



# Ein Paradigmenwechsel in der Neuropsychologie

Thomas Merten und Robbi Brockhaus

## LITERATUR:

**Gorissen, M., Sanz de la Torre, J.C. & Schmand, B.** (2003).

Effort in the neuropsychological evaluation of schizophrenia (ENES study) [Abstract]. *Journal of the International Neuropsychological Society*, 9, 512.

Green, P. (2003a). Welcoming a paradigm shift in neuropsychology. *Archives of Clinical Neuropsychology*, 18, 625-627.

Green, P. (2003b). Green's Word Memory Test for Windows: User's Manual. Edmonton, Canada: Green's Publishing.

Hartman, D.E. (2002). The unexamined lie is a lie worth fibbing. *Neuropsychological malingering and the Word Memory Test*. *Archives of Clinical Neuropsychology*, 17, 709-714.

Rohling, M.L., Green, P., Allen, L.M. & Iverson, G.L. (2002). Depressive symptoms and neurocognitive test scores in patients passing symptom validity tests. *Archives of Clinical Neuropsychology*, 17, 205-222.

Westhoff, K., Hornke, L.F. & Westmeyer, H. (2003). Richtlinien für den diagnostischen Prozess. Zur Diskussion gestellt. *Report Psychologie*, 28, 504-517.

»Welcoming a paradigm shift in neuropsychology«, so nannte Green (2003a) einen Artikel, der jüngst in den *Archives of Clinical Neuropsychology* erschien. Was sich hinter diesem Titel verbirgt, kann der Leser des *Report Psychologie* auch aus dem in dieser Ausgabe abgedruckten Aufsatz erfahren, den der Autor selbst zur Veröffentlichung in Deutschland vorschlug. Dahinter stand der Wunsch, dass eine wesentliche Entwicklungsrichtung, die die klinische Neuropsychologie in den vergangenen 10 bis 15 Jahren eingeschlagen hat, deutschsprachigen Psychologen leichter zugänglich würde.

Für viele, wenn nicht die meisten hier zu Lande tätigen klinischleistungsdiagnostisch arbeitenden Psychologen ist bislang eine routinemäßige Erfassung (Messung) der Testmotivation weit von der Realität entfernt. Damit laufen wir Gefahr, nicht nur individuelle Testwerte fälschlicherweise dort als valide zu betrachten, wo sie es nicht sind, sondern auch Forschung zu betreiben, die ohne Kontrolle dieser wichtigen Variablen zu irreführenden Ergebnissen kommen kann. Solche Forschungsergebnisse können ihrerseits in ungünstiger Weise die psychologische Theorienbildung beeinflussen. Auf der zurückliegenden Tagung der International Neuropsychological Society (INS) im Sommer 2003 in Berlin gab es beispielsweise Gelegenheit, von holländischen und spanischen Kollegen etwas über die Bedeutung der Testmotivation, wie sie mit dem Word Memory Test von Green (2003b) erfasst wird, für die Diagnostik kognitiver Störungen bei schizophren Erkrankten zu erfahren (Gorissen, Sanz de la Torre & Schmand, 2003).

Wie grundlegend die Arbeiten und das eindrucksvolle Engagement Paul Greens und seiner Kollegen auf die neuropsychologische Praxis und Forschung in Nordamerika in den zurückliegenden Jahren wirkten, wurde im Oktober 2003 mit der Verleihung des renommierten Nelson-Butters-Preises auf der Jahrestagung der National Academy of Neuropsychology (NAN) ge-

würdigt, und zwar für eine bemerkenswerte Arbeit zum Zusammenhang zwischen depressiven Symptomen und dem Abschneiden in psychologischen Testvariablen (Rohling, Green, Allen & Iverson, 2002).

Diese letztgenannte Arbeit macht darauf aufmerksam, dass die Untersuchung der Testmotivation weit über traditionelle Fragestellungen der Neuropsychologie hinausgeht. Inwieweit auch andere Disziplinen der angewandten Psychologie und der grundlagenorientierten Forschung von dem Paradigmenwechsel in der Neuropsychologie und der Entwicklung eines Methodenrepertoires für die Motivationserfassung profitieren, wird eine spannende Frage für die kommenden Jahre darstellen. In Zukunft immer weniger zu akzeptieren sind auf reinen Eindrucksurteilen beruhende Validitätsannahmen für die erhobenen Testdaten. Damit wird nicht zuletzt auch die Bedeutung einer an explizit formulierbaren Richtlinien für den diagnostischen Prozess orientierten Arbeit gestärkt werden. In einem jüngst im *Report Psychologie* zur Diskussion gestellten Grundsatzartikel heißt es dazu: »34. Der Diagnostiker prüft, ob alle Daten brauchbar und frei von ... Verzerrungen sind. 35. Der Diagnostiker bewertet die Qualität der Daten im Hinblick auf die diagnostischen Fragen. 36. Der Diagnostiker analysiert und interpretiert Daten aus Tests und anderen diagnostischen Verfahren nach dem neuesten Stand der Normen, Standards und des professionellen Wissens« (Westhoff, Hornke & Westmeyer, 2003, S. 512). Besser könnte das Feld, in das das Thema der im Folgenden abgedruckten Arbeit gehört, kaum abgesteckt werden.

Für alle weitergehend Interessierten sei auf eine vergleichende Betrachtung des Word Memory Tests durch Hartman (2002) sowie auf die Webseite [www.wordmemorytest.com](http://www.wordmemorytest.com) des Autors selbst hingewiesen. Die Übersetzer möchten schließlich Dr. Paul Green für diese Gelegenheit, seinen Artikel hier vorstellen zu können, nachdrücklich danken.

