



# Validierungshandbuch

Verhalten, Driving Forces

# Inhalt

## Verhalten

<b>1. Einführung und Hintergrund</b>	<b>7</b>
1.1. Messung des persönlichen »Stils« – ein geschichtlicher Abriss	7
1.2. Verbindung zwischen DISC und den publizierten Instrumenten von Target Training International und Success Insights	9
<b>2. Grundlagen</b>	<b>9</b>
2.1. Theorie	9
2.2. Validität	10
2.3. Bestätigung im Zuge der Anwendung	11
<b>3. Instrumentierung</b>	<b>12</b>
3.1. Überblick über das Instrument	12
3.2. Skalenstruktur	13
DISC-Skalen	13
Adaptive und natürliche Dimension	15
Interpretation von »Diskrepanzen« basierend auf einem Abgleich von adaptiver und natürlicher Dimension	15
3.3. Validität und Reliabilität	17
Antwortbasierte Reliabilität und interne Struktur	17
Kontextbasierte Validität und Beziehungen zu anderen Variablen	18
Konvergente und divergente Evidenz	20
3.4. Situation und Umfeld	20
Situationsanforderungen	21
Soziale Erwünschtheit	21
Kultureller Einfluss	21
3.5. Messbarkeitsgrenzen	22
Häufigkeitszählungen und Ergebnismessung	22
Ipsative Messverfahren	22
Einfluss situationsbedingter Erwartungen	23
<b>4. Technische Informationen</b>	<b>23</b>
4.1. Überprüfung und Überarbeitung	23
Untersuchung der theoretischen Kohärenz	24
Item- und Skalenreliabilitäten	25
4.2. Normen und Populationsparameter	25
4.3. Geschlechterverteilung	26
4.4. Sprachversionen	26

<b>5. Instrumentenprotokolle und Anwendung</b>	<b>26</b>
Administrierung, Anwendung	26
Skalierung & Diagramme	27
Report	28
Feedback und Interpretation	28
<b>6. Anhang</b>	<b>28</b>
<b>6.1. INSIGHTS MDI®: Deutsche Version 2021</b>	<b>28</b>
Hintergrund	28
Normierungsstichprobe	29
Skalenreliabilität	29
<b>6.2. INSIGHTS MDI®: Deutsche Version 2014</b>	<b>30</b>
Hintergrund	30
Normierungsstichprobe	30
Skalenreliabilität	31
Stellungnahme Prof. Dr. Karlheinz Zwerenz	31
<b>6.3. INSIGHTS MDI®: Deutsche Version 2009</b>	<b>32</b>
Zusammenfassung	32
Hintergrund	32
Referenzpopulation	33
Skalenreliabilität	34
Skalenbeziehungen – Korrelationen	35
Schlussfolgerungen	36
<b>6.4. INSIGHTS MDI®: Deutsche Version 2006.f10, .f24</b>	<b>36</b>
Hintergrund	36
Reliabilität	37
Beziehung zwischen den Anwendungen	37
Schlussfolgerung	38
<b>7. Literaturliste</b>	<b>38</b>

## Driving Forces

<b>1. Einführung und Hintergrund</b>	<b>41</b>
<b>2. Grundlagen</b>	<b>41</b>
<b>3. Instrumentierung</b>	<b>43</b>
3.1. Entwicklung – Theoretische Validität	43
3.2. Bestimmung einer Messstruktur	43
3.3. Konstruktvalidierung	44
Analyse ohne konzeptionelle Beschränkungen – Faktorenanalyse	44
Analyse zwecks Bestätigung einer vorgegebenen Struktur – Skalenreliabilität	45
Analyse der Korrelationen zwischen Skalen – Korrelationen zwischen Skalen	45
Entwicklung von Skalenwerten und Referenznormen	45
3.4. Freigabe und Follow-Up – Praxisbestätigung	45
<b>4. Überblick über das Instrument</b>	<b>46</b>
4.1. Skalenstruktur	46
4.2. Validität und Reliabilität	46
Antwortenbasierte Reliabilität und interne Struktur	46
Kontextbasierte Validität und Beziehungen zu anderen Variablen	48
Konvergente und divergente Evidenz	49
4.3. Situation und Umfeld	49
Situationsanforderungen	49
Soziale Erwünschtheit	50
Kultureller Einfluss	50
4.4. Messbarkeitsgrenzen	51
Itemgewichtung und Skalenkonstruktion	51
Ipsative Messverfahren	52
<b>5. Technische Informationen</b>	<b>52</b>
5.1. Überprüfung und Überarbeitung	52
Untersuchung der theoretischen Kohärenz	53
Item- und Skalenreliabilitäten	53
5.2. Normen und Populationsparameter	54
Geschlechterverteilung	54
Sprachversionen	55

