verwendung (34).

Bei der Simulation von Psychosen, die überwiegend im Bereich des Strafrechts beobachtet wird, stellt international die Fragebogendiagnostik eine wichtige Ergänzung zur klinischen Untersuchung dar. Im englischsprachigen Raum stützt sich die Forschung in diesem Bereich nach wie vor auf das Minnesota Multiphasic Personality Inventory (MMPI), dessen deutsche Fassung seit 1962 in Gebrauch ist. Vor allem der mit 566 Fragen häufig als alpträumhaft empfundene Umfang führte dazu, dass das MMPI in Deutschland gegenüber dem Freiburger Persönlichkeitsinventar (FPI) bei der Diagnostik von Persönlichkeitsstörungen zunehmend an Bedeutung verlor (1, 61). Das FPI verfügt zwar über eine so genannte Offenheitsskala, deren Items in etwa der L-Skala des MMPI entsprechen, im Bereich der Simulationsdiagnostik erwies sich der FPI jedoch als ungeeignetes Verfahren (45).

1.4. Das Minnesota Multiphasic Personality Inventory (MMPI) als Instrument zur Simulationsdiagnostik

Als Prototyp der heutigen Persönlichkeitsfragebogentests gilt das 1917 von einem Komitee der APA ausgearbeitete Verfahren zur Untersuchung der „emotional fitness“ bei Soldaten des US-Expeditionscorps. Es war als „screening device“ gedacht; auffällige Probanden sollten anschließend einer eingehenden psychiatrischen Untersuchung zugeführt werden (47). Bereits drei Jahre später wurde die Anfälligkeit des Instruments für Verfälschungen deutlich, als unmittelbar nach dem Waffenstillstand ein erheblicher Abfall der Scores bei den Soldaten zu beobachten war (35).

Die Bemühungen, Verfälschungstendenzen beim Bearbeiten von Persönlichkeitsfragebögen zu verhindern oder zumindest einzuschränken, waren zunächst allgemeiner Art. So wurden die Probanden ausdrücklich gebeten, offen und ehrlich zu antworten, und es wurde ihnen Anonymität zugesichert. Zudem wurde versucht, Fragen zu verwenden, deren diagnostischer Wert nicht direkt erkennbar ist, doch dies erwies sich als ausgesprochen schwierig. Bei den meisten Fragen, die zur Differenzierung zwischen „Normalität“ und psychischer Auffälligkeit dienen sollten, war die korrekte Antwort offensichtlich (44).

Bei der Entwicklung weiterer Persönlichkeitsfragebögen wurde mit verschiedenen Methoden versucht, der Gefahr einer Verfälschung direkt bei der Testkonstruktion Rechnung zu tragen.

Mit dem weitaus größten Umfang an Entwicklungs- und Forschungsarbeit, die bis dahin zu einem Persönlichkeitsfragebogen geleistet wurde, wurde Ende der 30er Jahre des 20. Jahrhunderts das Minnesota Multiphasic Personality Inventory (MMPI) von Hathaway und McKinley entwickelt.

Besondere Aufmerksamkeit wurde bei der Entwicklung des MMPI der Messung auffälliger Beantwortungstendenzen gezollt. Hierzu wurden dem Test vier sog. Validitätsskalen angegliedert. Methodisch berief man sich dabei auf Ruch, der als erster solche Fragen, die „susceptible for malingering“ schienen, in einer eigenen Skala zusammenführte (59).

1.5 Validitätsskalen und klinische Skalen des MMPI

„Cannot say“- oder „?“-Skala: Den Wert auf der „?“-Skala erhält man durch Auszählen der Fragen, deren Beantwortung ausgelassen wurde. Seine Höhe kann die Signifikanz der übrigen Skalen beeinflussen. Liegt die Anzahl nicht beantworteter Fragen an der oberen Normgrenze (T-Wert = 70), darf angenommen werden, dass

bei vollständig ausgefülltem Bogen die Werte für die einzelnen Skalen stärker vom Durchschnitt abweichen würden, als sich im Profil zeigt. Ist die Anzahl der ausgelassenen Fragen sehr hoch, so sind alle anderen Skalen ungültig. Die „?“-Skala verlor im Verlauf der Weiterentwicklung des MMPI zunehmend an Bedeutung und wurde bei der Weiterentwicklung des MMPI zum MMPI-2 nicht übernommen.

F-Skala: Die F-Skala wurde primär entwickelt, um Probanden zu identifizieren, die nicht lesen können oder aus sonstigen Gründen die Fragen nicht verstehen, oder solche, die sich zwar vordergründig kooperativ zeigen, aber die Fragen nachlässig oder ohne jeden persönlichen Bezug beantworten. Hierzu wurden aus dem ursprünglichen Fragenpool solche Fragen ausgewählt, die von nahezu allen Probanden in eine Richtung beantwortet wurden, beispielsweise: „Ich kann alle Menschen gleich gut leiden“ (Standardantwort „Nein“). 64 dieser Fragen wurden zur F-Skala zusammengestellt. Bei der Auswertung werden diejenigen Antworten erfasst, die von der Üblichen abweichen. Die Vermutung, dass die F-Skala auch Probanden erfasst, die sich bewusst als psychisch auffällig darstellen („fake bad“), konnte bald darauf experimentell bestätigt werden (10, 23). Auch bei Menschen mit Persönlichkeitsstörungen und psychiatrischen Patienten ist der F-Wert in der Regel erhöht (z.B. 2, 3, 26, 30, 31).

L-Skala: Die 15 Fragen der L-Skala beziehen sich sämtlich auf Verhaltensweisen, die als sozial erwünscht gelten, aber tatsächlich nur selten anzutreffen sind. So spiegelt beispielsweise die Verneinung der Aussage „Ich werde manchmal wütend“ ein durchaus erstrebenswertes Verhalten wieder, das aber dennoch nur für einen kleinen Prozentsatz der Bevölkerung tatsächlich zutreffen wird. Anspruch der L-Skala war es, solche Probanden zu erfassen, die bemüht sind, ein möglichst positives Bild von sich darzustellen („fake good“). In Studien zeigte sich jedoch, dass differenziertere Probanden den Fallen der L-Skala häufig ausweichen konnten. Ein sehr hoher Wert auf der L-Skala spricht also relativ sicher für eine übertrieben positive Selbstdarstellung, bei einem unauffälligen Wert auf dieser Skala kann aber eine manipulative Beantwortung des Gesamttests in Richtung „fake good“ nicht sicher ausgeschlossen werden (44). Leicht erhöhte Werte auf der L-Skala wurden zudem auch in einigen Studien bei „fake-bad“ gefunden (2, 9, 16, 23).

K-Skala: Als letzte der Validitätsskalen wurde die K-Skala eingeführt, als man sich zunehmend eingestand, dass eine manipulative Testbearbeitung zumindest in bestimmten Bereichen, wie beim Militär, häufiger war als zunächst angenommen und dass die F- und die L-Skala beschönigende Tendenzen („fake good“) oder übertriebenes Pathologisieren („fake bad“) nicht mit ausreichender Sicherheit erkennen bzw. ausschließen konnten (44). Die K-Skala wurde entwickelt, indem man die Antworten von Probanden analysierte, die trotz klinisch diagnostizierter psychischer Störungen unauffällige Profile im MMPI zeigten. Davon ausgehend, dass

diese Probanden durch ein besonders defensives Antwortverhalten bei bestimmten Fragen die Auffälligkeiten bei anderen Fragen ausgleichen, verglich man ihre Antworten einzeln mit denen einer „Normalgruppe“. Die Fragen mit den häufigsten Abweichungen wurden in der K-Skala versammelt. So finden sich erhöhte Werte auf der K-Skala bei Menschen mit psychischen Störungen, aber unauffälligem Testergebnis, während niedrige Werte bei psychisch Gesunden gefunden werden, die pathologische Tendenzen in übertriebener Weise herausstellen. Die K-Skala wurde daraufhin als ein Korrekturinstrument für die klinischen Skalen eingesetzt, indem bestimmte Skalen in Abhängigkeit von der Höhe des individuellen K-Wertes nach oben korrigiert werden. Von besonderer Bedeutung ist die K-Korrektur in den Fällen, in denen die Werte der klinischen Skalen im Grenzbereich zwischen normal und pathologisch liegen. Hier können durch die K-Korrektur dissimulierte psychische Störungen offensichtlich werden (11, 23, 43, 44).

Für die eigentliche Persönlichkeitsdiagnostik verfügt der MMPI über zehn klinische Skalen. Diese lauten – angelehnt an die Kraepelinsche psychiatrische Klassifizierung: Hypochondriasis (Hs), Depression (D), Hysteria (Hy), Psychopathic Deviate (Pd), Paranoia (Pa), Psychasthenia (Pt), Schizophrenia (Sc), Hypomania (Ma). Weitere Kategorien sind Maskulinität – Femininität (MF) und, 1946 in die revidierte Fassung aufgenommen, soziale Introversion – Extroversion (Si).

1.6 Spezielle Skalen des MMPI zur Simulationsdiagnostik

Schon bald nach der Entwicklung des MMPI wurden über die ursprünglichen Validitätsskalen hinaus weitere Skalen und Indices zur Simulationsdiagnostik entwickelt.

Bereits 1947 wurde der F-K-Index durch Gough eingeführt (24). Während der F-Wert die Tendenz abbildet, psychische Probleme übersteigert darzubieten, zeigt die K-Skala die Bemühung an, Störungen zu negieren. Findet sich kein Hinweis auf eine der beiden Bestrebungen, liegt der F-K-Index um 0. Bei Simulation erhöht sich der F-Wert gegenüber K, so dass ein F-K-Index mit positivem Vorzeichen resultiert. Ein F-K-Index mit negativem Vorzeichen hingegen, kann als Hinweis auf Dissimulation gelten.

1948 veröffentlichte Wiener eine erste Studie zu den Subtle-Obvious-Scales, die seither weite Verbreitung fanden (74). Sie basieren auf verschiedenen klinischen Skalen und unterscheiden zwischen Items, die leicht erkennbar auf eine psychische Störung hinweisen und solche, deren Zusammenhang mit psychischen Störungen eher schwer erkennbar ist. Im direkten Vergleich zeigten sich jedoch die F-Skala und die F-K-Differenz bei der Simulationsdiagnostik den Subtle-Obvious-Scales überlegen (5, 66).

1954 entwickelte Gough zur Diagnostik simulierter Neurosen die Dissimulation Scale (Ds) (25). Die ursprüngliche Skala wurde in den folgenden Jahrzehnten mehrfach überarbeitet, so dass derzeit mehrere Unterformen in Gebrauch sind. Da nicht alle Untersuchungen mit derselben Unterform arbeiten, lässt sich eine übergreifende Aussage zur Effizienz der Ds-Skala nur schwer treffen (5, 28). Ihre Eignung zur Simulationsdiagnostik zeigte sich in neuerer Zeit vor allem in Studien zur Simulation bestimmter Krankheitsbilder (55, 73).

Mit dem MMPI-2 wurden zwei modifizierte Skalen zur Erfassung inkonsistenten Antwortverhaltens eingeführt: die Variable-Response-Inconsistency-Scale (VRIN) und die True-Response-Inconsistency-Scale (TRIN). Beide Skalen erwiesen sich als sensibel in Bezug auf zufälliges Antwortverhalten, sie sind jedoch zur Simulationsdiagnostik weniger geeignet (18, 62, 64).

1.7 Studien zur Simulationsdiagnostik mit dem MMPI

Das MMPI ist weltweit der meist verbreitete Persönlichkeitsfragebogen, auch wenn er aufgrund seines Umfangs und seiner Skalenüberlappungen selbst im Ursprungsland nicht unumstritten ist (8, 33). Er wurde in 30 Sprachen übersetzt und gehört zu den am meisten beforschten Persönlichkeitsfragebögen (15). Auch seine Eignung im Hinblick auf die Simulationsdiagnostik wurde seit seiner Entwicklung in zahlreichen Studien untersucht. Nach der Einführung des MMPI-2, 1989, wurden die Studien fortgeführt und zum Teil wiederholt, um zu überprüfen in wie weit sich die bisherigen Erfahrungen auf das modifizierte Verfahren übertragen lassen. Einen Überblick über die wichtigsten Grundlagenstudien geben zwei Metaanalysen zur Simulationsdiagnostik mit dem MMPI (5) und MMPI-2 (56) sowie eine Arbeit von Schretlen in der auch auf andere Testverfahren eingegangen wird (63).

Ziel der beiden o.g. Metaanalysen war es, die Eignung verschiedener MMPI-Skalen zur Simulationsdiagnostik zu überprüfen. In beiden Untersuchungen erwies sich die F-Skala als der geeignetste Parameter zum Simulationsnachweis, unabhängig davon, ob die Rohwerte oder die T-Werte dieser Skala betrachtet wurden. Auch die Eignung des F-K-Index wurde in beiden Metaanalysen bestätigt. Berry et al. führten daneben noch die Ds-Skala an, Rogers et al. die Obvious-Subtle-Scales.

Explizit schlüsseln die Metaanalysen die verglichenen Studien nach Untersuchungspopulation und Versuchsplan auf und geben so einen Überblick über die gängigen Untersuchungsmethoden zum Thema Simulationsforschung. Am häufigsten wurde das Simulationsverhalten von Studenten untersucht. Von den 28 Studien der Metaanalyse von Berry wurden 14 an Studenten durchgeführt. Von den 15 Studien der Metaanalyse von Rogers arbeiteten 10 mit Studenten. Zum Vergleich wurden entweder die authentischen Angaben derselben oder einer anderen

Studentengruppe hinzugezogen (z.B. 7, 10, 16, 49, 72) oder auch die Fragebögen psychiatrischer Patienten ausgewertet (z.B. 25, 26, 29, 55, 66, 73). Auch in verschiedenen Untersuchungen, die nicht in den Metaanalysen erfasst sind, setzen sich die Untersuchungspopulationen aus Studenten, ggf. in Kombination mit psychiatrischen Patienten zusammen (z.B. 11, 39, 64).

Alternativ zu Studenten wurden in frühen Studien auch Angehörige des Militärs zu Untersuchungen herangezogen (22, 24, 37). Besonders realitätsnahe Ergebnisse liefern die Untersuchungen an Häftlingen (31, 38, 50, 70, 71). Eine frühe Studie beschreibt das Simulationsverhalten von Experten (Psychiatern, Psychologen, Sozialpädagogen) (23).

Auf ein mögliches geschlechtsspezifisches Simulationsverhalten wird lediglich in einer der in den Metaanalysen aufgeführten Studien eingegangen (26).