Berlin und Krefeld, November 2003



# Testmotivation und ihre Messung

**»Der Patient war anscheinend motiviert...,  
doch die Testergebnisse sind ungültig«**

Symptomvalidierungstestung in der Neuropsychologie und der Word Memory Test<sup>1</sup>

**PAUL GREEN, Edmonton, Canada**

Deutsche Übersetzung von

**THOMAS MERTEN (Berlin)**

**ROBBI BROCKHAUS (Krefeld)**

## **Warum beschäftigen wir uns mit Testmotivation?**

Wenn wir einen Befund oder ein Gutachten abfassen, verwenden wir alle, wie wir es gelernt haben, solche Formulierungen wie: »Der Patient (oder das Kind) hat allem Anschein nach gut motiviert die Aufgaben bearbeitet, so dass kein Zweifel an der Gültigkeit der Ergebnisse besteht.« Wenn wir eine solche Feststellung tref-

fen, anerkennen wir implizit die Möglichkeit, dass auch das Gegenteil wahr sein könnte. Wie oft aber haben wir allein auf der Grundlage unserer klinischen Beobachtungen zu Papier gebracht: »Die Ergebnisse sind als ungültig zu betrachten«? Die meisten von uns können derartige Fälle an den Fingern einer Hand abzählen, weil wir beharrlich von der Annahme ausgehen, dass eine ausreichende Leistungsmotivation (effort) vorliegt

<sup>1</sup> Modifizierte und erweiterte Fassung eines Aufsatzes, der erstmals im August 2001 im Newsletter »Psymposium« der Alberta Psychological Association veröffentlicht wurde. Kurz zuvor war dem Autor der Dick-Pettifor-Gedächtnispreis für seinen herausragenden Beitrag zur Entwicklung der Psychologie in Alberta verliehen worden.

## Autor

Dr. PAUL GREEN,  
BSc in Psychologie an  
der Univ. Wales,  
MSc und PhD in  
Psychologie an der  
Univ. Birmingham,  
England.  
Seit 27 Jahren Tätigkeit  
als Klinischer  
Neuropsychologe in  
England und Kanada.  
Zahlreiche  
Veröffentlichungen zu  
so mannigfaltigen  
Themen wie  
Interhemisphärentrans-  
fer bei Schizophrenen,  
Sprachverständnis bei  
lernbehinderten Kin-  
dern und Einfluss von  
zerebralen Verletzun-  
gen auf den Geruchs-  
sinn. Erlangte große  
Bekanntheit durch die  
Entwicklung des Word  
Memory Test (WMT),  
der für verschiedene  
Sprachen,  
einschließlich der  
deutschen und  
der türkischen,  
adaptiert wurde.

Dr. THOMAS MERTEN,  
Klinikum im  
Friedrichshain,  
Klinik für Neurologie,  
Landsberger Allee 49,  
10249 Berlin  
T 030 – 4221-1663  
E thomas.merten@  
vivantes.de

und die Testergebnisse gültig sind. Dagegen hat uns die Leistungsmotivationsforschung in den zurückliegenden 15 Jahren gezeigt, dass Personen, die Leistungstests bearbeiten, recht häufig keine ausreichende Anstrengungsbereitschaft aufbringen, gültige Testergebnisse zu produzieren. Ohne spezialisierte Messinstrumente sind wir jedoch bedauerlicherweise nicht in der Lage, ungültige Ergebnisse als solche zu erkennen.

Liest man in einer Fachzeitschrift zum ersten Mal davon, dass sich eine ungenügende Motivation ungünstig auf die Testergebnisse von Probanden auswirkt, so wird wohl die natürliche Reaktion sein, zwar die grundsätzliche Richtigkeit dieser Aussage zu akzeptieren, gleichzeitig jedoch davon überzeugt zu sein, dass die Testdaten, die man selbst bei Patienten erhoben hat, davon ausgenommen sind. Man könnte glauben, die Aussage treffe höchstens auf die Daten zu, die andere Kollegen ermitteln. Ungenügende Anstrengungsbereitschaft – so etwas könnte es schon einmal bei ein paar Erwachsenen geben, die sich gerade im Rahmen eines Berentungsverfahrens einer Gedächtnisprüfung unterziehen müssen. Wir würden keineswegs erwarten, dass es sich um ein weit verbreitetes Phänomen handelt. Bei Kindern beispielsweise, die ja keinerlei finanzielle Beweggründe für eine Aggravation kognitiver Störungen haben können, würde man einen solchen Einflussfaktor nicht erwarten.

Aber das Phänomen, von dem wir sprechen, ist weit verbreitet. Eingeschränkte Leistungsmotivation kann bei einer Reihe von Kindern beobachtet werden. Leistungsmotivation erweist sich als eine der stärksten Variablen überhaupt, die Testergebnisse beeinflussen.