In den genannten Metaanalysen erst ansatzweise erfasst ist der Trend, zu einer spezielleren Fragestellung. Während die dort angeführten Studien als Aufgabe im allgemeinen die Simulation einer unspezifischen psychiatrischen Erkrankung oder einer „gravierenden“ psychiatrischen Erkrankung vorgeben, zielen neuere Studien zunehmend auch auf die Simulation bestimmter Krankheitsbilder wie Schizophrenie, Depression oder posttraumatische Belastungsstörung und berücksichtigen dabei die Möglichkeit einer gezielten Vorbereitung auf die Simulation (14, 55, 65, 73).

Auffallend ist, dass im deutschen Sprachraum bisher keine Studien zur Simulationsdiagnostik mit dem MMPI vorliegen. Lediglich das „Manual für die Deutsche Kurzform“ (21) brachte den deutschen Anwender des MMPI mit einer Auswahl formaler Interpretationsregeln in Kontakt, die sich auch auf die Simulationsdiagnostik beziehen. Verlangt wird zunächst eine F-K-Differenz von mindestens 10 T-Wertpunkten. Ist diese Bedingung gegeben wird zusätzlich ein F-Wert von über 60 gefordert. Zudem müssen vor der Diagnose Simulation noch Verständnisschwierigkeiten, zufälliges Antworten sowie psychische Erkrankungen ausgeschlossen werden. Leider wurden für diese Interpretationsregeln keine Belege angegeben (15).

1.8 Fragestellung

Im Rahmen einer experimentellen Studie mit Studenten der Universität zu Düsseldorf sollte untersucht werden, wie sich eine auf Simulation gerichtete absichtliche Verfälschungstendenz auf das MMPI-Ergebnis auswirkt. Dabei konzentrierte sich die Arbeit auf die Validitätsskalen L, F, K mit der Frage, ob die Werte auf diesen Skalen Schlüsse auf Simulation zulassen und ob die Berechnung der F-K-Differenz zur Unterscheidung von Simulation versus gültiger Selbstbeschreibung beitragen kann.

Ergänzend sollte die Frage behandelt werden, inwieweit sich Männer und Frauen in der Ausprägung der Veränderungen auf den Validitätsskalen unterscheiden. Um dem gesamten MMPI Rechnung zu tragen galt es weiterhin, die klinischen Skalen nach Auffälligkeiten bei Simulation zu untersuchen.

2. Methode

2.1 Probanden

Die Studie wurde durchgeführt mit 70 Teilnehmern des Kurses „Medizinische Psychologie“ für Medizinstudenten im dritten vorklinischen Semester. Die Studentinnen und Studenten hatten zuvor weder im Studium noch privat Erfahrungen mit dem MMPI oder anderen Persönlichkeitsfragebögen gemacht. Da zum Bearbeiten des MMPI eine gute Beherrschung der deutschen Sprache erforderlich ist, wurden diejenigen Studenten für die Studie als geeignet betrachtet, die im Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B) mindestens 90 IQ-Punkte erreichten.

2.2 Versuchsplan

Es wurde ein abhängiger Zwei-Gruppenplan realisiert. Alle Probanden bearbeiteten die Deutsche Kurzform des MMPI zweifach, zum einen im Sinne einer gültigen Selbstbeschreibung, zum anderen mit Verfälschungsabsicht.

2.3 Messverfahren

2.3.1 MWT-B

Der Mehrfachwahl-Wortschatz-Intelligenztest MWT-B (Lehrl 1995) dient zur Abschätzung der kristallinen Intelligenz bei deutschsprachigen Probanden. Den Testpersonen wird die Aufgabe gestellt, bekannte Wörter wieder zu erkennen und von unbekanntem Wörtern zu unterscheiden. Hierzu steht in 37 Wortzeilen je ein Wort aus der deutschen Sprache unter vier ähnlichen, aber sinnlosen Wörtern. Jede richtig markierte Zeile wird mit einem Punkt bewertet; nach dem resultierendem Rohwert kann der IQ ermittelt werden. Die Durchführungsdauer beträgt etwa fünf Minuten

2.3.2 MMPI – Die Deutsche Kurzform für Handauswertung

Verwendet wurde für unsere Untersuchung die Deutsche Kurzform des MMPI für Handauswertung (in der 2. Auflage 1993). Gegenüber der Langform des MMPI, die 566 Items enthält, ist die Deutsche Kurzform zur Handauswertung auf 221 Items verkürzt. Die Kürzung wirkt sich vor allem auf die Bearbeitungsdauer aus, die mit 30 bis 40 Minuten deutlich unter der ursprünglichen Form liegt. Die Auswertungszeit beträgt fünf bis zehn Minuten.

2.4 Versuchsdurchführung

2.4.1 Versuchsablauf

Die Untersuchungsteilnehmer wurden an fünf aufeinanderfolgenden Tagen in Gruppen von je 13 bis 15 Probanden untersucht. Zunächst erfolgte mittels MWT-B die Bestimmung des IQ und des deutschen Sprachverständnisses. Anschließend wurde jedem Probanden ein Testheft zur Deutschen Kurzform des MMPI vorgelegt, mit der Aufforderung, den Testbogen zweimal auszufüllen: einmal als gültige Selbstbeschreibung, das andere mal mit absichtlicher Verfälschung. Zudem sollten auf jedem Testbogen die Basisdaten Alter, Geschlecht sowie der ermittelte IQ-Wert vermerkt werden, jedoch ausdrücklich nicht der Name. Ein fester Zeitrahmen zum Ausfüllen der Testbögen wurde nicht festgesetzt.

2.4.2 Gültige Selbstbeschreibung

Bedingung für die gültige Selbstbeschreibung war das Bearbeiten des Fragebogens gemäß der folgenden Anleitung im Testheft:

„In diesem Heft finden Sie eine Reihe von Feststellungen, die sich auf Ihr Befinden, auf Ihre Meinungen, Interessen, Erfahrungen, Vorlieben und Abneigungen beziehen. Lesen Sie diese Feststellungen und entscheiden Sie möglichst spontan, ob eine Feststellung jeweils eher auf Sie zutrifft oder nicht. Dabei gibt es keine „richtigen“ oder „falschen“ Antworten, weil jedermann seine eigenen Ansichten haben kann. Für jede Feststellung sind zwei Antworten möglich: „Ja“, wenn es auf Sie zutrifft, „Nein“, wenn es nicht auf Sie zutrifft. Beantworten Sie bitte jede Feststellung nach Ihrer eigenen Meinung. In Fällen, wo Sie nicht absolut mit „Ja“ oder mit „Nein“ Stellung nehmen können, entscheiden Sie sich trotzdem für das eine oder andere; entscheiden Sie sich für diejenige Aussage, die eher zutrifft, als die andere. Nur wenn eine Feststellung sich auf etwas bezieht, was Sie nicht wissen, verzichten Sie auf die Stellungnahme. Unmittelbar hinter jeder Fragestellung finden Sie zwei Kästchen: Stimmen Sie einer Feststellung zu, setzen Sie ein Kreuz bei „Ja“; verneinen Sie einer Feststellung, kreuzen Sie „Nein“ an.“

2.4.3 Absichtliche Verfälschung

Für die absichtliche Verfälschung erhielten die Untersuchungsteilnehmer folgende Anweisungen: „Stellen Sie sich bitte vor, Sie hätten aus persönlichen Motiven heraus einen Menschen getötet. Sie wollten dafür aber keine Verantwortung übernehmen. Sie wollten vielmehr vortäuschen, dass sie wegen einer „krankhaften seelischen Störung“, wegen einer „tiefgreifenden Bewusstseinsstörung“ oder wegen „Schwachsinn“ oder einer „schweren anderen seelischen Abartigkeit“ unfähig waren,

das Unrecht der Tat einzusehen oder nach dieser Einsicht zu handeln. Sie wollen den Schutz der Psychiatrie in Anspruch nehmen.“

2.5. Auswertung

Die Auswertung der Bögen erfolgte zunächst von Hand durch Übertragung der Rohwerte in die Profilbögen (differenziert nach m/w), wodurch die entsprechenden T-Werte ermittelt werden konnten. Anschließend wurden die Basisdaten sowie die T-Werte der Validitätsskalen und der klinischen Skalen in einer Datenbank erfasst und mittels SPSS ausgewertet. Die statistische Auswertung umfasste das Errechnen von Mittelwerten und Standardabweichungen (SD) sowie die zufallskritische Bewertung der Bedingungs differenzen mittels t-Test für abhängige Stichproben.

Zur Vermeidung multiplen Testens wurden entscheidungsstatistisch nur die Bedingungs differenzen bezüglich der Validitätsskalen L,F,K bewertet. Für die klinischen Skalen erfolgte die Beschreibung lediglich deskriptiv.

3. Ergebnisse

3.1 Probandenstichprobe

Es fand sich eine Kurszusammensetzung von 43 Frauen (61,4%) und 27 Männern (38,6%). Die Gesamtzahl der Kursteilnehmer betrug 70. Das Alter der Männer lag im Durchschnitt bei 23,48 Jahren (SD 1,847), das der Frauen bei 23,65 (SD 6,575). Insgesamt fand sich ein Durchschnittsalter von 23,59 (SD 5,25).

Unter den Studenten waren 48 (68,6%) mit deutscher Muttersprache, 22 Studenten (31,4%) waren mit einer anderen Muttersprache aufgewachsen. Im Mehrfachwahlwortschatztest lag der Durchschnitts-IQ insgesamt bei 117,53 (SD 12,09). Für die Männer lag der Wert bei 117,00 (SD 12,33), für die Frauen bei 117,86 (SD 12,076). Für alle Studienteilnehmer lag der Durchschnitts-IQ im MWT-B über 90.

Alle Kursteilnehmer füllten die Antwortbögen zum MMPI vollständig aus und versahen ihn mit den geforderten Basisdaten Alter, Geschlecht und IQ.

3.2 Die Validitätsskalen L, F, K sowie die F-K-Differenz und ihre Beeinflussung durch Simulation

3.2.1 Die L-Skala

Auf der L-Skala ließ sich unter Simulation ein Absinken des Mittelwertes von 47,24 auf 43,61 T-Wert-Punkte beobachten. Der t-Test bestätigte die Signifikanz der Veränderung.

Tabelle 1: Veränderung der L-Skala durch Simulation

	Mittelwert	SD Mittelwert	Differenz	SD Differenz	t	sig. (2-seitig)
sim L	43,614	8,5				
echt L	47,243	7,53	-3,629*	01,2	-2,54	0,013

* signifikant $\leq 0,05$

Die Häufigkeitenanalyse für die L-Skala zeigt auf, wie viele Probanden jeweils einen bestimmten T-Wert auf dieser Skala erreichten. Hierbei zeigte sich, dass die Veränderungen des Mittelwertes zwischen Simulation und gültiger Selbstbeschreibung auf einer Häufigkeitsverschiebung innerhalb eines fast identischen Skalenbereichs beruht. Der günstigste Cut-off-Wert zur Differenzierung von simulierten Skalen und gültiger Selbstbeschreibung liegt bei 46 mit 28,6% falsch-negativen und 42% falsch-positiven Ergebnissen.

Tabelle 2: Häufigkeitenanalyse L-Skala simuliert und echt

L simuliert			L echt		
T-Wert L-Skala	Häufigkeit	Cumulative Prozent	T-Wert L-Skala	Häufigkeit	Cumulative Prozent
30	5	7,1	34	2	2,9
34	6	15,7	35	2	5,7
37	7	25,7	37	2	8,6
39	9	38,6	39	8	20,0
41	6	47,1	41	7	30,0
43	10	61,4	43	7	40,0
46	7	71,4	46	2	42,9
47	3	75,7	47	11	58,6
50	4	81,4	50	6	67,1
52	2	84,3	52	8	78,6
55	1	85,7	55	4	84,3
56	3	90,0	56	4	90,0
59	1	91,4	59	3	94,3
60	5	98,6	60	2	97,1
64	1	100,0	64	1	98,6
65			65	1	100,0
Gesamt	70		Gesamt	70	

— günstiger Cut-off-Wert

Tabelle 3: Cut-off-Wert L-Skala

Cut-off L		
Cut-off 46	falsch positiv	falsch negativ
simuliert		N20 28,6%
echt	N30 42,9%	

3.2.2. Die F-Skala

Die ausgeprägteste Differenz zwischen Simulation und gültiger Selbstbeschreibung fand sich auf der F-Skala. Mit einem Anstieg von 50,53 auf 98,60 T-Wert-Punkte lag der Mittelwert bei Simulation hoch signifikant über dem Wert bei gültiger Selbstbeschreibung.

Tabelle 4: Veränderung der F-Skala durch Simulation

	Mittelwert	SD Mittelwert	Differenz	SD Differenz	t	sig. (2-seitig)
sim F	98,60	17,10	48,07**	02,0	19,75	0,000
echt F	50,53	8,18				

* * hoch signifikant $\leq 0,01$

Die Häufigkeitenanalyse für die F-Skala zeigt die große Streuung der Werte bei Simulation, die schon durch die hohe Standardabweichung für sim F in Tabelle 4 zum Ausdruck kam.

Im Handbuch zur Deutschen Kurzform des MMPI wird ein F-Wert von >60 als Bedingung für die Verdachtsdiagnose Simulation gefordert. Wie die Häufigkeitenanalyse zur F-Skala zeigt, ist dies auch in unserer Untersuchung der günstigste Cut-off-Wert. Unter Simulation fand sich bei einem Probanden ein Wert <60 (falsch-negativ) bei gültiger Selbstbeschreibung waren die Werte von sechs Probanden auf >60 erhöht (falsch-positiv).

Tabelle 5: Häufigkeitenanalyse F-Skala simuliert und echt

F simuliert			F echt		
T-Wert F-Skala	Häufigkeit	Cumulative Prozent	T-Wert F-Skala	Häufigkeit	Cumulative Prozent
57	1	1,4	43	12	17,1
62	2	4,3	44	4	22,9
69	3	8,6	46	8	34,3
74	1	10,0	47	10	48,6
75	2	12,9	49	2	51,4
78	1	14,3	50	14	71,4
80	2	17,1	52	1	72,9
81	1	18,6	53	3	77,1
82	2	21,4	55	6	85,7
85	3	25,7	56	1	87,1
87	1	27,1	59	2	90,0
88	1	28,6	60	1	91,4
91	2	31,4	63	1	92,9
94	1	32,9	69	1	94,3
96	2	35,7	71	1	95,7
97	3	40,0	72	1	97,1
99	4	45,7	77	1	98,6
100	2	48,6	82	1	100,0
102	3	52,9			
103	2	55,7			
104	3	60,0			
106	2	62,9			
107	1	64,3			
109	2	67,1			
110	2	70,0			
113	3	74,3			
116	10	88,6			
119	8	100,0			
Gesamt	70		Gesamt	70	

— günstigster Cut-off-Wert

Tabelle 6: Cut-off-Wert F-Skala

Cut-off F		
Cut-off 60	falsch positiv	falsch negativ
simuliert		1
echt	6	

3.2.3. Die K-Skala

Mit einer Differenz von $-14,97$ T-Wert-Punkten lag der Mittelwert auf der K-Skala bei Simulation hoch signifikant unter dem Mittelwert bei gültiger Selbstbeschreibung.

Tabelle 7: Veränderung der K-Skala durch Simulation

	Mittelwert	SD Mittelwert	Differenz	SD Differenz	t	sig. (2-seitig)
sim K	40,64	7,83	-14,97**	01,2	-10,78	0,000
echt K	55,60	7,96				

** hoch signifikant $\leq 0,01$

Die Häufigkeitenanalyse für die K-Skala zeigt eine mäßige Überschneidung der Skalenbereiche von Simulation und gültiger Selbstbeschreibung. Der günstigste Cut-off-Wert ließ sich bei 40 festlegen, mit 42,9% falsch-negativen Ergebnissen bei Simulation und 2,9% falsch-positiven Ergebnissen bei der gültigen Selbstbeschreibung.

Tabelle 8: Häufigkeitenanalyse K-Skala simuliert und echt

K simuliert			K echt		
T-Wert K-Skala	Häufigkeit	Cumulative Prozent	T-Wert K-Skala	Häufigkeit	Cumulative Prozent
28	2	2,9	39	1	1,4
30	3	7,1	40	1	2,9
32	3	11,4	42	4	8,6
33	4	17,1	44	4	14,3
34	2	20,0	46	2	17,1
35	10	34,3	48	3	21,4
36	4	40,0	50	3	25,7
37	4	45,7	51	4	31,4
38	1	47,1	52	2	34,3
39	4	52,9	53	3	38,6
40	3	57,1	54	2	41,4
42	3	61,4	55	4	47,1
44	7	71,4	56	2	50,0
46	5	78,6	57	6	58,6

48	6	87,1
51	2	90,0
52	1	91,4
53	1	92,9
54	1	94,3
55	2	97,1
62	2	100,0
Gesamt	70	

— günstigster Cut-off-Wert

59	3	62,9
60	4	68,6
61	3	72,9
62	3	77,1
63	3	81,4
64	4	87,1
65	2	90,0
66	3	94,3
67	1	95,7
68	1	97,1
70	2	100,0
Gesamt	70	

Tabelle 9: Cut-off-Wert K-Skala

Cut-off K		
Cut-off 40	falsch positiv	falsch negativ
simuliert		N30 42,9%
echt	N2 2,9%	

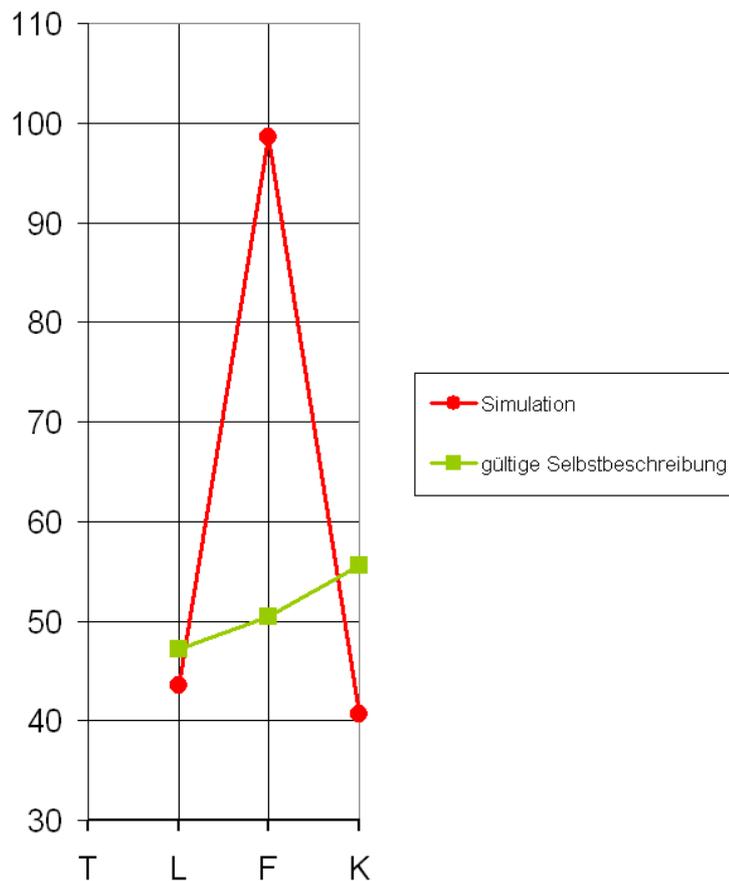
3.2.4 Profilvergleich von L, F, K

Bei der Zusammenführung der Mittelwerte unter Simulation und bei gültiger Selbstbeschreibung und einer graphischen Darstellung in Anlehnung an die Auswertungsbögen des MMPI wird das Veränderungsmuster der einzelnen Validitätsskalen besonders deutlich:

Tabelle 10: Veränderung der Validitätsskalen in der Synopsis

	Mittelwert	SD Mittelwert	Differenz
sim L	43,614	8,5	-3,629
echt L	47,243	7,53	
sim F	98,600	17,1	48,071
echt F	50,529	8,18	
sim K	40,643	7,83	-14,957
echt K	55,600	7,96	

Abb.1:



3.3 Die F-K-Differenz und ihre Beeinflussung durch Simulation

Der deutliche Anstieg des Mittelwertes auf der F-Skala unter Simulation bei gleichzeitigem Absinken des Mittelwertes auf der K-Skala findet seinen Niederschlag in einem gegenüber der gültigen Selbstbeschreibung hoch signifikanten Anstieg der F-K-Differenz.

Tabelle 11: Veränderung des F-K-Differenz durch Simulation

	Mittelwert	SD Mittelwert	Differenz	SD Differenz	t	sig. (2-seitig)
sim F-K	58,10	22,58	63,17**	28,80	18,35	0,000
echt F-K	-5,07	12,83				

** hoch signifikant $\leq 0,01$

Gemäß dem Handbuch zur Deutschen Kurzfassung des MMPI ist eine F-K-Differenz von ≥ 10 T-Wert-Punkten als erster Hinweis auf Simulation zu werten. Zur Überprüfung der Sensibilität und Sensitivität dieses Cut-off-Wertes führten wir eine Häufigkeitsanalyse für die F-K-Differenzen unter Simulation und bei gültiger Selbstbeschreibung durch. Dabei ergab unsere Studie bei dem vorgegebenen Cut-off-Wert von 10 eine falsch-negative Einschätzung für drei der simulierten Kurven, während 11 Probanden mit gültiger Selbstbeschreibung als simulationsverdächtig galten. Die besten Ergebnisse fanden sich in unserer Untersuchung bei einem Cut-off-Wert von 19.

Tab.12: Häufigkeit der F-K-Differenzwerte simuliert und echt

simuliert			echt		
F - K Differenz	Häufigkeit	Cumulative Prozent	F - K Differenz	Häufigkeit	Cumulative Prozent
-3	1	1,4	-27	1	1,4
0	1	2,9	-25	1	2,9
7	1	4,3	-23	2	5,7
11	1	5,7	-21	3	10,0
20	1	7,1	-19	4	15,7
22	1	8,6	-17	8	27,1
25	1	10,0	-15	3	31,4
30	2	12,9	-14	2	34,3
32	1	14,3	-12	1	35,7
34	1	15,7	-10	4	41,4
37	1	17,1	-9	1	42,9
38	2	20,0	-8	3	47,1
39	2	22,9	-7	4	52,9
40	1	24,3	-6	2	55,7
42	2	27,1	-5	1	57,1
45	1	28,6	-4	2	60,0
46	2	31,4	-3	3	64,3
47	1	32,9	-1	2	67,1

49	2	35,7	2	1	68,6
50	1	37,1	5	4	74,3
51	1	38,6	6	1	75,7
54	1	40,0	7	1	77,1
55	1	41,4	8	4	82,9
56	3	45,7	9	1	84,3
58	1	47,1	10	3	88,6
64	1	48,6	11	3	92,9
66	1	50,0	13	1	94,3
67	4	55,7	15	1	95,7
68	3	60,0	18	1	97,1
71	2	62,9	26	1	98,6
72	3	67,1	29	1	100,0
73	3	71,4			
75	1	72,9			
76	1	74,3			
77	2	77,1			
78	1	78,6			
79	1	80,0			
80	1	81,4			
81	4	87,1			
82	1	88,6			
83	1	90,0			
84	3	94,3			
86	1	95,7			
88	1	97,1			
89	1	98,6			
91	1	100,0			
Gesamt	70		Gesamt	70	

— empfohlener Cut-off-Wert
— günstigster Cut-off-Wert

Tabelle 13: Sensibilität und Sensitivität bei unterschiedlichen Cut-off-Werten

Cut-off F-K		
Cut-off 10	falsch positiv	falsch negativ
simuliert		N3 4,3%
echt	N11 15,7%	

Cut-off F-K		
Cut-off 19	falsch positiv	falsch negativ
simuliert		N4 5,7%
echt	N2 1,4%	

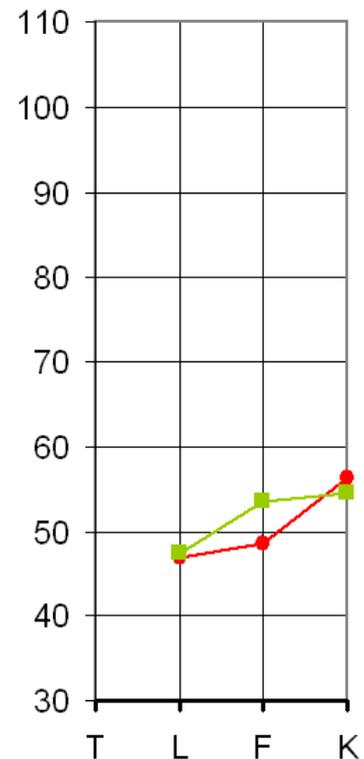
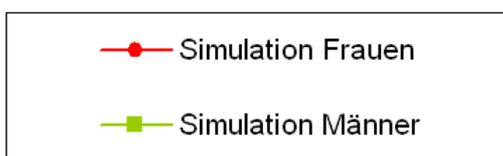
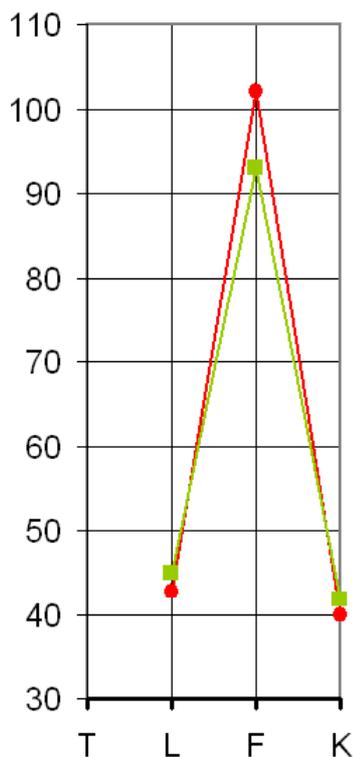
3.4 Geschlechterspezifische Beeinflussung der Validitätsskalen L, F, K und der F-K-Differenz

Differenziert nach dem Geschlecht, fanden sich Unterschiede vor allem auf der F-Skala. Während die Frauen bei der gültigen Selbstbeschreibung mit 48,62 im Mittel einen niedrigeren Wert hatten als die Männer mit 53,55, lag ihr F-Wert unter Simulation mit 102,00 T-Wert-Punkten deutlich über dem der männlichen Probanden mit 92,85. Bei den K-Werten zeigte sich eine umgekehrte Tendenz. Hier hatten die

Frauen bei der gültigen Selbstbeschreibung höhere Werte als die Männer, unter Simulation sanken die K-Werte im Mittel unter die der männlichen Probanden. Diese Ergebnisse addieren sich in der F-K-Differenz, die die geschlechtsspezifischen Unterschiede im Simulationsverhalten noch einmal verdeutlicht.

Tabelle 14: Veränderung der Validitätsskalen nach Geschlechtern differenziert

	Frauen		Männer	
	Mittelwert	SD	Mittelwert	SD
sim L	42,84	9,24	44,92	7,05
echt L	47,04	7,15	47,55	8,23
sim F	102,00	18,58	92,85	12,60
echt F	48,62	5,57	53,55	10,58
sim K	39,93	8,65	41,85	6,19
echt K	56,23	7,88	54,59	8,11
sim F - K	62,07	25,71	51,00	16,11
echt F - K	-7,60	11,11	-1,03	14,47



In Hinblick auf die Unterschiede zwischen den F-Werten männlicher und weiblicher Probanden führten wir für diese Skala eine Häufigkeitenanalyse getrennt nach den Geschlechtern durch.

Mit dem in Handbuch zur Deutschen Kurzform des MMPI vorgeschlagenen Cut-off-Wert von >60 konnten 100% der simulierten Kurven bei den Frauen korrekt identifiziert werden. Bei den gültigen Selbstbeschreibungen fand sich ein falsch-positives Ergebnis. 60 erwies sich damit bei Frauen als der günstigste Cut-off-Wert.

Tabelle 16: Häufigkeitenanalyse F-Skala bei weiblichen Probanden

F simuliert		Frauen	F echt		Frauen
T-Wert F -Skala	Häufigkeit	Cumulative Prozent	T-Wert F-Skala	Häufigkeit	Cumulative Prozent
62	2	4,7	43	12	27,9
69	3	11,6	47	10	51,2
75	2	16,3	50	14	83,7
78	1	18,6	53	3	90,7
81	1	20,9	56	1	93,0
87	1	23,3	59	2	97,7
94	1	25,6	72	1	100
97	3	32,6			
100	2	37,2			
103	2	41,9			
106	2	46,5			
109	2	51,2			
113	3	58,1			
116	10	81,4			
119	8	100,0			
Gesamt	43		Gesamt	43	

Tabelle 17: Cut-off-Werte F-Skala für Frauen

Cut-off F Frauen		
Cut-off 60	falsch positiv	falsch negativ
simuliert		N0
echt	N1 2,3%	

Bei den männlichen Probanden führte die Festlegung des Cut-off-Wertes auf 60 zu einem falsch-negativen und fünf falsch-positiven Ergebnissen. Der günstigste Cut-off-Wert für die Männer lag in unserer Untersuchung bei 73.