Testmotivation als Gruppenmerkmal wurde zuerst an Personen untersucht, die eine leichte Schädel-Hirn-Verletzung erlitten hatten. Wissenschaftler wie Binder (z.B. 1993) konnten zeigen, dass paradoxerweise eine Gruppe von Patienten nach minimalem Schädel-Hirn-Trauma gegenüber solchen mit schweren zerebralen Verletzungen signifikant schlechtere Ergebnisse in einfachen Motivationsmaßen erreichte. Zu einem ähnlichen Resultat kamen Green, Iverson und Allen (1999) unter Verwendung des Word Memory Tests (WMT), der weiter unten beschrieben wird. Biologisch macht es keinen Sinn, dass die Patienten mit den schwersten Hirnverletzungen am besten abschneiden. Die Erklärung für diese völlig absurd erscheinenden Ergebnisse liegt vielmehr darin, dass eine Anzahl von Patienten nach leichten Verletzungen dazu neigten, ihre kognitiven Störungen zu übertreiben, während diejenigen mit schweren Hirnverletzungen für solche Übertreibungen weniger anfällig waren. Autoren wie Larrabee (2000) schätzen die Prävalenz der Beschwerdenaggravation bei Personen nach leichtem Schädel-Hirn-Trauma, die sich im Rahmen einer Begutachtung einer neuropsychologischen Diagnostik unterziehen, gegenwärtig auf etwa 30 bis 40 Prozent. Das bedeutet, dass mitgeteilte neuropsychologische Testbefunde in ca. 35 Prozent der Fälle von Folgezuständen leichter Schädel-Hirn-Verletzungen wahrscheinlich als ungültig anzusehen sind. Eben hieraus resultiert die große Aufmerksamkeit, die dieses Thema verdient.

## Symptomvalidierungstests: CARB und WMT

Als vollzeitbeschäftigter Klinischer Neuropsychologe in eigener Praxis und ohne jegliche Förderung durch Forschungsgelder begann ich im Jahre 1992 mit dem routinemäßigen Einsatz von Instrumenten zur Messung der Leistungsmotivation (die einer gängigen Sprachregelung folgend auch Symptomvalidierungstests, SVT, genannt werden). Diese Leistungsmotivationstestung wurde bei jeder vorgestellten Person unabhängig von ihrer Diagnose durchgeführt. Der von mir zuerst benutzte Symptomvalidierungstest war das CARB (Computerized Assessment of Response Bias = Computerbasierte Diagnostik von Negativen Antwortverzerrungen) von Conder, Allen und Cox (später Conder, Allen, Green & Cox, 1997). Diese Aufgabe ist objektiv außerordentlich einfach, weshalb Personen, die eine schwere Schädel-Hirn-Verletzung erlitten, rund 98 Prozent Richtige erreichten. Viele Patienten nach minimalem Schädel-Hirn-Trauma erzielten dagegen in diesem Test Werte weit unterhalb des Mittelwerts der Schwerverletzten, was die Möglichkeit von mangelhafter Leistungsmotivation und folglich ungültigen Testergebnissen nahe legt (Green & Iverson, 2001a). Viele Patienten mit anderen Diagnosen wie Major Depression, chronischem Müdigkeitssyndrom oder posttraumatischer Belastungsstörung verhielten sich aber im Test genauso. Dabei sei angemerkt, dass das subjektiv geschilderte Ausmaß der Gedächtnisstörungen, gemessen mithilfe eines standardisierten Fragebogens, in all diesen Patientengruppen größer war als bei Patienten nach unzweifelhaftem schwerem Schädel-Hirn-Trauma oder bei Patienten mit einer eindeutig nachweisbaren, schweren zerebralen Erkrankung. In meiner Untersuchungsreihe von 1207 aufeinander folgenden Fällen machten 97 Prozent aller nicht-hirngeschädigten, nicht-neurologischen Patienten geltend, sie litten an Gedächtnisstörungen, die in ungünstiger Weise ihre Arbeitsfähigkeit herabsetzten.

Um 1994 herum schuf ich einen Gedächtnistest namens Word Memory Test (Green & Astner, 1995), der Motivationsmaße einschloss. Dieser Test, der später in einer ersten, unter DOS laufenden Version programmiert wurde (Green, Allen & Astner, 1996), wird inzwischen auf der ganzen Welt von hunderten von Neuropsychologen benutzt, zum Beispiel in Kanada, den Vereinigten Staaten, Neuseeland, Australien, Großbritannien, den Niederlanden, Spanien, Deutschland, Argentinien oder Litauen. Im WMT werden 20 Wortpaare entweder mündlich oder auf dem Bildschirm dargeboten. Jedes Paar regt zwei verschiedene, miteinander assoziativ verknüpfte alltagsnahe Vorstellungen an (z.B. Tisch – Stuhl oder Tasse – Kaffee). Nach zweimaliger Darbietung der Wortliste soll der Proband unter jeweils zwei möglichen Antworten (z.B. Tasse – Henkel) diejenige herausuchen, die zur Lernliste gehörte. Eine ganz ähnliche Aufgabe wird nach Ablauf einer halben Stunde gestellt, gefolgt von einer Reihe von Gedächtnistests.

## Unterschiedliches Abschneiden verschiedener Diagnosegruppen in SVT

Personen, die zweifelsfrei an einer schweren Hirnverletzung bzw. einer neurologischen Erkrankung wie Hirntumor oder Aneurysmablutung litten, konnten ohne Mü-

he die beschriebene Wiedererkennungsaufgabe erfüllen und erreichten einen mittleren Testwert von rund 96 Prozent Richtigen (Green, Lees-Haley & Allen, 2002). Die Leistungsmotivation eines Probanden muss dann ernsthaft infrage gestellt werden, wenn seine erreichten Punkte unterhalb eines Trennwertes (cutoff) fallen. Dieser Trennwert liegt etwa drei Standardabweichungen unterhalb des Mittelwertes jener Personen, die schwere Hirnverletzungen oder neurologische Erkrankungen aufwiesen. Wenn man den Intelligenzquotienten zum Vergleich heranzieht, würde dies einem IQ von weniger als 55 entsprechen, wobei allerdings zu berücksichtigen ist, dass die WMT-Werte nicht normalverteilt sind. Ungültige Testergebnisse sind besonders dann wahrscheinlich, wenn ein Proband, der den Symptomvalidierungstest nicht bestanden hat, eine Diagnose wie etwa leichtes Schädel-Hirn-Trauma, Depression, Fibromyalgie oder chronisches Müdigkeitssyndrom hat. Auch andere Diagnosen, zu denen üblicherweise keine ausgesprochen schweren kognitiven Störungen gehören, sondern bei denen im Allgemeinen überhaupt keine kognitiven Defizite in Leistungstests erscheinen, kommen hier infrage.