Tabelle 18: Häufigkeitenanalyse F-Skala bei männlichen Probanden

F simuliert		Männer	F echt		Männer
T-Wert F -Skala	Häufigkeit	Cumulative Prozent	T-Wert F-Skala	Häufigkeit	Cumulative Prozent
57	1	3,7	44	4	14,8
74	1	7,4	46	8	44,4
80	2	14,8	49	2	51,9
82	2	22,2	52	1	55,6
85	3	33,3	55	6	77,8
88	1	37,0	60	1	81,5
91	2	44,4	63	1	85,2
96	2	51,9	69	1	88,9
99	4	66,7	71	1	92,6
102	3	77,8	77	1	96,3
104	3	88,9	82	1	100,0
107	1	92,6			
110	2	100,0			
Gesamt	27		Gesamt	27	

— empfohlener Cut-off-Wert

— günstigster Cut-off-Wert

Tabelle 19: Cut-off-Werte F-Skala für Männer

Cut-off F Frauen			Cut-off F Frauen		
Cut-off 60	falsch positiv	falsch negativ	Cut-off 73	falsch positiv	falsch negativ
simuliert		N1 3,7%	simuliert		N1 3,7%
echt	N5 18,5%		echt	N2 7,4%	

Auch die Häufigkeitenanalyse der F-K-Differenzen führten wir ergänzend getrennt für weibliche und männliche Probanden durch. Bei den Frauen ergab der vorgegebene Cut-off-Wert von 10 gleichermaßen 3 falsch-positive wie 3 falsch-negative Ergebnisse. Der günstigste Cut-off-Wert ließ sich für die Frauen bei 16 festlegen.

Tabelle 20: Häufigkeit der F-K-Differenzen bei weiblichen Probanden simuliert und echt

F - K simuliert		Frauen		F - K echt		Frauen	
F - K Differenz	Häufigkeit	Cumulative Prozent		F - K Differenz	Häufigkeit	Cumulative Prozent	
-3	1	2,3		-27	1	2,3	
0	1	4,7		-25	1	4,7	
7	1	7,0		-23	2	9,3	
20	1	9,3		-21	1	11,6	
22	1	11,6		-19	3	18,6	
25	1	14,0		-17	5	30,2	
30	1	16,3		-15	2	34,9	
34	1	18,6		-14	2	39,5	
37	1	20,9		-12	1	41,9	
46	2	25,6		-10	1	44,2	
49	1	27,9		-9	1	46,5	
50	1	30,2		-8	2	51,2	
56	1	32,6		-7	4	60,5	
58	1	34,9		-5	1	62,8	
64	1	37,2		-3	3	69,8	
67	2	41,9		-1	2	74,4	
68	1	44,2		2	1	76,7	
71	2	48,8		5	3	83,7	
72	1	51,2		6	1	86,0	
73	2	55,8		8	3	93,0	
75	1	58,1		10	1	95,3	
77	2	62,8		11	1	97,7	
78	1	65,1		15	1	100,0	
79	1	67,4					
80	1	69,8					
81	4	79,1					
82	1	81,4					
83	1	83,7					
84	3	90,7					
86	1	93,0					
88	1	95,3					
89	1	97,7					
91	1	100,0					
Gesamt	43			Gesamt	43		

empfohlener Cut-off-Wert

günstigster Cut-off-Wert

Tabelle 21: Cut-off-Werte F-K für Frauen

Cut-off F-K Frauen		
Cut-off 10	falsch positiv	falsch negativ
simuliert		N3 7%
Echt	N3 7%	

Cut-off F-K Frauen		
Cut-off 16	falsch positiv	falsch negativ
simuliert		N3 7%
echt	N0	

Bei den Männern wurden mit dem vorgegeben Cut-off-Wert von 10 alle Simulanten richtig erkannt, während 8 gültige Selbstbeschreibungen falsch-positiv eingeschätzt wurden. Das günstigste Verhältnis von falsch-positiven zu falsch-negativen Probanden ergab sich durch die Anhebung des Cut-off-Wertes auf 29.

*Tabelle 22: Häufigkeit der F-K-Differenzen bei männlichen Probanden
simuliert und echt*

simuliert			echt		
F - K	Häufigkeit	Cumulative Prozent	F - K	Häufigkeit	Cumulative Prozent
11	1	3,7	-21	2	7,4
30	1	7,4	-19	1	11,1
32	1	11,1	-17	3	22,2
38	2	18,5	-15	1	25,9
39	2	25,9	-10	3	37,0
40	1	29,6	-8	1	40,7
42	2	37,0	-6	2	48,1
45	1	40,7	-4	2	55,6
47	1	44,4	5	1	59,3
49	1	48,1	7	1	63,0
51	1	51,9	8	1	66,7
54	1	55,6	9	1	70,4
55	1	59,3	10	2	77,8
56	2	66,7	11	2	85,2
66	1	70,4	13	1	88,9
67	2	77,8	18	1	92,6
68	2	85,2	26	1	96,3
72	2	92,6	29	1	100,0
73	1	96,3			
76	1	100,0			
Gesamt	27		Gesamt	27	

— empfohlener Cut-off-Wert

— günstigster Cut-off-Wert

Tabelle 23: Cut-off-Werte F-K für Männern

Cut-off Männer		
Cut-off 10	falsch positiv	falsch negativ
simuliert		N0
Echt	N8 29,6%	

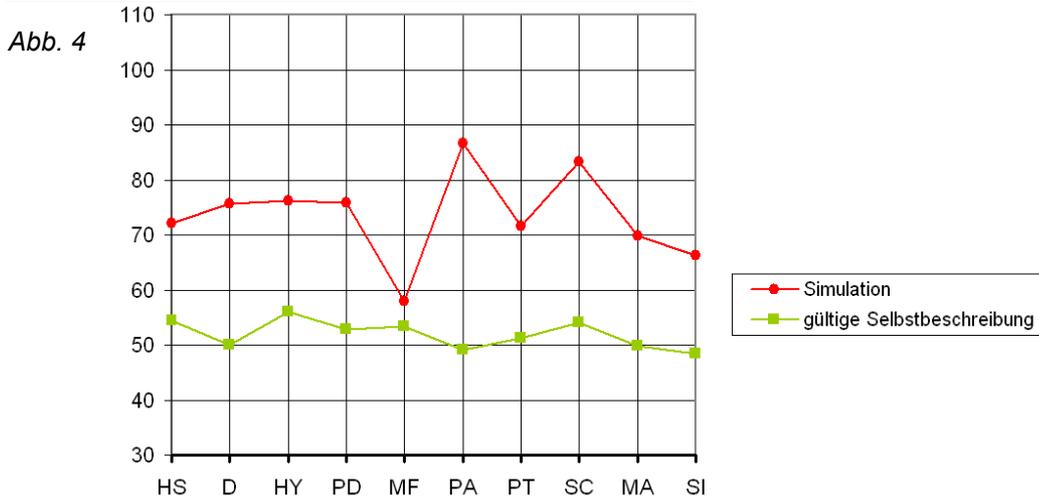
Cut-off Männer		
Cut-off 29	falsch positiv	falsch negativ
simuliert		N1 3,7%
echt	N1 3,7%	

3.5 Die klinischen Skalen und ihre Beeinflussung durch Simulation

Bei den klinischen Skalen lagen im Gesamtkollektiv die Werte der gültigen Selbstbeschreibung sämtlich unter denen bei Simulation. Der höchste Anstieg bei Simulation fand sich auf den Skalen Pa und Sc.

Tabelle 24: Veränderung der klinischen Skalen durch Simulation

	Mittelwert	SD Mittelwert	Differenz	SD Differenz
sim Hs	72,15	10,05	17,74	13,21
echt Hs	54,41	9,57		
sim D	75,77	10,68	25,73	14,16
echt D	50,04	8,33		
sim Hy	76,23	8,80	20,24	11,77
echt Hy	55,99	6,91		
sim Pd	75,82	8,21	23,01	12,05
echt Pd	52,81	8,27		
sim MF	57,96	10,10	4,69	13,32
echt MF	53,27	7,98		
sim Pa	86,59	13,13	37,53	18,35
echt Pa	49,06	9,73		
sim Pt	71,65	10,38	20,35	12,23
echt Pt	51,30	10,26		
sim Sc	83,26	12,29	29,23	14,64
echt Sc	54,02	8,72		
sim Ma	69,92	10,58	20,00	15,03
echt Ma	49,92	7,82		
sim Si	66,21	8,24	17,79	12,27
echt Si	48,42	7,33		

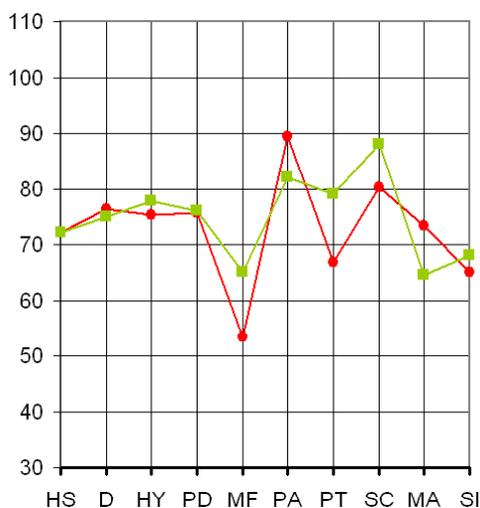


Bei der nach Geschlechtern differenzierten Beurteilung der klinischen Skalen fällt auf, daß die Werte der männlichen Probanden sämtlich über denen der Frauen liegen. Bei Simulation lagen die Werte der Frauen hingegen auf 4 der 10 klinischen Skalen (Hs, D, Pa, Ma) über denen der Männer.

Tabelle 25: Werte der klinischen Skalen unter Simulation nach Geschlechtern differenziert

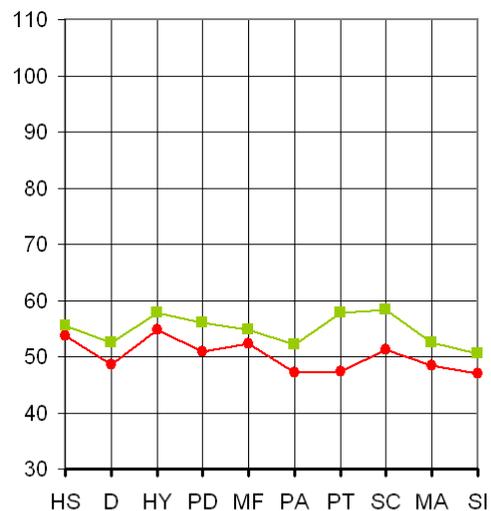
	Frauen		Männer	
	Mittelwert	SD	Mittelwert	SD
sim Hs	72,17	9,34	72,11	11,28
echt Hs	53,73	7,99	55,50	11,74
sim D	76,30	11,61	74,93	9,16
echt D	48,51	7,81	52,48	8,69
sim Hy	75,26	8,51	77,78	9,19
echt Hy	54,84	5,52	57,81	8,47
sim Pd	75,63	8,89	76,13	7,14
echt Pd	50,78	7,48	56,04	8,86
sim MF	53,42	7,08	65,19	10,06
echt MF	52,35	7,67	54,74	8,40
sim Pa	89,44	13,54	82,04	11,26
echt Pa	47,07	8,36	52,22	11,02
sim Pt	66,96	8,44	79,11	8,79
echt Pt	47,27	7,06	57,72	11,36
sim Sc	80,24	11,38	88,06	12,34
echt Sc	51,34	5,97	58,29	10,64
sim Ma	73,34	11,15	64,48	6,84
echt Ma	48,37	8,34	52,39	6,28
sim Si	65,05	9,02	68,07	6,63
echt Si	47,03	7,01	50,63	7,41

Abb. 5



—●— Simulation Frauen
—■— Simulation Männer

Abb. 6



—●— gültige Selbstbeschreibung Frauen
—■— gültige Selbstbeschreibung Männer

4. Diskussion

Ziel der Untersuchung war es festzustellen, wie sich auf Simulation gerichtete absichtliche Verfälschungstendenzen auf das MMPI-Ergebnis auswirken. Das Hauptaugenmerk galt dabei den Validitätsskalen L, F, K.

4.1 Zur Methode

Die untersuchte Gruppe setzte sich zusammen aus 61,4 % Frauen (N 43) und 38,6 % Männern (N 27). Vor allem die frühen amerikanischen Studien wurden z.T. mit ausschließlich männlichen Versuchspersonen durchgeführt (23, 37) oder es finden sich keine Angaben zum Geschlecht der Versuchspersonen (10, 24, 25). Die überwiegende Anzahl der zum Vergleich herangezogenen Studien arbeitete jedoch mit gemischtgeschlechtlichen Gruppen, wobei die prozentuale Verteilung schwankte. In den neueren Untersuchungen an Studenten fand sich z.T. ebenfalls eine deutliche Überzahl weiblicher Probanden (z.B. 2, 64, 72). Auf ein mögliches geschlechtsspezifisches Simulationsverhalten wurde jedoch in den allermeisten Studien nicht eingegangen.

Das Durchschnittsalter der Probanden in unserer Untersuchung entsprach im wesentlichen dem in vergleichbaren amerikanischen Studien. Differenziert nach Männern und Frauen fand sich für beide Gruppen ein nahezu identischer Altersdurchschnitt (w 23,69 / m 23,48), allerdings lag die Standardabweichung bei den Frauen mit 6,575 deutlich über der der männlichen Probanden (SD 1,847). Dieser Effekt ergab sich im wesentlichen durch zwei Studentinnen im Alter von 40 und 60 Jahren.

Zur Intelligenz der Probanden fanden sich in den vorliegenden amerikanischen Studien keine Angaben. Wir ermittelten den IQ mithilfe des MWT-B, der gleichzeitig Schlüsse auf die Beherrschung der deutschen Sprache zulässt. Durchführung, Auswertung und Interpretation des Testes können als objektiv gelten (32). Sowohl in Bezug auf den mittleren IQ als auch auf die Standardabweichung ließen sich bei Frauen und Männern in unserer Untersuchung keine deutlichen Unterschiede feststellen. Für die gesamte Gruppe lag der mittlere IQ bei 117,53.

Neuere amerikanische Studien schlüsseln regelmäßig die ethnische Herkunft ihrer Probanden auf (z.B. White, African-American, Asian-American). Auf ein möglicherweise spezifisches Simulationsverhalten ethnischer Subgruppen wird jedoch nur in Ausnahmefällen eingegangen (z.B. 41). Nachdem wir durch den MWT ein ausreichendes Sprachverständnis auch bei allen nichtmuttersprachlichen Probanden sichergestellt hatten, blieb der ethnische Hintergrund bei der Auswertung unserer Untersuchung ebenfalls unberücksichtigt.

Die Durchführung als abhängiger 2-Gruppenplan bot sich an, um im Rahmen des Kurses den Probanden die Veränderungen des MMPI-Profiles durch Simulation anhand der jeweils eigenen Werte zu verdeutlichen. Durch das Auswerten der jeweils eigenen Bögen wurde zudem der Anonymisierung in höchstem Maße Rechnung getragen. Obwohl bei Untersuchungen zum Thema Simulation häufig unabhängige Stichproben durchgeführt werden, finden sich vergleichbare gepaarte Stichprobenuntersuchungen an Studenten auch bei mehreren amerikanischen Grundlagenstudien (7, 10, 16, 26, 49). Bei der Aufgabenstellung blieb es den Probanden überlassen, welche Form der Beantwortung – simuliert oder gültig – sie zuerst vornehmen wollten. Nachdem die frühe Studie von Cofer gezeigt hatte, dass bei gepaarten Stichproben die Reihenfolge der Beantwortung ohne Bedeutung für das Ergebnis ist (10), wurde dieser Aspekt auch in zahlreichen weiteren Studien vernachlässigt

Die Gesamtform des MMPI wird aufgrund ihres Umfanges von 566 Items und des damit verbundenen Zeitaufwandes zur Bearbeitung und Auswertung kaum zu Screening-Untersuchungen angewandt (61). Wir verwendeten für unsere Untersuchung die deutlich anwenderfreundlichere Deutsche Kurzform des MMPI, die leider inzwischen nicht mehr auf dem Markt ist. Die Korrelation zwischen Kurzform und Gesamtform auf Skalenebene sind alle hoch: bei den klinischen Skalen D, Hy, Pd, MF, Pa, Ma ist $r > 0,8$; bei den Validitätsskalen L, K und F und den klinischen Skalen Hs, Pt, Sc und Si ist $r > 0,9$. Die Validation der Kurzform ergab zudem, dass nicht nur die Mittelwerte, sondern auch die Streuung der Skalenwerte identisch sind. Somit kann von einer großen Übereinstimmung zwischen Werten der Deutschen Kurzform und der Gesamtform des MMPI ausgegangen werden (21).

Die Aufgabenstellung in unserer Studie orientiert sich an den Kriterien der Schuldfähigkeit. Sie ist somit im Hinblick auf das Krankheitsbild ungerichtet – den Studenten war es freigestellt, beispielsweise eine akute Psychose als schwere seelische Störung oder Schwachsinn zu simulieren. Ausdrücklich wurde jedoch aufgegeben, eine so ausgeprägte Störung vorzugeben, dass der Proband für sein Handeln nicht mehr zur Verantwortung gezogen werden kann. Diese Aufgabenstellung kommt dem ebenfalls offengehaltenen „fake-bad“-Begriff vieler amerikanischen Studien sehr nahe, zumal dort häufig die Aufgabenstellung um den Aspekt der fehlenden Verantwortlichkeit erweitert wird (z.B. 16, 29, 72). Voraussetzung für eine solche Aufgabenstellung ist die „Naivität“ der Probanden, die bei der Simulation ihre eigene persönliche Vorstellung von einer schweren psychischen Störung wiedergeben, so dass nahezu alle entsprechenden Studien mit Studenten im Vorstudium arbeiten.

4.2 Zum Einfluss von Simulation auf die Validitätsskalen L, F, K und die F-K-Differenz

Mit drei (in der Frühzeit mit der Skala „?“ sogar vier) Validitätsskalen setzte der MMPI als erster Persönlichkeitsfragebogen einen eindeutigen Schwerpunkt auf die Diagnostik verfälschenden Antwortverhaltens. Als Ziel wurde formuliert, das Kontinuum zwischen Simulation („fake-bad“) zur Dissimulation („fake-good“) abzubilden, mit Übertreibungen und Bagatellisierung in verschiedener Ausprägung als Zwischenschritten.

4.2.1 L-Skala

Für den Bereich der übertrieben positiven Selbstdarstellung, bis hin zur Dissimulation, wurde die L-Skala konstruiert. Zu erwarten wäre demgemäß bei Simulation ein gegenüber der Norm niedrigerer Wert auf der L-Skala. Diese Annahme bestätigt sich in unserer Untersuchung durch einen mittleren Abfall von 47,24 auf 43,61 T-Wertpunkten. Die Differenz war bei zweiseitigem t-Test signifikant. Auch Wetter und Shores fanden in ihren Studien an Studenten bei Simulation einen Abfall auf der L-Skala (64, 72). Hawk, der die MMPI von Häftlingen verglich und dabei „entlarvte“ Simulanten gesunden und psychisch kranken „ehrlichen“ Inhaftierten gegenüberstellte, fand ebenfalls bei den Simulanten einen niedrigeren L-Wert als bei der authentischen Vergleichsgruppe (32). Verschiedene Untersuchungen an Studenten erbrachten jedoch auch entgegen gesetzte Veränderungen. So fand sich bei Exner et al. (16) unter Simulation ein L-Anstieg von 43,32 auf 47,16 und bei Bagby et al. (2) von 47,44 auf 49,33. Diese Studien untersuchten neben den Auswirkungen von „fake-bad“ auch die Veränderungen bei „fake-good“. Im direkten Vergleich zeigte sich, dass bei Dissimulation ein noch wesentlich stärkerer Anstieg der L-Werte zu beobachten war. Unter Berücksichtigung der oben genannten Ergebnisse ließe sich postulieren, dass ein Abfall der Werte auf der L-Skala zwar für Simulation spricht, aber zur Verdachtsdiagnose Simulation nicht gefordert werden darf. Dementsprechend spielt die L-Skala generell bei der Simulationsdiagnostik eine untergeordnete Rolle. Auch die hohe Fehlerquote von 71,5% bei unserem Versuch mittels eines Cut-off-Wertes auf der L-Skala simulierte Kurven von gültiger Selbstbeschreibung zu unterscheiden spricht gegen die Eignung der L-Skala zur Simulationsdiagnostik.

4.2.2 F-Skala

Durch hohe Werte auf der F-Skala werden sowohl Unverständnis der Fragen wie auch nachlässiges, „zufälliges“ Ausfüllen und Simulation abgebildet. Die erste Möglichkeit wurde in unserer Untersuchung durch den vorgeschalteten Mehrfachwahlwortschatztest ausgeschlossen. Allerdings fanden sich bei der gültigen

Selbstbeschreibung sechs Probanden mit F-Werten über 60. Da der F-Wert dieser Probanden bei Simulationen jeweils noch darüber hinaus erhöht war, kann angenommen werden, dass die Aufgabenstellung verstanden und zumindest die Simulation korrekt ausgeführt wurde. Darüber, ob die betreffenden Studenten die gültige Selbstbeschreibung nach dem Zufallsprinzip ausfüllten oder bewusst verfälschten, um möglichst wenig von sich preiszugeben, oder ob der F-Wert im Rahmen einer latenten psychischen Störung erhöht war, wie bei Gynther (30) beschrieben, kann nur spekuliert werden.

Lediglich ein Proband hatte auch bei Simulation einen F-Wert <60 , wobei sein F-Wert bei gültiger Selbstbeschreibung noch darunter lag.

Trotz der genannten einzelnen hohen F-Werte lag der Durchschnittswert für die F-Skala bei gültiger Selbstbeschreibung in unserer Untersuchung mit 50,42 T-Wert-Punkten nahe dem Mittelwert der Normierungsstichprobe und im unteren Bereich der vergleichbaren Studien, bei denen die F-Werte zwischen 46,3 und 63,65 angesiedelt waren (64, 2). Deutlich höher lag der F-Wert mit 72 T-Wert-Punkten bei den nachweislich psychisch gesunden und nichtsimulierenden Häftlingen in der Untersuchung von Hawk et al. (31).

Unter Simulation fand sich in unserer Untersuchung ein sehr deutlicher Anstieg auf 98,6 T-Wert-Punkte. Allerdings bleibt dieser Mittelwert unter dem der vergleichbaren amerikanischen Studien. Vor allem Wetter et al. (119,37), Shores et al. (118,6) und Graham et al. (117,5) fanden wesentlich höhere Werte (72, 64, 26).

Deutlich niedriger lag der durchschnittliche F-Wert bei Simulation in der frühen Studie von Gough, der seine Untersuchung an psychiatrisch erfahrenen „Experten“ (Psychiater, Psychologen, Sozialarbeiter an psychischen Kliniken) durchführte (23). Bei diesem Kollektiv lag der F-Wert bei gültiger Selbstbeschreibung bei 52,18 T-Wert-Punkten, Simulation hatte einen Anstieg auf 79,87 zur Folge. Der letztgenannte Wert kommt den F-Werten, die bei psychiatrischen Patienten erhoben wurden, bereits sehr nahe. In der Untersuchung von Bagby et al. lag der F-Werte der Patientengruppe bei 71,84, in der Studie von Bagby et al. sogar bei 79,7. Diese Annäherung der Werte macht deutlich, dass bei Simulationsverdacht insbesondere mäßig erhöhte F-Werte eine zusätzliche Diagnostik erfordern, um das tatsächliche Vorliegen einer psychiatrischen Erkrankung auszuschließen bzw. zu beweisen.

Die Interpretationshilfe zur Deutschen Kurzform des MMPI arbeitet bei der Simulationsdiagnostik mit der Verknüpfung zweier Parameter (F-K und F), die in den amerikanischen Studien stets getrennt betrachtet werden. Sowohl in der Metaanalyse von Berry et al. zum MMPI I als auch in der Metaanalyse von Rogers et al. zum MMPI II zeigte sich die F-Skala zur Beurteilung von Simulation allen anderen Skalen

überlegen. In unserer Studie lag die Gesamtfehlerquote unter der alleinigen Voraussetzung des günstigsten Cut-off-Wertes von $F > 60$ bei 10% (N7).

4.2.3 K-Skala

Mit einem durchschnittlichen K-Wert von 55,60 T-Wert-Punkten bei der gültigen Selbstbeschreibung lag unsere Untersuchung im Vergleich mit amerikanischen Studien im mittleren Bereich. Sechs ähnliche Untersuchungen erbrachten hierzu Werte zwischen 45,01 (2) und 57,44 (16). Obwohl Grow et al. darauf hinweisen, dass K-Scores bei Studenten gewöhnlich ein Sigma höher liegen als in der Normalpopulation (29), lagen auch die Ergebnisse der Untersuchung an Häftlingen bei Iverson et al. (38) und Hawk et al. (31) mit 46,5 bzw. 51 in diesem Bereich. Auffallend hohe Werte auf der K-Skala fanden sich bei den von Gough untersuchten „Experten“, bei ihnen lag der K-Wert bei gültiger Selbstbeschreibung im Mittel bei 62, Werte über 60, die für eine übertrieben positive Selbstdarstellung sprechen, werden ansonsten nur vereinzelt bei „fake-good“ gefunden (16).

Auch der K-Wert bei Simulation lag bei unserer Untersuchung mit 40,64 T-Wert-Punkten im mittleren Bereich. Die Werte der amerikanischen Studien an Studenten, variierten zwischen 35,49 (72) und 44,04 (10). Das Absinken der K-Scores unter Simulation stimmt mit der Annahme überein, dass niedrige K-Werte bei psychisch Gesunden zu erwarten sind, die pathologische Tendenzen in übertriebener Weise herausstellen.

Obwohl bei allen Untersuchungen an Studenten ein Abfall des K-Wertes unter Simulation zu beobachten ist, lassen sich keine Schlüsse aus der absoluten Höhe des K-Wertes ziehen. So lagen der K-Wert bei gültiger Selbstbeschreibung in der Studie von Bagby et al. (45,01) und der Wert unter Simulation bei Exner et al. (44,04) nur wenig auseinander (2, 16). Hinzu kommt, dass bei Kenntnis psychischer Erkrankungen zum Beispiel bei den Experten in der Studie von Gough der K-Wert unter Simulation noch höher ($K=45$) liegen kann (23). Das selbe gilt für eine weniger spielerische Beschäftigung mit dem Thema wie bei den „echt simulierenden“ Häftlingen der Hawk-Studie, hier lag der K-Wert bei 46 (31). Entsprechend wurde von Graham et al. (26) und Cassisi et al. (9) formuliert, dass die K-Skala für sich genommen zum Nachweis von Simulation nicht geeignet ist.

Die Festsetzung eines Cut-off-Wertes zur Differenzierung simulierter und gültiger Kurven erbrachte in unserer Untersuchung eine Fehlerquote von 45,8%. Auch unter diesem Aspekt kann die K-Skala allein wenig zur Simulationsdiagnostik beitragen.

4.2.4 Profilvergleich L, F, K

Bei der Handauswertung der MMPI-Fragebögen wird zunächst die Anzahl der auffällig beantworteten Items für jede Skala ermittelt. Anschließend werden die Werte in das Koordinatensystem der Profilbögen übertragen. Durch das Verbinden der einzelnen Markierungen ergibt sich eine Kurve, die im Bereich der Validitätsskalen unter Simulation ein charakteristisches zackenartiges Profil aufweist. Auch in zahlreichen amerikanischen Veröffentlichungen wird explizit durch Graphiken auf diesen Kurvenverlauf hingewiesen, durch den auffällige Werte in besonders leicht erkennbarer Form visualisiert werden (z.B. 11, 14, 26, 27, 31, 64, 73). Das Koordinatensystem in Abb.1 orientiert sich an den oben beschriebenen Profilbögen, so dass die Kurvenverläufe in dieser Arbeit den optischen Eindruck von Simulation und Normalbefund bei der Handauswertung des MMPI wiedergeben.

4.2.5 F-K- Differenz

Gough beschrieb erstmals 1947 die Anwendung des F-K-Index und führte diese 1950 weiter aus (23, 24). In den folgenden Jahren beschäftigten sich zahlreiche weitere Studien mit dieser Größe (z.B. 10, 37), wobei in sämtlichen Untersuchungen zur Berechnung des F-K-Index die Rohwerte herangezogen wurden, auch wenn die übrigen Skalen anhand ihrer T-Werte beurteilt wurden. In unserer Untersuchung wurde die F-K-Differenz hingegen gemäß der Interpretationshilfe zur Deutschen Kurzform des MMPI anhand der T-Werte für die F- und die K-Skala ermittelt.

Auch wenn es technisch problemlos möglich gewesen wäre, in unserer Untersuchung zum Erzielen einer formal besseren Vergleichbarkeit auf die Rohwerte zurückzugreifen, sprachen statistische Gründe gegen dieses Vorgehen. Zwar enthält die K-Skala sowohl in der amerikanischen Langform als auch in der Deutschen Kurzform 22 Fragen, die F-Skala ist in der ursprünglichen Version mit 64 Fragen jedoch deutlich umfangreicher als in der von uns verwendeten Kurzform, in der sie nur 24 Fragen umfasst. Würde man nun die Rohwerte der Kurzform in Beziehung setzen, entspräche unter Simulation der K-Abfall zwar der Langform, auf der F-Skala wären jedoch deutlich niedrigere Werte zu erwarten. Auch diese Art der Berechnung würde also keine vergleichbaren Werte liefern. Die Ergebnisse unserer Untersuchung können also nur in ihrer Tendenz mit den amerikanischen Studien verglichen werden.