Man könnte natürlich dagegenhalten, dass ein Mensch unter Schmerzen einfach nicht in der Lage ist, sich während einer Testung ausreichend zu konzentrieren – und wenn die Aufgaben noch so leicht sein mögen. Diese Hypothese wurde jedoch in einer im Journal of Forensic Neuropsychology veröffentlichten Studie strikt zurückgewiesen. Die Autoren (Gervais, Green, Allen & Iverson, 2001) konnten vielmehr zeigen, dass 42 Prozent aller Patienten mit chronischen Schmerzsyndromen im CARB nicht bestanden, wenn dieser Test in seiner standardisierten Form durchgeführt wurde. Wenn dagegen vergleichbaren Patienten mitgeteilt wurde, dass der Test lediglich die Leistungsmotivation misst, fielen nur noch 6 Prozent durch (Gervais et al., 2001). In ähnlicher Weise waren 35 Prozent der Patienten mit Fibromyalgie, die eine Berentung anstrebten, im CARB und/oder im WMT auffällig. Unter den Patienten mit Fibromyalgie, die sich nicht in einem Rentenverfahren befanden, lag der Anteil indessen nur bei 4 Prozent (Gervais, Russell, Green, Ferrari, Allen & Pieschl, 2001). Die letztgenannte Studie veranschaulicht, dass sich Misserfolg in Symptomvalidierungstests nicht auf bestimmte diagnostische Gruppen beschränken lässt, sondern dass es vielmehr eine charakteristische Einstellung gegenüber den Leistungstests selbst ist, die sich hier niederschlägt. Ausgenommen sind natürlich einige wenige sehr schwere Erkrankungen wie etwa fortgeschrittene Demenzen.

Nachdem mehr als eintausend unausgelesene Patienten das CARB und den WMT im Rahmen einer jeweils zweitägigen neuropsychologischen Diagnostik in meiner Praxis bearbeitet hatten, konnten wir sehr viel mehr über Testmotivation und die Bedeutung neuropsychologischer Testergebnisse erfahren. Als eines der wichtigsten Ergebnisse zeigte sich, dass Leistungsmotivation, wie sie durch den WMT gemessen wird, in einer Untersuchung an 904 Gutachtenpatienten 50 Prozent der Varianz aller anderen Testergebnisse erklärte

(Green, Rohling, Lees-Haley & Allen, 2001). Um dieses Resultat in den richtigen Kontext zu setzen, sei erwähnt, dass das Bildungsniveau (Ausbildungsjahre) nur 11 Prozent der Varianz und das Lebensalter lediglich 4 Prozent der Varianz in diesem Datensatz aufklärten. Das bedeutet also, dass für die hier betrachtete Patientenstichprobe die während der Untersuchung entfaltete Testmotivation mit großem Abstand die wichtigste Einzelvariable darstellte, die Einfluss auf die anderen Testergebnisse ausübte. In dieser Stichprobe, die für die Klientel privater, mit der Begutachtung von Invaliditäts- und Entschädigungsfällen befasster Psychologenpraxen in Nordamerika als repräsentativ gelten kann, fielen rund 28 Prozent der Untersuchten in Symptomvalidierungstests durch, und wir können zweifelsfrei belegen, dass diese Patienten in den eigentlichen Leistungstests ungültige Ergebnisse produzierten.

### Die Bedeutung auffälliger Werte in einem SVT

Was bedeutet es aber, wenn jemand einen Symptomvalidierungstest nicht besteht? Eine Art, sich der Beantwortung dieser Frage zu nähern, liegt darin, dass man sich eine große Anzahl von Ergebnissen in Leistungstests anschaut, und zwar getrennt für solche Personen, die einen SVT wie den WMT bestehen, und solche, die durchfallen. Dr. Martin Rohling hat ein Maß entwickelt, das er OTBM nannte (overall test battery mean = Gesamtmittel der Testbatterie). Dazu wandelte er jeden einzelnen von insgesamt 43 neuropsychologischen Testwerten in einen z-Wert um, für den der Erwartungswert null wäre und individuelle Ergebnisse als Abweichungseinheiten ober- oder unterhalb dieses Erwartungswertes ausgedrückt werden (Green, Rohling et al., 2001). Wir fanden, dass Personen nach leichten Schädel-Hirn-Verletzungen, die den WMT bestanden, Werte nur 0,1 Standardabweichungen (SD) unterhalb des Erwartungswertes der 43 Testergebnisse erzielten (d.h. sie schnitten fast identisch mit dem Erwartungswert ab). Demgegenüber lagen die Werte derjenigen Personen nach leichtem Schädel-Hirn-Trauma, die auffällige Werte im WMT erreichten, 1,25 Standardabweichungen unter dem OTBM-Erwartungswert. Das bedeutet, dass die Untergruppe der leicht Verletzten, die ihre kognitiven Störungen übertrieben, im Vergleich zu jenen, die keine Motivationsprobleme hatten und valide Ergebnisse lieferten, erheblich weiter unterhalb des Erwartungswertes abschnitt.