Diese Tendenz findet ihren Ausdruck in erster Linie durch das Vorzeichen. So zeigt der negative F-K-Index bei der gültigen Selbstbeschreibung in unserer Untersuchung auf den ersten Blick, dass der K-Wert hier im Mittel über dem F-Wert lag, während der positive F-K-Index bei Simulation die gegenüber K erhöhten F-Werte darstellt. In seiner Grundlagenstudie zur Einführung des F-K-Index fand Gough diese Skalenverhältnisse bei 85,3 % der authentischen Probanden und bei 88,4 % der Simulanten (24). Bei der gültigen Selbstbeschreibung ergab sich in der Mehrzahl der

vergleichbaren Studien ein negativer F-K-Index. Lediglich in der Studie von Wetter et al. fand sich auch bei gültiger Selbstbeschreibung ein, wenn auch eher niedriger, positiver F-K-Index (72). Mögliche Erklärungen für dieses Phänomen werden in der Arbeit leider nicht diskutiert.

In Bezug auf den F-K-Index bei Simulation gab es keine grundlegenden Abweichungen zwischen den amerikanischen Studien und unserer Untersuchung. Es fanden sich sämtlich positive Durchschnittswerte, deren Höhe wie bei der vorliegenden Arbeit das deutliche Überwiegen der F- vor den K-Werten klar zum Ausdruck bringt. Unter Berücksichtigung der o.g. Studie von Wetter et al. lässt sich daher formulieren, dass ein negativer F-K-Index mit großer Sicherheit gegen Simulation spricht, unabhängig davon, ob mit der ursprünglichen Fassung oder mit der Deutschen Kurzform des MMPI gearbeitet wird, dass jedoch zum Simulationsnachweis nicht nur das Vorzeichen, sondern auch die Höhe des F-K-Index berücksichtigt werden muss.

4.2.6 Cut-off-Werte F-K

Wie bereits beschrieben, ist die Höhe der F-K-Differenz von ausschlaggebender Bedeutung für die Simulationsdiagnostik. Zahlreiche amerikanische Studien haben sich in den Jahrzehnten nach 1949 bemüht, durch die Festlegung geeigneter Cut-off-Werte die Anzahl falschpositiver und falschnegativer Verdachtsdiagnosen möglichst gering zu halten. Dabei schwanken die Empfehlungen zwischen 5 und 15 T-Wert-Punkten (10, 18). Rogers et al. geben in ihrer Metaanalyse des MMPI-2 als günstigsten Cut-off-Wert 10 an und sind damit nahe an der ursprünglichen Empfehlung von Gough, der den Cut-off-Wert bei 9 festlegte (56, 24). Auch die Interpretationshilfe zur Deutschen Kurzform des MMPI fordert für die Verdachtsdiagnose Simulation eine F-K-Differenz von ≥ 10 . Der Anschein einer engen Übereinstimmung mit den Ergebnissen der amerikanischen Studien relativiert sich jedoch dadurch, dass in den amerikanischen Untersuchungen wie oben beschrieben die Rohwerte korreliert werden, während die Deutsche Kurzform des MMPI mit den T-Werten arbeitet.

In Bezug auf die Sensitivität erwies sich der vom Handbuch vorgegebene Cut-off-Wert in unserer Untersuchung als sehr empfindlich. 95,7% (N67) der durch Simulation verfälschten MMPI wurden hiermit richtig identifiziert. Auf der anderen Seite wurden jedoch 15,7% (N11) der Probanden falsch positiv eingeordnet, wodurch sich eine Gesamtfehlerquote von 20% (N14) ergab.

Das günstigste Verhältnis von falschpositiv und falschnegativ diagnostizierten Probanden fand sich in unserer Untersuchung bei einem Cut-off-Wert von 19. Mit diesem wurden immer noch 94,3% (N66) der simulierten Kurven richtig erkannt,

während lediglich 2,9% der authentischen Probanden (N2) unter Simulationsverdacht gerieten, so dass sich eine Gesamtfehlerquote von 8,6% (N6) ergab.

4.3. Zu den geschlechterspezifischen Unterschieden bei Simulation bezogen auf die Validitätsskalen L, F, K und die F-K-Differenz

Der unter Simulation bei der Gesamtgruppe gefundene Abfall auf der L-Skala fand sich bei geschlechtsdifferenzierter Untersuchung sowohl bei den Frauen als auch bei den Männern wieder, wobei die Differenz zur gültigen Selbstbeschreibung bei den Frauen größer war. Bei ihnen war auf der L-Skala ein Abfall von 47,04 auf 42,94 T-Wert-Punkten zu beobachten, bei den Männern lediglich von 47,55 auf 44,92. Der Vergleich mit der Studie von Graham et al. zeigt, dass dort die Differenz bei den Frauen von 50,5 bei der gültigen Selbstbeschreibung auf 43,1 bei Simulation noch deutlicher ausgeprägt ist, während sich bei den Männern sogar ein leichter Anstieg von 49,2 auf 50,9 darstellt (26). Das wirft ein neues Licht auf die Betrachtung der übrigen Vergleichsstudien, wenn man in Betracht zieht, dass jene Untersuchungen, in denen ebenfalls ein L-Abfall beschrieben ist, mit überwiegend weiblichen Probanden arbeiteten (64, 72), während solche Studien, in denen ein L-Anstieg beobachtet wurde, geschlechtsausgeglichen waren (16) oder mehr männliche als weibliche Probanden beinhalteten (2, 9). Die Untersuchungen von Hawk et al. an Häftlingen hat hingegen gezeigt, dass bei männlichen echten Simulanten ebenfalls ein niedriger L-Wert zu erwarten ist (31). Dass amerikanische männliche Studenten bei Simulation davon abweichendes Antwortverhalten zeigen, während die männlichen Studenten unserer Untersuchung eher wie erwartet simulieren, unterstreicht die Bedeutung Kulturkreis spezifischer Studien.

Auf der F-Skala zeigte sich sowohl in der Gesamtgruppe als auch bei differenzierter Betrachtung der weiblichen und männlichen Probanden überall der zu erwartende deutliche F-Anstieg unter Simulation. Besonders ausgeprägt war der Anstieg bei den Frauen von 48,62 T-Wert-Punkte bei gültiger Selbstbeschreibung auf 102,0 bei Simulation. Bei den Männern war ein Anstieg von 53,55 auf 92,85 zu beobachten. Der Vergleich mit den amerikanischen Studien an Studenten zeigt bei der Betrachtung der F-Skala, wie schon bei der L-Skala, ebenfalls eine deutlichere Ausprägung der Differenz bei den Studien mit überwiegend weiblichen Probanden (64, 72). Dabei ergab sich die große Differenz nicht allein aus besonders hohen F-Werten unter Simulation, sondern auch durch auffallend niedrige Werten bei gültiger Selbstbeschreibung. Als Ursache für den letztgenannten Befund könnte vor dem Hintergrund des Aufbaus der F-Skala eine größere Sorgfalt der Frauen beim Beantworten des Fragebogens angenommen werden. Allerdings wird diese Tendenz in der Untersuchung von Graham nicht bestätigt, in der beide Subgruppen nahezu identische Werte bei gültiger Selbstbeschreibung wie bei Simulation zeigen.

Bei der nach Geschlechtern differenzierten Untersuchung der K-Skala fand sich kein signifikanter Unterschied zwischen den Werten der weiblichen und denen der männlichen Probanden. Tendenziell lagen bei der gültigen Selbstbeschreibung die K-Werte der Frauen mit 56,23 T-Wert-Punkten über denen der Männer mit 54,59, was im gegebenen Rahmen zwischen 50 und 60 gemäß dem Handbuch zur Deutschen Kurzform des MMPI für einen Hang zur positiven Selbstdarstellung bzw. einer Abwehr gegenüber dem Test spricht. Auch der erwartete Abfall des K-Wertes unter Simulation war bei den Frauen etwas deutlicher ausgeprägt (Frauen 39,93, Männer 41,85). Solche Unterschiede zwischen männlichem und weiblichem Simulationsverhalten stellte auch Graham et al. in seiner Studie dar. Vergleichbar hohe K-Werte bei gültiger Selbstbeschreibung fanden sich allerdings auch in Studien ohne auffallend hohen Frauenanteil (10, 16), desgleichen ähnlich niedrige Werte unter Simulation (3, 9), so dass davon ausgegangen werden kann, dass die Geschlechtszugehörigkeit eines Probanden bei der Beurteilung der K-Werte von untergeordneter Bedeutung ist.

Colby, der lediglich die gültigen Selbstbeschreibungen von Frauen und Männern verglich, stellte in seinem Artikel ebenfalls die L-F-K-Kurvenverläufe graphisch dar (11). Dabei ergaben sich im Vergleich mit unserer Untersuchung nahezu identische Profile.

Bei den F-K-Werten unter Simulation fand sich in unserer Untersuchung ein augenscheinlicher Unterschied zwischen den weiblichen und den männlichen Probanden. Dies legt die Vermutung nahe, dass es sinnvoll sein könnte, bei der Festlegung von Cut-off-Werten zur Simulationsdiagnostik das Geschlecht der Probanden zu berücksichtigen.

Auch Graham et al. formulierten in ihrer Studie, dass bei bestimmten Fragestellungen die optimalen Cut-off-Scores für Männer und Frauen unterschiedlich sein können (26). So konnten in ihrer Untersuchung mit einem Cut-off-Score von 13 90% der männlichen psychiatrischen Patienten korrekt identifiziert werden, während bei Frauen ein Cut-off-Wert von 22 notwendig war, um gleichermaßen eine Sensitivität von 90% zu erzielen. In Bezug auf Simulationsdiagnostik war der Unterschied in der Graham-Studie allerdings weniger deutlich.

Unter Verwendung des für die Deutsche Kurzform des MMPI vorgeschlagenen Cut-off-Wertes von 10 zeigte sich, dass sämtliche drei falschnegativ diagnostizierten Probanden weiblich waren. Die Falschpositiven setzten sich zusammen aus 3 Frauen und 8 Männern, was noch einmal die erhöhte Tendenz der Männer zu F-K-Werten mit positivem Vorzeichen bei gültiger Selbstbeobachtung unterstreicht.

Bei den Frauen ließ sich in unserer Untersuchung eine Optimierung des Cut-off-Wertes bereits durch eine Anhebung auf 16 erzielen. Dadurch fand sich zwar keine Verbesserung der Sensitivität, es konnte jedoch eine Reduktion der falschpositiv

bewerteten Probanden auf 0 erzielt werden, so dass eine Gesamtfehlerquote von 5,3% (N3) resultierte.

Bei den Männern ließ sich ein ähnliches Ergebnis nur mit einem deutlich höheren Cut-off-Wert erreichen. Durch die Anhebung des Wertes auf 29 konnte die Anzahl der falschpositiv bewerteten Probanden auf 1 reduziert werden. Ein Simulant (3,7%) wurde nun falsch negativ bewertet. Die Gesamtfehlerquote lag somit bei 3,4%.

Die Festlegung geschlechtsspezifischer Cut-off-Werte für die F-K-Differenz erscheint nach unseren Ergebnissen sinnvoll, angesichts der relativ geringen Aussagekraft einer einzigen Untersuchung bedarf es aber weiterer Untersuchungen und klinischer Belege.

4.4. Zum Einfluss von Simulation auf die klinischen Skalen

Die Werte der klinischen Skalen lagen bei der gültigen Selbstbeschreibung in unserer Untersuchung zwischen 48,42 T-Wert-Punkten als niedrigstem Wert für Si/Soziale Introversion und 55,99 als höchsten Wert für Hy/Hysterie. Die älteren Studien beinhalten die erst 1946 eingeführte Si-Skala noch nicht, bei drei vergleichbaren neueren Studien (2, 64, 72) war Si ebenfalls der niedrigste Wert. Shores et al. führten auch Hy als höchsten Wert an (64).

Unter Simulation fand sich in unserer Untersuchung der höchste Anstieg auf der Skala Pa/Paranoia. Der Mittelwert lag hier bei 86,59 T-Wert-Punkten, die Differenz zur gültigen Selbstbeschreibung betrug 37,53. Eine zweite Spitze fand sich auf der Sc/Schizophrenia-Skala mit einem Mittelwert von 83,26 und einer Differenz von 29,23 T-Wert-Punkten zur gültigen Selbstbeschreibung. Dies spricht dafür, dass „naive“ Probanden mit forensisch relevanten psychiatrischen Erkrankungen am ehesten eine paranoid-halluzinatorische Schizophrenie assoziieren.

In den amerikanischen Studien ist unter Simulation generell Sc/Schizophrenia die am stärksten erhöhte Skala, und zwar sowohl bei den Untersuchungen an Studenten als auch bei Studien mit Häftlingen und bei Goughs Arbeit über „Experten“. Pa/Paranoia ist in allen diesen Studien an zweiter Stelle erhöht (z.B. 2, 10, 23, 31, 38, 64, 72). Dabei liegen die Werte in den amerikanischen Studien in der Regel über denen unserer Untersuchung. Spitzen bei Sc und Pa finden sich jedoch auch bei psychiatrischen Patienten in Studien von Graham et al. sowie Hawk et al. (26, 31). Die Werte bei den Patienten in diesen Studien lagen jedoch deutlich unter denen der amerikanischen Studenten im Simulationsversuch. Graham et al. kamen zu dem Schluss, dass Studenten bei Simulation dazu neigen, Symptome zu „over-reporten“. Gemessen an den Studenten zeigten sich die „Experten“ in Goughs Studie zurückhaltender (23). Doch auch sie lagen mit ihren Werten deutlich über denen von

psychiatrischen Patienten. Eine Abgrenzung der Werte unserer Untersuchung von den Werten der amerikanischen Patienten ist hingegen schwieriger. Hier wäre es sinnvoll, wenn zum Vergleich die Werte deutscher Patienten, erhoben mit der MMPI-Kurzform, hinzugezogen werden könnten.

Auch auf fünf weiteren klinischen Skalen fand sich bei unserer Untersuchung ein Anstieg der mittleren T-Werte auf über 70: Hy/Hysteria (76,23), Pd/Psychopathic Deviate (75,82), D/Depression (75,77), Hs/Hypochondriasis (72,15) und Pt/Psychasthenia (71,65). Ma/Hypomania (69,92) und Si/soziale Introversion-Extroversion (66,21) blieben unter der 70er Perzentile. Der MF-Wert war mit 57,96 T-Wert-Punkten unauffällig. In vergleichbaren Studien ließ sich kein einheitliches Muster im Anstieg dieser Werte erkennen. Überproportional häufig fand sich lediglich noch eine deutliche Erhöhung auf der Pd-Skala. Die Werte auf der Si-Skala blieben auch in amerikanischen Studien meist unter 70. Am wenigsten durch Simulation beeinflusst wurde in sämtlichen Studien die MF-Skala. Dies zeigt, dass die Geschlechtsidentität in der allgemeinen Vorstellung nicht mit psychiatrischen Erkrankungen in Zusammenhang gebracht wird.

Im Rahmen der Obvious-Subtle-Scales werden vor allem im amerikanischen Raum auch die klinischen Skalen zur Simulationsdiagnostik herangezogen. Ohne diese Spezifizierung kann die Betrachtung dieser Skalen jedoch lediglich grobe Hinweise geben. So wären im Rahmen einer Begutachtung bei einer Erhöhung von Pa und Sc über 80 T-Wert-Punkte deutliche Verhaltensauffälligkeiten zu erwarten. Durch eine ergänzende klinische Beobachtung könnte gegebenenfalls eine Diskrepanz zwischen angegebenen Symptomen und gezeigtem Verhalten festgestellt werden.

Bei der Aufschlüsselung nach Geschlechtern zeigte sich, dass die dominierende Höhe des Pa-Peaks im Gesamtkollektiv durch die hohen Werte der weiblichen Probanden auf dieser Skala bestimmt wird. Bei den Frauen war Pa im Mittel auf 89,34 T-Wert-Punkte erhöht, Sc auf 80,20. Bei den männlichen Probanden fand sich – wie in den amerikanischen Studien – die stärkste Erhöhung auf der Sc-Skala mit 88,42 und ein geringerer Anstieg auf Pa mit 81,92. Die Höhe des Sc-Peaks vor der des Pa-Peaks in amerikanischen Studien lässt sich jedoch nicht auf das Überwiegen männlicher Probanden zurückführen. In mehreren Studien überwogen dort gleichfalls Frauen (64, 72). Bei Graham (26), der die Werte von Männern und Frauen ebenfalls getrennt untersuchte, war auch bei den Frauen Sc mit 117,1 T-Wert-Punkten vor Pa mit 115,7 erhöht, wenn auch die Differenz weniger ausgeprägt war als bei den Männern (Sc 115,4, Pa 111,7).

Bei den Frauen in unserer Untersuchung fanden sich auf den übrigen klinischen Skalen weder im Vergleich mit den Männern, noch im Vergleich mit den Probanden der Graham-Studie Auffälligkeiten, mit denen der ungewöhnlich hohe Pa-Anstieg in Zusammenhang gebracht werden könnte. Eine nachträgliche Befragung der

Probandinnen zu den Fragen der Skalen Pa und Sc war aufgrund des anonymisierten Verfahrens nicht möglich. Nur weitere Untersuchungen können zeigen, ob deutsche Studentinnen generell forensisch relevante psychische Störungen vorwiegend auf der Pa-Skala simulieren.

5. Zusammenfassung

Simulation bedeutet, dass bewusst Symptome geschildert oder produziert werden, um einen äußeren Vorteil zu erreichen. Von besonderer Relevanz ist der Aspekt möglicher Simulation bei psychiatrischen Begutachtungen in denen der Selbsteinschätzung der Probanden eine vorherrschende Rolle zukommt.

Dem MMPI – im amerikanischen Raum der meist benutzte Selbstbeurteilungsbogen zur Diagnostik psychischer Störungen – wurden deshalb drei sogenannte Validitätsskalen vorgeschaltet. Diese werden mit L, F, K bezeichnet und erfassen auffälliges Antwortungsverhalten sowohl im Sinne von „fake bad“ (Simulation) als auch von „fake good“ (Dissimulation). Die Eignung dieser Skalen zur Simulationsdiagnostik wurden in Amerika immer wieder in Studien überprüft.

Gemäß dem Handbuch zur Deutschen Kurzfassung des MMPI sind bestimmte Auffälligkeiten auf diesen Skalen – eine F-K-Differenz von ≥ 10 in Verbindung mit einem F-Wert von > 60 T-Wert-Punkten als Hinweis auf bewusstes Verfälschen zu werten. Für den deutschen Sprachraum liegen bisher jedoch keine Studien zur Simulationsdiagnostik mit dem MMPI vor.

Die vorliegende Untersuchung beschäftigt sich mit der Frage, wie das MMPI-Ergebnis durch absichtliche Verfälschung im Sinne von Simulation beeinflusst wird. Hierzu wurde eine Untersuchung an Studenten in Form eines abhängigen 2-Gruppen-Plans realisiert. Der Versuchsplan orientierte sich an amerikanischen Grundlagenstudien zu diesem Thema.

Untersucht wurde ein Kollektiv von 43 Studentinnen und 27 Studenten des dritten vorklinischen Semesters Humanmedizin. Sie bearbeiteten im Rahmen des Kurses „medizinische Psychologie“ jeweils zwei mal die Deutsche Kurzform des MMPI: einmal mit der gültigen Selbstbeschreibung und einmal unter der Vorgabe eine psychische Erkrankung zu simulieren, die in ihrer Ausprägung Schuldunfähigkeit entspricht. Alle Studenten erfüllten diese Aufgabe.

Unsere Auswertung konzentrierte sich vor allem auf die drei Validitätsskalen. Auf allen drei Skalen fanden sich signifikante Unterschiede zwischen den Mittelwerten bei Simulation und bei gültiger Selbstbeschreibung.

Da erhöhte Werte auf der L-Skala für Dissimulation sprechen ist unter Simulation ein geringes Absinken der Mittelwerte zu erwarten. Diese fand sich in unserer Untersuchung, jedoch war die Differenz mit $-3,629$ zwischen dem Mittelwert bei gültiger Selbstbeschreibung (47,243) und unter Simulation (43,614) im Vergleich zur F- und K-Skala am geringsten ausgeprägt. Auch amerikanische Studien haben

gezeigt, dass die L-Skala zur Simulationsdiagnostik gegenüber der F-Skala und der F-K-Differenz vernachlässigt werden kann.

Auf der F-Skala fand sich im Mittel ein massiver Anstieg zwischen den Werten bei gültiger Selbstbeschreibung (43,614) und bei Simulation (98,6). Damit lag der Mittelwert bei Simulation deutlich über dem in der Interpretationshilfe zur Deutschen Kurzfassung des MMPI geforderten Wert von >60 T-Wert-Punkten. Dieser Parameter erwies sich als sehr sensibles Messinstrument zur Simulationsdiagnostik. Bei 69 Probanden (98,6%) stieg der Wert unter Simulation über 60 an. Die Spezifität war mit 6 falsch positiven Probanden (8,6%) weniger ausgeprägt. Mit einer Gesamtfehlerquote von N7 (10%) erwies sich jedoch die F-Skala auch für sich alleine genommen als geeignetes Instrument zur Simulationsdiagnostik. Dies deckt sich mit den Ergebnissen der amerikanischen Forschung, während die Interpretationshilfe zur Deutschen Kurzform des MMPI die F-Skala im Hinblick auf Simulation nicht isoliert betrachtet.

Parallel zum Anstieg auf der F-Skala zeigten sich auf der K-Skala bei Simulation mit 40,64 T-Wert-Punkten im Mittel deutlich niedrigere Werte als bei gültiger Selbstbeschreibung mit 55,60. Die K-Skala für sich genommen ist jedoch wenig aussagekräftig.

Bei der Handauswertung des MMPI mithilfe von Profildbögen ergibt sich durch die genannten Veränderungen auf den Validitätsskalen ein charakteristischer Kurvenverlauf, der auffällige Werte bei Simulation leicht erkennbar visualisiert.

Durch die entgegen gesetzte Veränderung der F- und der K-Werte bei Simulation nimmt die F-K-Differenz zu und verändert dabei in der Regel ihr Vorzeichen. In unserer Untersuchung fand sich bei gültiger Selbstbeschreibung eine mittlere F-K-Differenz von $-5,0714$, unter Simulation lag der Wert bei $57,957$. Die von der Interpretationshilfe im Handbuch zur Deutschen Kurzform des MMPI geforderte F-K-Differenz von ≥ 10 wurde lediglich von 3 Simulanten (4,3%) nicht erreicht. Auf der anderen Seite waren 13 Probanden (18,5%) bei gültiger Selbstbeschreibung auf diese Weise auffällig. Mit einer Gesamtfehlerquote von 22,9% war dieser Parameter wenig geeignet zur Simulationsdiagnostik. Eine Verbesserung der Spezifität ließ sich mit einer Erhöhung des Cut-off-Wertes für die F-K-Differenz auf 19 erzielen. Mit jetzt 4 falsch negativen und 2 falsch positiven Probanden (Gesamtfehlerquote N6 (8,6%)) erwies sich die F-K-Differenz nun der F-Skala sogar überlegen.

Bei der Untersuchung der Validitätsskalen nach Geschlechtern fand sich bei den Frauen unter Simulation eine noch stärkere Akzentuierung der typischen Veränderungen als bei den Männern: Bei den Frauen stieg der F-Wert im Mittel auf 102,00 T-Wert-Punkte bei den Männern lediglich auf 92,85. Auch das Absinken der Werte auf der K-Skala war bei den Frauen mit 39,93 stärker ausgeprägt als bei den

Männern mit 41,85. Die unterschiedliche Ausprägung auf den einzelnen Skalen schlug sich auch auf die F-K-Differenz nieder. Diese betrug bei den Frauen im Mittel 62,06, bei den Männern 51.

Insgesamt erwies sich der F-Wert als besonders geeignet zur Simulationsdiagnostik bei weiblichen Probanden. Alle Simulantinnen wurden mithilfe dieses Parameters korrekt identifiziert, bei den gültigen Selbstbeschreibungen fand sich ein falsch-negatives Ergebnis. Weniger günstig war das Ergebnis für die F-Skala bei den männlichen Probanden mit einer Fehlerquote von N6 (22,2%). Bei ihnen bewährte sich ein Cut-off-Wert von 73 mit einer Fehlerquote von N3 (11,1%).

Mit dem vorgegebenen Cut-off-Wert von 10 für die F-K-Differenz konnten 37 Frauen (86,1%) und 19 Männer (81,4%) korrekt identifiziert werden. Der deutlich geeignetere Cut-off-Wert für Frauen lag bei 18 (3 falsch negative Probandinnen) für Männer bei 29 (1 falsch negativ und 1 falsch positiver Proband).

Die klinischen Skalen des MMPI dienen im amerikanischen Raum nur unter speziellen Gesichtspunkten der Simulationsdiagnostik. In der Interpretationshilfe zum Handbuch der Deutschen Kurzfassung des MMPI finden sie im Hinblick auf Simulation keine Beachtung. Wie in allen amerikanischen Studien fand sich jedoch auch in unserer Untersuchung unter Simulation ein Anstieg auf allen 10 klinischen Skalen mit stärkster Ausprägung auf den Skalen Sc/Schizophrenia und Pa/Paranoia. Auffällig war jedoch, dass die größte Differenz zwischen gültiger Selbstbeschreibung und Simulation mit 37,53 T-Wert-Punkten auf Pa zu beobachten war, während in amerikanischen Studien der stärkste Anstieg stets für Sc beschrieben wird. Bei der Aufschlüsselung nach Geschlechtern wurde deutlich, dass diese Verschiebung durch die Werte der weiblichen Probanden hervorgerufen wird. Eine Erklärung für dieses auffällige Antwortverhalten ließe sich möglicherweise durch nachträgliches Explorieren der Probandinnen zu den Fragen der relevanten Skalen erreichen. Dies war aufgrund des anonymisierten Studiendesigns in unserer Untersuchung nicht möglich.

Zusammenfassend führten im Rahmen der experimentellen Studie absichtliche Verfälschungstendenzen zu charakteristischen Veränderungen auf den Validitätsskalen des MMPI. Insbesondere auf der F-Skala und bei der F-K-Differenz fanden sich unter Simulation deutlich höhere Werte. Dieser Befund bestätigt die Ergebnisse amerikanischer Grundlagenstudien zu dieser Fragestellung und ergänzt die Interpretationsvorschläge im Handbuch zur Deutschen Kurzform des MMPI.

6. Literatur

- (1) Angleitner A. Testrezension zu Minnesota Multiphasic Personality Inventory (MMPI). *Z Diff Diag Psychol* 1997; 18: 4-10
- (2) Bagby RM, Rogers R, Buis T, Kalemka V. Malingered and defensive response styles on the MMPI-2: an examination on validity scales. *Assess* 1994; 1: 31-38
- (3) Bagby MR, Rogers R, Buis T, Nicholson RA, Cameron SL, Rector NA. Detecting feigned depression and schizophrenia on the MMPI-2. *J Personal Assess* 1997; 68: 650-666
- (4) Berger M (Hrsg). *Psychiatrie und Psychotherapie*. München, Wien: Urban und Fischer 1999
- (5) Berry DT, Baer RA, Harris MJ. Detection of malingering on the MMPI: a metaanalysis. *Clin Psychol Rev.* 1991; 11 : 585-598
- (6) Bleuler E. *Lehrbuch der Psychiatrie*. Berlin: Springer 1918
- (7) Burkhard B, Christian W, Gynther M. Item subtlety and faking on the MMPI. *J Psychol Assess* 1978; 42: 76-80
- (8) Butcher JN, Hostetler K. Abbreviating MMPI item administration: what can be learned from the MMPI for the MMPI-2. *Psychol Assess* 1990; 2: 12-21
- (9) Cassisi JE, Workman DE. The detection of malingering and deception with a short form of the MMPI-2 based on the L, F, and K scales. *J Clin Psychol* 1992; 48: 54-58
- (10) Cofer CN, Chance J, Judson AJ. A study of malingering on the minnesota multiphasic personality inventory. *J Psychol* 1949; 27: 491-499
- (11) Colby F. Usefulness of the k correction in MMPI profiles of patients and nonpatients. *Psychol Assess* 1989; 1: 142-145
- (12) Eckhardt A. *Selbstmanipulierte Krankheiten*. In: Faust V. (Hrsg): *Psychiatrie*. Stuttgart, Jena, New York: Gustav Fischer 1995
- (13) Ekman P, O'Sullivan M. Who can catch a liar. *Am Psychol* 1991; 46: 913-920
- (14) Elhai JD, Gold SN, Sellers AH, Dorfman WI. The detection of malingered posttraumatic stress disorder with MMPI-2 fake bad indices. *Assess* 2001; 8: 221-236
- (15) Engel R. Replik zu Angleitners Rezension des MMPI. *Z Diff Diag Psychol* 1997; 18: 10-15
- (16) Exner JE, McDowell E., Pabst J, Stackmann W, Kirk L. On the detection of willful falsification in the MMPI. *J consult Psychol* 1963; 27 : 91-94
- (17) Faust D, Ziskin J. The expert witness in psychology and psychiatry. *Sc* 1988; 241: 31-35