Personen mit der Diagnose einer schweren Hirnverletzung oder eines Hirntumors erreichten dagegen im Durchschnitt Werte, die lediglich 0,4 SD unter dem Erwartungswert lagen. Sie hatten also erheblich bessere Testwerte als die Patienten nach leichtem Schädel-Hirn-Trauma, die in den SVT auffällig waren. Die Personen nach leichter zerebraler Verletzung, die die SVT nicht bestanden, schnitten gegenüber Personen mit nachgewiesenen schweren Hirnschädigungen oder Hirntumoren bei den gemittelten 43 Einzeltests deutlich schlechter ab. Diese aufsehenerregenden Ergeb-

nisse, die sich auf Daten von 500 Fällen von Schädel-Hirn-Verletzungen stützten, lieferten gleichsam den Titel des Artikels: »Effort has a greater effect on test scores than severe brain injury in compensation claimants« (Green, Rohling et al., 2001), das heißt, in Begutachtungsfällen wirkt sich die Leistungsmotivation stärker als eine schwere Hirnverletzung auf das Abschneiden in der Testdiagnostik aus! Die Ergebnisse sind vor kurzem in einer Studie an einer sehr großen unabhängigen Stichprobe amerikanischer Patienten mit Schädel-Hirn-Trauma repliziert worden, wenngleich die Daten noch unveröffentlicht sind (Dr. Rohling, persönliche Mitteilung).

Diese Forschungsergebnisse machen uns sehr deutlich, dass wir ohne eine routinemäßige Messung der Leistungsmotivation im Rahmen unserer Diagnostik zu dem Schluss gelangen, dass viele Menschen kognitive Störungen aufweisen – obwohl dies gar nicht der Fall ist. Anders ausgedrückt: wenn wir fälschlicherweise ungültige Testdaten als gültig betrachten, überschätzen wir die Schwere der kognitiven Störungen der entsprechenden Person. In vielen Fällen liegt der Grund für das schlechte Abschneiden einer Person in einem Leistungstest darin, dass sie nicht ausreichend motiviert war, gültige Testergebnisse zu produzieren. Versicherungsgesellschaften in Nordamerika haben jetzt die Wichtigkeit der Motivationsmessung anerkannt, so dass manche bereits den Einsatz mehr als eines SVT im Rahmen jedes Gutachtens fordern.

### **Leistungsmotivation und die Verwendung des WMT im Kindesalter**

Lehrer schreiben häufig in ein Schulzeugnis »Könnte besser sein!«. Neuere, maßgeblich von Dr. Lloyd Flaro und Dr. John Courtney ausgeführte Forschungsarbeiten zeigen, dass einige Kinder im Rahmen einer neuropsychologischen Untersuchung auch besser abschneiden könnten, wenn sie es nur wollten. In einer Studie mit kanadischen Kindern im Alter von 7 bis 18 Jahren (Green & Flaro, 2003) hatten weder das Lebensalter noch die sprachliche Intelligenz einen signifikanten Einfluss auf die Werte in den Symptomvalidierungsteilen des WMT. Für jede der klinisch unterschiedenen sechs Gruppen von Kindern, darunter solche mit Diagnosen wie fötales Alkoholsyndrom, Schizophrenie des Kindesalters, Aufmerksamkeits-Defizit-Syndrom, neurologische Erkrankungen und Verhaltensstörungen, lagen die Mittelwerte der Symptomvalidierungsteile des WMT bei nahezu 95 Prozent Richtigen, selbst unter Einschluss jener Fälle, in denen der WMT nicht bestanden wurde. Die Gesamtgruppe der Kinder erreichte ebenso hohe Werte wie eine Gruppe von Erwachsenen, die im Rahmen eines Verfahrens zur Ausübung des Sorgerechts für ihre Kinder untersucht wurden. Unter den Kindern, die ohne Probleme die Symptomvalidierungsteile des WMT bestanden, war zum Beispiel eine 17-jährige Jugendliche mit einem verbalen IQ von 52. In den beiden entscheidenden Symptomvalidierungs-Untertests des WMT gab sie 100 Prozent richtige Antworten.

Obwohl die meisten Kinder in der genannten Studie

insgesamt gute Leistungen erzielten, gab es 14 Prozent unter ihnen, für die eine unzureichende Leistungsmotivation ermittelt wurde. Es konnte für diese Kinder gezeigt werden, dass ihre Ergebnisse in den tatsächlichen Gedächtnisleistungstests aufgrund mangelnder Leistungsanstrengung zu niedrig ausfielen. Dass diese Leistungseinschränkungen am ehesten als Artefakt zu betrachten sind, ist durch die Kinder auf Befragen hin freimütig eingestanden worden. Sie gaben im Übrigen ganz kindgemäße Motive dafür an (Green & Flaro, 2003). Dass solche Kinder durchaus befähigt sind, bei den Symptomvalidierungsmaßen des WMT gut abzuschneiden, konnte Dr. Flaro im Rahmen einer experimentellen Studie zeigen, die mit Erlaubnis der Eltern durchgeführt wurde. Er ließ all jene Kinder, die zunächst ein unzureichendes Ergebnis im CARB oder im WMT gezeigt hatten, diese Tests am gleichen Tage noch einmal machen. Dieses Mal wurde eine kleine Belohnung ausgesetzt. Mit einer Ausnahme bestanden alle! Damit konnte gezeigt werden, dass die Kinder mühelos in der Lage waren, die Aufgaben gut auszuführen, dies aber einfach beim ersten Mal gar nicht erst versucht hatten. Der einzige Junge, der bei der zweiten Durchführung erneut schlecht abschnitt, bewies in anderer Weise, dass auch er die Aufgabe prinzipiell hätte erfüllen können. Er erzielte ein Ergebnis im WMT, das signifikant unterhalb des Zufallsniveaus lag. Dies gelingt jedoch mit hoher Wahrscheinlichkeit nur bei Kenntnis der richtigen Antworten und gezielter Wahl der jeweils falschen Wörter. Die klinische Diagnose des Jungen lautete übrigens oppositionelle Störung des Sozialverhaltens.

Derartige Forschungsfragestellungen müssen noch ausgeweitet und die bisherigen Ergebnisse repliziert werden, aber es deutet alles darauf hin, dass selbst für die Untersuchung von Kindern, die keinen finanziellen Anreiz zur Symptomausweitung haben, eine Messung ihrer Leistungsmotivation von Bedeutung sein wird. Die Anstrengungsbereitschaft wird in bestimmten Fällen gering ausfallen und folglich werden Testergebnisse, die diese Kinder erreichen, ihre wirklichen Fähigkeiten unterschätzen.

### **Analogstudien**

Auch Analogstudien, die einen wichtigen methodischen Zugang zur Überprüfung der Güte von Symptomvalidierungstests darstellen, belegen die Brauchbarkeit des Word Memory Tests in überzeugender Weise. In einer ersten von Green, Lees-Haley und Allen (2002) berichteten Untersuchung konnten alle 20 Personen, die die Rolle eines Leistungsausfälle vortäuschenden Patienten übernahmen, mithilfe einer einfachen Entscheidungsregel korrekt identifiziert werden. Eine weitere Studie der Autoren wurde mit solchen »experimentellen Simulanten« (wie sie in der Literatur genannt werden) durchgeführt, die beispielsweise als Psychologen und Ärzte erfahren im Umgang mit zerebral geschädigten Patienten waren. 24 der 25 Probanden, die experimentell kognitive Störungen vortäuschten, wurden durch den WMT identifiziert. Der eine verbliebene Fall wies ein auffälliges Muster in den vier

Untertests des WMT auf, die Gedächtnis messen. Wenn also beide Indikatoren, d.h. (1) ein auffälliges Muster in den Symptomvalidierungsteilen des WMT und (2) ein auffälliges Muster in den Gedächtnisteilen des WMT, zur Klassifikation herangezogen werden, so wurde auch für diese Studie eine 100prozentig richtige Identifikation erreicht.

Ähnlich überzeugende Ergebnisse sind durch weitere Analogstudien mit der englischen (Tan, Slick, Strauss & Hultsch, 2002), türkischen (Brockhaus & Peker, 2003) und deutschen (Brockhaus & Merten, im Druck) Version des WMT erzielt worden.

### **Konsequenzen für Gruppenuntersuchungen**

Die Messung der Leistungsmotivation ist jedoch nicht nur für die klinische Einzelfalldiagnostik von Bedeutung. Möglicherweise noch weiter reichende Konsequenzen aus derartigen Forschungsergebnissen sind für Gruppenuntersuchungen zu erwarten. Wenn die Testmotivation nicht überprüft wird, sind wahrscheinlich die an vielen Stichproben erhaltenen Daten durch ungültige Testergebnisse kontaminiert. Dies wiederum kann die aus den Forschungsergebnissen gezogenen Schlussfolgerungen in ungünstiger Weise verzerren. Ein gutes Beispiel für dieses Phänomen ist die Untersuchung des Geruchssinns bei Personen nach Schädel-Hirn-Trauma. Doty et al. (1997) kamen beispielsweise zu dem Schluss, dass es bei Patienten nach leichten Verletzungen ohne Bewusstlosigkeit ebenso viele Störungen der Geruchsdiskrimination gibt wie bei Patienten nach Schädel-Hirn-Verletzungen mit Bewusstlosigkeit von mehr als einem Tag. Dies erscheint nicht nur unlogisch, sondern wir wissen inzwischen, dass es in dieser Verallgemeinerung schlicht falsch ist. Die genannten Autoren hatten es versäumt, eine Messung der Leistungsmotivation in ihre Untersuchung einzuschließen, und konnten deshalb auch aufgrund einer Symptomübertreibung ungültige Daten nicht aus ihrer Analyse ausschließen.

Neuere Untersuchungen konnten dagegen zeigen, dass in Symptomvalidierungstests auffällig abscheidende Personen häufig bei einer Überprüfung des Geruchssinnes ebenso auffällige Werte aufweisen. Dies bedeutet, dass ihre Ergebnisse aufgrund eingeschränkter Testmotivation bzw. Symptomaggravation wahrscheinlich als ungültig anzusehen sind (Green & Iverson, 2001b). Wenn alle Fälle von eingeschränkter Leistungsanstrengung ausgeschlossen wurden, ergab sich für Patienten nach mittelschweren und schweren Kopfverletzungen eine Inzidenz der Störung der Geruchsdiskrimination, die mindestens zehnmal höher lag als für Patienten nach leichten Kopfverletzungen. Dieses Ergebnis wäre nicht mit einem konventionellen Untersuchungsplan ohne Symptomvalidierungstests erhalten worden. Der starke Einfluss, den die Schwere einer Kopfverletzung auf die Ergebnisse im Geruchstest ausübt, wurde erst nach Ausschluss all jener Fälle sichtbar, deren Ergebnisse wegen unzureichender Testmotivation oder einer Symptomaggravation ungültig waren. Wenn wir also die Beziehung zwischen der Störung des Geruchssinnes und der Schwere einer Kopfverletzung untersuchen,

dann hängt das Ergebnis davon ab, ob wir die Testmotivation in der Analyse berücksichtigen. Die davon abhängende unterschiedliche Befundlage stellt die Grundlage für unsere Theorienbildung dar. Letzten Endes fallen unsere Theorien über den Geruchssinn und das Gehirn unterschiedlich aus, je nachdem, ob wir die Leistungsmotivation messen oder nicht (Green, Rohling, Iverson & Gervais, 2003).