- (18) Fox DD, Gerson A, Lees-Haley PR. Interrelationship of MMPI-2 validity scales in personal injury claims. *J Clin Psychol* 1995; 51: 42-47
- (19) Gaebel W, Müller-Spahn F. Lehrbuch der Diagnostik und Therapie psychischer Störungen. Stuttgart, Köln, Berlin: Kohlhammer 2002
- (20) Gavin H. On feigned and factitious diseases chiefly of sldiers and seamen. London: J. Churchill 1843
- (21) Gering A, Blaser A. MMPI; Deutsche Kurzform für Handauswertung; Handbuch. Bern, Stuttgart, Wien: H. Huber 1982
- (22) Gough HG. Diagnostic patterns on the MMPI. *J Clin Psychol* 1946; 2: 23-37
- (23) Gough HG. Simulated Patterns on the MMPI. *J Abnorm Soc Psychol* 1947; 42: 215-225
- (24) Gough HG. The F minus K dissimulation index for the MMPI. *J Consult Psychol* 1950; 14: 408-413
- (25) Gough HG. Some common misperception about neuroticism. *J Consult Psychol* 1954; 18: 287-292
- (26) Graham JR, Watts D, Timbrook RE. Detecting fake-good and fake-bad MMPI-profiles. *J Personal Assess* 1991; 57: 265-277
- (27) Grayson HM, Olinger LB. Simulation of "normalcy" by psychiatric patients on the MMPI. *J Consult Psychol* 1957; 21: 73-77
- (28) Greene RL. Assessment of malingering and defensiveness by multiscale personality inventories. In: Rogers R (Ed.). *Clinical assessment of malingering and deception*. 2nd ed. New York: Guilford, 1997: 169-207
- (29) Grow R, McVough W, Eno TD. Faking and the MMPI. *J Clin Psychol* 1980; 36: 910-917
- (30) Gynther MD. The clinical utility of invalid MMPI F-scores. *J Consult Psychol* 1961; 25: 540-542
- (31) Hawk GL, Cornell DG. MMPI profiles of malingerers diagnosed in pretrial forensic evaluations. *J clin Psychol* 1989; 45: 673-678
- (32) Heffner M. Mehrfachwahl-Wortschatz-Intelligenztest : MWT-B.
www.prometheus.uni-tuebingen.de
- (33) Helmes E, Reddon JR. A perspective on development in assessing psychopathology: A critical review on MMPI and MMPI-2. *Psychological Bulletin*. 1993; 113: 453-471
- (34) Heubrock D. Neuropsychologische Diagnostik bei Simulationsverdacht- ein Überblick über Forschungsergebnisse und Untersuchungsmethoden. *Diagn* 1995; 41: 303-321

- (35) Hollingworth HL. Judging human character. New York: D. Appleton 1922
- (36) Huber G. Psychiatrie. Stuttgart: Schattauer, 1999: 492-495
- (37) Hunt HF. The effect of deliberate deception on MMPI performance. J Consult Psychol 1948; 3: 396-402
- (38) Iverson GL, Franzen MD. Examination of inmates' ability to malingering on the MMPI-2. Psychol Assess 1995; 7: 118-12
- (39) Lanyon RI. Simulation of normal and psychopathic MMPI personality patterns. J Consult Psychol 1967; 31: 94-97
- (40) Lees-Haley PR, Fox DD. MMPI subtle-obvious scales and malingering: clinical versus simulated scores. Psychol Rep 1990; 66: 907-911
- (41) Lucio E, Duran C, Graham JR, Ben Porath YS. Identifying faking bad on the MMPI-A with mexican adolescents. Assess 2002; 9: 62-69
- (42) Mayers Konversationslexikon 6. Auflage; Leipzig, Wien 1909
- (43) Mc Kinley JC, Hathaway SR, Meehl PE. The MMPI: VI. The K scale. J consult psychol 1948; 12: 20-31
- (44) Meehl PE, Hathaway SR. The K factor as a suprasor variable in the MMPI. J. Appl. Psychol 1946; 30: 525-564
- (45) Merten T. Fragen der neuropsychologischen Diagnostik bei Simulationsverdacht. Fortschr. Neurol. Psychiat 2002; 70: 126-138
- (46) Merten T. Über Simulation, artefizielle und somatoforme Störungen- eine konzeptionelle Verwirrung. Z Klin Psychol Psychiat Psychother 2001; 4: 21-39
- (47) Mittenecker E. Subjektive Tests zur Messung der Persönlichkeit. In: Enzyklopädie der Psychologie, Themenbereich B, Serie II, Band 3; Göttingen: Verlag für Psychologie, 1982: 57-91
- (48) Möller HJ, Laux G, Kapfhammer HP (Hrsg). Psychiatrie und Psychotherapie. 2. Aufl. Berlin, Heidelberg, New York: Springer 2005
- (49) Peterson GW, Clark DA, Bennet B. The utility of MMPI subtle, obvious scales for detecting fake good and fake bad response sets. J Clin Psychol 1989; 45: 575-583
- (50) Pollock PH, Quigley B, Worley KO, Bashford C. Feigned mental disorder in prisoners referred to forensic mental health sevices. Journal of psychiatric and mental health nursing 1997; 4: 9-15
- (51) Rasch W. Forensische Psychiatrie. 2. Aufl. Stuttgart, Berlin, Köln: Kohlhammer 1999: 338-343
- (52) Resnick PJ. Malingered psychosis. In Rogers R (Ed). Clinical assessment of malingering and deception. 2nd ed. New York: Guilford 1997: 47-67

- (53) Rogers R. Development of a new classificatory model of malingering. *Bulletin of the American Academy of Psychiatry and Law* 1990; 18: 323-333
- (54) Rogers R. Current status of clinical methods. In Rogers R (Ed). *Clinical assessment of malingering and deception*. 2nd ed. New York: Guilford, 1997: 373-397
- (55) Rogers R, Bagby RM, Chakraborty D. Feigning schizophrenic disorders on the MMPI-2: Detection of coached simulators. *J Personal Assess* 1993; 60: 215-226
- (56) Rogers R, Sewell KW, Salekin RT. A meta-analysis of malingering on the MMPI-2. *Assess* 1994; 1: 227-237
- (57) Rosenhan DL. On being sane in insane places. *Sc* 1973; 179: 250-258
- (58) Rothke SE, Dahlstrom WG, Greene RL, Arredondo R, Mann AW. MMPI-2 normative data for the F-K-Index: Implications for clinical, neuropsychological, and forensic practice. *Assess* 1994; 1: 1-15
- (59) Ruch FL. A technique for detecting attempts to fake performance on a self-inventory type of personality test. In: McNemar Q, Merrill MA. *Studies in personality*. New York: McGraw-Hill, 1942: 229-234
- (60) Schmidt LR. *Lehrbuch der klinischen Psychologie*. Stuttgart: Enke 1984: 71-72
- (61) Schorr A. Stand und Perspektiven diagnostischer Verfahren in der Praxis. *Diagn* 1995; 41: 3-20
- (62) Sewell KW, Rogers R. Response consistency and the MMPI-2: Development of a simplified screening scale. *Assess* 1994; 1: 293-299
- (63) Schretlen DJ. The use of psychological tests to identify malingered symptoms of mental disorder. *Clin Psychol Rev* 1988; 8: 451-476
- (64) Shores EA, Carstairs JR. Accuracy of the MMPI-2 computerized minnesota report in identifying fake-good and fake-bad response sets. *Clinical Neuropsychol* 1998; 12: 101-106
- (65) Steffan JS, Clopton JR, Morgan RD. An MMPI-2 scale to detect malingered depression. *Assess* 2003; 10: 282-292
- (66) Timbrook RE, Graham JR, Keiller SW, Watts D. Comparison of the Wiener-Harmon subtle-obvious scales and the standart validity scales in detecting valid and invalid MMPI-2 profiles. *Psychol Assess* 1993; 5 : 53-61
- (67) Tölle R, Windgassen K. *Psychiatrie*. 14. Aufl. Berlin, Heidelberg, New York. Springer 2006
- (68) Turner M. Malingering. *Br J Psychiat* 1997; 171: 409-411
- (69) Venzlaff U, Förster K. *Psychiatrische Begutachtung*. 3. Aufl. München, Wien: Urban und Fischer 2000: 88-89

- (70) Walters GD, White TW, Greene R. Use of the MMPI to identify malingering and exaggeration of psychiatric symptomatology in male prison inmates. *J Consult Clin Psychol* 1988; 56: 111-117
- (71) Wasylw OE, Grossmann LS, Haywood TW, Cavanaugh JL. The detection of malingering in criminal forensic groups: MMPI validity scales. *J Personal Assess* 1988; 52: 321-333
- (72) Wetter MW, Baer RA, Berry DTR, Smith GT, Larsen LH. Sensitivity of MMPI-2 validity scales to random responding and malingering. *Psychol Assess* 1992; 4: 369-374
- (73) Wetter MW, Baer RA, Berry DTR, Robinson LH, Sumpter J. MMPI-2 profiles of motivated fakers given specific symptom information: A comparison to matched patients. *Psychol Assess* 1993; 5: 317-323
- (74) Wiener DN. Subtle and obvious keys for the MMPI. *J Consult Psychol* 1948; 12: 164-170

LEBENS LAUF

Luise Heitmann von Franqué
geboren am 22.11.1962 in Bonn
evangelisch
ledig, 1 Tochter

Schul Ausbildung

1969-1973 Städtische Grundschule Wetzlar
1973-1982 Gesamtschule und Gymnasiale Oberstufe in Wetzlar

Hochschulstudium

1983-1984 Studium der Philosophie und Sprachwissenschaft
an der Universität Heidelberg
1985-1993 Studium der Humanmedizin an der
Heinrich-Heine Universität Düsseldorf

Beruf

01.08.1993 - 14.11.1996 ÄiP und anschl. Assistenzärztin in der Ev. Stiftung
Tannenhof, Remscheid, Klinikbereich Psychiatrie
18.11.1996 - 31.10.1998 Mutterschutz und Erziehungsurlaub
01.11.1998 – 31.06.2001 Assistenzärztin in der Ev. Stiftung Tannenhof
davon 1 Jahr im Klinikbereich Neurologie
02.05.2001 Fachärztin für Psychiatrie und Psychotherapie
(Ärztekammer Nordrhein)
01.07.2001 – 01.12.2001 Oberärztin in der Ev. Stiftung Tannenhof
01.01.2002 – 30.08.2003 Stadtärztin beim Sozialpsychiatrischen Dienst der
Stadt Köln
seit dem 01.02.2002 Konsiliarpsychiaterin der Justizvollzugsanstalt Köln
seit dem 01.09.2003 Ärztin in der Tagesklinik Siegburg
Klinik für Psychiatrie und Psychotherapie

Wirkungen experimenteller Simulation auf das MMPI

Luise Heitmann - von Franqué

Zusammenfassung

Anhand einer experimentellen Studie wird gezeigt, wie die Simulation einer psychischen Erkrankung das MMPI-Ergebnis verfälscht.

43 Studentinnen und 27 Studenten des 3. vorklinischen Semesters Humanmedizin wurden im Rahmen eines abhängigen 2-Gruppenplans untersucht. Sie erhielten die Aufgabe, die Deutsche Kurzform des MMPI jeweils zwei mal zu bearbeiten: einmal mit der gültigen Selbstbeschreibung und einmal mit der Vorgabe, eine psychiatrische Erkrankung zu simulieren, die in ihrer Ausprägung Schuldunfähigkeit entspricht. Die Angaben wurden anonymisiert erhoben, lediglich Alter und Geschlecht wurden angegeben. Zudem wurde durch das Vorschalten des MWTB-2 ein ausreichendes Sprachverständnis bei allem Probanden sichergestellt.

Bei der Auswertung der erhobenen Daten mittels SPSS fanden sich für das Gesamtkollektiv auf allen 3 Validitätsskalen des MMPI (L,F,K) signifikante Unterschiede zwischen den Werten bei Simulation und bei gültiger Selbstbeschreibung. Besonders eindrücklich waren die Veränderungen auf der F-Skala mit einem Anstieg von 43,614 bei gültiger Selbstbeschreibung auf 98,6 unter Simulation. Auf der K-Skala war ein Absinken der Mittelwerte von 55,6 bei gültiger Selbstbeschreibung auf 40,643 unter Simulation zu beobachten. Die F-K-Differenz stieg entsprechend von -5,071 auf 57,957.

Unter der Anwendung eines Schwellenwertes von $F > 60$, wie er in der Interpretationshilfe zur Deutschen Kurzform des MMPI vorgeschlagen wird, konnten 98,6% der simulierenden Probanden korrekt identifiziert werden. 8,6% der gültigen Selbstbeschreibungen wurden unter dieser Voraussetzung falsch positiv bewertet. (Gesamtfehlerquote 10%)

Für die F-K-Differenz erwies sich ein Cut-off-Wert von 19 mit 4 falsch negativen und 4 falsch positiven Probanden als günstigste Vorgabe. Dies deckt sich nicht mit der Empfehlung der Interpretationshilfe zur Deutschen Kurzform des MMPI; dort wird ein Cut-off-Wert von 10 vorgeschlagen.

Differenziert nach Geschlechtern zeigte sich unter Simulation sowohl der Anstieg der Werte auf der F-Skala als auch das Absinken der Werte auf der K-Skala bei Frauen stärker ausgeprägt als bei Männern. Zum Simulationsnachweis erwiesen sich unterschiedliche Cut-off-Werte für Männer und Frauen als sinnvoll. (Frauen: $F > 60$, $F - K > 18$. Männer: $F > 73$, $F - K > 29$)

Auf allen klinischen Skalen des MMPI lagen die Werte unter Simulation über den Werten bei gültiger Selbstbeschreibung. Am deutlichsten ausgeprägt war die Veränderung auf den Skalen Pa (Paranoia) und Sc (Schizophrenie).

Die Ergebnisse der Untersuchung entsprechen im Wesentlichen den Ergebnissen amerikanischer Grundlagenstudien zur Simulationsdiagnostik mit dem MMPI. Sie ergänzen die Empfehlungen in der Interpretationshilfe zur Deutschen Kurzform des MMPI, für die bisher keine Belege vorlagen.



Prof. Dr. Erlo Lehmann