Dies gilt auch für eine Reihe anderer Studiendesigns. So glauben beispielsweise viele Menschen, dass eine Major Depression kognitive Störungen verursacht. Ohne Berücksichtigung der Leistungsanstrengung, also in fast allen früheren Untersuchungen, fand sich, dass Patientengruppen mit Major Depression signifikant schlechter als gesunde Kontrollpersonen abschnitten (z.B. Veiel, 1997). Wenn die Testmotivation jedoch kontrolliert wird, verschwindet dieser Unterschied. Rohling, Green, Allen und Iverson (2002) untersuchten den Einfluss einer Depression auf neuropsychologische Testwerte anhand einer großen Patientenstichprobe. Sobald alle Fälle mit einer unzureichenden Motivation, valide Testergebnisse zu erreichen, ausgeschlossen wurden, gab es keinen Zusammenhang mehr zwischen den Werten im Beck Depression Inventory und der Leistung in irgendeinem der eingesetzten neuropsychologischen Tests.

Der Einsatz von Symptomvalidierungstests im Rahmen von Gruppenstudien ist deshalb von Bedeutung, weil sie es uns gestatten, die größte identifizierbare Fehlerquelle, die wir kennen, aus unseren Daten zu entfernen. Die Entfernung eines solchen Artefakts wirft ein neues Licht auf althergebrachte Glaubenssätze und veranlasst uns zum Umdenken bei einigen unserer Theorien, die sich auf frühere Forschungsergebnisse gründen, ohne dass die Leistungsmotivation gemessen oder kontrolliert wurde. Das Konzept der Testmotivation beeinflusst gegenwärtig viele Forschungszweige, die Untersuchung von Personen nach einem sog. Schleudertrauma der Halswirbelsäule (Schmand, Lindeboom, Schagen, Heijt, Koene & Hamburger, 1997), Patienten mit chronischem Schmerz (Gervais, Green et al., 2001), Fibromyalgie (Gervais, Russell et al., 2001) und chronischem Müdigkeitssyndrom (van der Werf, Prins, Jongen, van der Meer & Bleijenberg, 2000) eingeschlossen. Auch bei Kindern werden SVT zunehmend eingesetzt. Mit großer Wahrscheinlichkeit wird die Erfassung der Testmotivation in klinischen Untersuchungen von Kindern in Zukunft genauso routinemäßig erfolgen, wie dies bei Erwachsenen immer mehr zur Routine wird. Dennoch werden weiterhin Studien veröffentlicht, in denen eine unzureichende Testmotivation als möglicher starker Einflussfaktor nicht ausgeschlossen wurde. So berichteten beispielsweise Niemi, Portin, Aalto, Hakala und Karlsson (2002) von schlechteren neuropsychologischen Testergebnissen bei Personen mit einer somatoformen Störung gegenüber Normalen. Eine ungenügende Leistungsanstrengung wurde als Erklärung für die niedrigen Testwerte jedoch durch die Autoren nicht in Betracht gezogen. Störungsbilder wie somatoforme Störungen sind aber gerade deswegen be-

merkwürdig, weil die subjektiven Beschwerden meist übertrieben erscheinen und bisher keine rationale biologische Erklärung für sie gefunden wurde.

### Schlussfolgerung

Zusammenfassend lässt sich sagen, dass die Messung der Testmotivation deshalb von Bedeutung ist, weil eine geringe Leistungsanstrengung während einer Testuntersuchung ganz offensichtlich keine Seltenheit darstellt. Bei den von uns untersuchten Personen erweist sich der Einfluss der Leistungsmotivation auf die Testwerte als weitaus stärker als der des Vorliegens einer schweren Schädel-Hirn-Verletzung oder neurologischen Erkrankung (Green et al., 2001). Wenn demnach die Testmotivation nicht korrekt gemessen wird, kommt man sowohl im klinischen Einzelfall als auch in Gruppenstudien zu falschen Schlussfolgerungen. Iverson und Binder (2000) bekräftigen, dass die Messung der Testmotivation im Rahmen von Begutachtungen eine notwendige Bedingung für gültige Aussagen darstellt. Die oben diskutierten Forschungsergebnisse legen nahe, dass eine Motivationsmessung im Kontext psychiatrischer Fragestellungen und selbst bei Kindern erforderlich sein kann. Eine niedrige Leistungsanstrengung kann in jedwedem Kontext die Testwerte beeinflussen.

Die an inzwischen mehr als 3000 Erwachsenen erhaltenen Daten zeigen, dass 50 Prozent der Varianz in neuropsychologischen Tests allein durch die Testmotivation erklärt werden könnten. In einer Übersichtsarbeit zum Word Memory Test, die in den Archives of Clinical Neuropsychology erschien, kommt Hartman (2002) zu dem Schluss, dass in Zukunft Studien, die ohne Messung der Testmotivation Zeitschriften zur Veröffentlichung angeboten werden, als nicht interpretierbar und damit nicht veröffentlichungswürdig betrachtet werden könnten. Die Forschungsergebnisse von Dr. Flaro und Dr. Courtney werfen die Frage auf, wie fehlerhaft auch die Testergebnisse von Kindern im Ergebnis fluktuierender Leistungsmotivation eigentlich sind. Ohne die Messung der Leistungsanstrengung laufen wir Gefahr, unsere Schlussfolgerungen auf der Grundlage ungültiger Daten aus Gruppenstudien oder klinisch untersuchten Einzelfällen zu ziehen. Wir wissen inzwischen, dass sowohl für Erwachsene als auch für Kinder ungültige Daten weit häufiger vorliegen, als wir früher dachten. Aus diesem Grunde ist ohne eine tatsächlich erfolgende Motivationsmessung eine Feststellung wie etwa »Der Patient hat die Tests voll motiviert bearbeitet« nicht länger hinzunehmen. Wenn die Leistungsanstrengung nicht adäquat gemessen wird, könnten die Testdaten und ihre Interpretation bei einer Tiefenprüfung nicht bestehen.

## L I T E R A T U R

- Allen, L., Conder, R.L., Green, P. & Cox, D.R. (1997) CARB '97 Manual for the Computerized Assessment of Response Bias. Durham, NC: CogniSyst.
- Binder, L.M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. *Journal of Clinical and Experimental Neuropsychology*, 15, 170-183
- Brockhaus, R. & Merten, T. (im Druck). Neuropsychologische Diagnostik suboptimalen Leistungsverhaltens mit dem Word Memory Test. *Nervenarzt*.
- Brockhaus, R. & Peker, Ö. (2003). Testing effort in Turkish-speaking subjects: validation of a translation of the Word Memory Test (WMT) [Abstract]. *Journal of the International Neuropsychological Society*, 9, 520.
- Doty, R.L., Yousem, D.M., Pham, L.T., Kreshak, A.A., Geckle, R. & Lee, W.W. (1997). Olfactory dysfunction in patients with head trauma. *Archives of Neurology*, 54, 1131-1140.
- Gervais, R., Green, P., Allen, L.M. & Iverson, G. (2001). Effects of coaching on symptom validity testing in chronic pain patients presenting for disability assessments. *Journal of Forensic Neuropsychology*, 2, 1-19.
- Gervais, R.O., Russell, A.S., Green, P., Allen, L.M., Ferrari, R. & Pieschl, S.D. (2001). Effort testing in patients with fibromyalgia and disability incentives. *Journal of Rheumatology*, 28, 1892-1899.
- Green, P. & Astner, K. (1995). Manual for the Oral Word Memory Test. Edmonton, Canada: Neurobehavioural Associates.
- Green, P., Allen, L. & Astner, K. (1996). The Word Memory Test: A User's Guide to the Oral and Computer-Administered Forms. US version 1.1. Durham, NC: CogniSyst.
- Green, P. & Flaro, L. (2003). Word Memory Test performance in children. *Child Neuropsychology*, 9, 189-207.
- Green, P. & Iverson, G. (2001a). Validation of the Computerized Assessment of Response Bias in litigating patients with head injuries. *The Clinical Neuropsychologist*, 15, 492-497.
- Green, P. & Iverson, G.L. (2001b). Effects of injury severity and cognitive exaggeration on olfactory deficits in head injury compensation claims. *NeuroRehabilitation*, 16, 237-243.
- Green, P., Iverson, G.L. & Allen, L. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury*, 13, 813-819.
- Green, P., Lees-Haley, P.R. & Allen, L.M. (2002). The Word Memory Test and the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology*, 2, 97-124.
- Green, P., Rohling, M.L., Lees-Haley, P.R. & Allen, L.M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, 15, 1045-1060.
- Green, P., Rohling, M.L., Iverson, G. & Gervais, R. (2003). Relationships between olfactory discrimination and head injury severity. *Brain Injury* 17, 479-496.
- Hartman, D.E. (2002). The unexamined lie is a lie worth fibbing: neuropsychological malingering and The Word Memory Test. *Archives of Clinical Neuropsychology*, 17, 709-714.
- Iverson, G.L. & Binder, L.M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma Rehabilitation*, 15, 829-858.
- Larrabee, G.J. (2000). Neuropsychology in personal injury litigation. *Journal of Clinical and Experimental Neuropsychology*, 22, 702-707.
- Niemi, P.M., Portin, R., Aalto, S., Hakala, M. & Karlsson, H. (2002). Cognitive functioning in severe somatization – a pilot study. *Acta Psychiatrica Scandinavica*, 106, 461-463.
- Rohling, M.L., Green, P., Allen, L.M. & Iverson, G.L. (2002). Depressive symptoms and neurocognition in patients passing symptom validity tests. *Archives of Clinical Neuropsychology*, 17, 205-222.
- Schmand, B., Lindeboom, J., Schagen, S., Heijt, R., Koene, T. & Hamburger, H.L. (1998). Cognitive complaints in patients after whiplash injury: the impact of malingering. *Journal of Neurology, Neurosurgery and Psychiatry*, 64, 339-343.
- Tan, J.E., Slick, D.J., Strauss, E. & Hultsch, D. (2002). How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist*, 16, 495-505.
- van der Werf, S.P., Prins, J.B., Jongen, P.J., van der Meer, J.W. & Bleijenberg, G., (2000). Abnormal neuropsychological findings are not necessarily a sign of cerebral impairment: A matched comparison between chronic fatigue syndrome and multiple sclerosis. *Neuropsychiatry, Neuropsychology and Behavioural Neurology*, 13, 199-203.
- Veiel, H.O.F. (1997). A preliminary profile of neuropsychological deficits associated with major depression. *Journal of Clinical and Experimental Neuropsychology*, 19, 587-606.

ANZEIGE HOGREVE 1/1