<b>6. Instrumentenprotokolle und Einsatz</b>	<b>55</b>
Verwaltung und Einsatz	55
Report	55
Skalierung und Diagramme	56
Training und Interpretation	57
<b>7. Literaturliste</b>	<b>57</b>
<b>8. Anhang</b>	<b>58</b>
<b>8.1. PIW / Motive (Driving Forces), Deutsche Version 2021</b>	<b>58</b>
Hintergrund	58
Normierungsstichprobe	58
Skalenreliabilität	59
<b>8.2. PIW – Persönliche Interessen und Werte (PIW2): Deutsch 2014</b>	<b>59</b>
Hintergrund	59
Normierungsstichprobe	60
Skalenreliabilität	60
<b>8.3. PIW – Persönliche Interessen und Werte (PIW2): Deutsch 2008</b>	<b>61</b>
Schlussfolgerungen für den Einsatz	61
Daten und Methodik	61
Beschreibung der Stichprobe	61
Reliabilität und Itemkohärenz	61
Beziehung zwischen den Skalen	62
Referenznormen	63
<b>8.4. PIW – Persönliche Interessen und Werte PIW 2 – Deutsche Version: 2003</b>	<b>63</b>
Zusammenfassung	63
Hintergrund	63
Reliabilität	64
Korrelationen	65
Schlussfolgerungen	65
Stellungnahme Prof. Dr. Karlheinz Zwerenz	66
<b>Copyright</b>	<b>67</b>





# Verhalten

## 1. Einführung und Hintergrund

Dieses Handbuch wurde zu dem Zweck entwickelt, die Konstruktreliabilität und die Validität des von Target Training International, Ltd., (TTI) und Success Insights, Ltd., Scottsdale, Arizona, USA, veröffentlichten **INSIGHTS MDI®**-Instruments zu demonstrieren. Es richtet sich an die Nutzer des Instruments, an potentielle Klienten und Kunden, die mehr erfahren wollen zu den wissenschaftlichen Spezifikationen des Instruments und der Reports. Ziel ist es dabei, mittels eines verifizierbaren Evidenzmusters die Reliabilität und Validität des **INSIGHTS MDI®**-Instruments sowie seine Eignung für unterschiedliche Bereiche der persönlichen Entwicklung und zur Unternehmensentwicklung, sowie seine Anwendbarkeit in einer Vielzahl von unternehmensrelevanten Situationen (z.B: Teamzusammensetzung) aufzuzeigen.

Die wichtigste Absicht des Handbuchs ist es, die erforderlichen Informationen zu liefern, ohne allzu sehr in die Länge zu gehen. Das erhöht die Praktikabilität des Handbuchs und kommt der Präzision der vermittelten Informationen zugute. Einige Teile des Handbuchs handeln von konkreten statistischen Verfahren und Methoden. Deren Darstellung wird komprimiert sein und die Informationen auf empirische und statistische Weise präsentieren. Es wird angenommen, dass einige aber nicht alle Leser über testtheoretisches Grundwissen verfügen. Deshalb werden die Informationen sowohl in statistisch-tabellarischer Form als auch in erklärender Textform dargeboten.

Die in diesem Handbuch bereitgestellten Forschungsergebnisse und Statistiken wurden nach den gemeinsam von der American Educational Research Association, der American Psychological Association und dem National Council on Measurement in Education herausgegebenen „Standards for Educational and Psychological Testing“ (1999) abgefasst und durchgeführt. Die Richtlinien liefern die Standards, nach denen viele US-basierte aber auch internationale Testverfahren entwickelt und validiert werden. Absicht dieses Handbuchs ist es, diese Spezifikationen zu erläutern und den Leser zu unterstützen, die Standards genauer kennenzulernen. Der Leser wird zudem ermuntert, aktiv Fragen zu den im Markt verfügbaren Tests zu stellen und sich ein Bild davon zu machen, inwieweit diese Tests ähnliche Richtlinien befolgen wie das **INSIGHTS MDI®**-Instrument und seine Reports.

### 1.1. Messung des persönlichen »Stils« – ein geschichtlicher Abriss

Das **INSIGHTS MDI®**-Instrument gehört seiner Art nach in die Kategorie von Instrumenten, die zusammengefasst als »Persönlichkeitstests« bezeichnet werden. Die Autoren dieses Handbuchs und die Herausgeber von Success Insights und Target Training International Performance Systems (TTI) bevorzugen aus diversen Gründen den Begriff »Stil« anstelle von »Persönlichkeit«. Erstens ist »Persönlichkeit« ein sehr komplexer und globaler Begriff, der sich auf eine große Bandbreite von Verhaltensweisen und Eigenschaften eines Menschen bezieht. Zweitens beschreibt der Begriff »Stil« eher, wie jemand konkret etwas tut – eine Gedanke, den wir ursprünglich Fritz Perls verdanken –, und eignet sich deshalb besser für die Zwecke und Ziele des **INSIGHTS MDI®**-Instruments und seiner Ergebnisreports.

Historisch betrachtet existieren diverse Methoden, um Persönlichkeit respektive Verhaltensweisen und »Stil« eines Menschen zu messen. In den frühen Arbeiten von Kraepelin (1892) mit seinen freien Assoziationstests wurde der Testperson eine Liste mit Wörtern vorgelegt, zu der sie jeweils das erste Wort nennen sollte, das ihr dazu in den Sinn kam. Die Methode der freien Assoziation wurde und wird bis heute für eine Vielzahl von Testzwecken verwendet. Kritiker der Methode verweisen auf Probleme mit der Ergebnisermittlung, der Beurteiler-Reliabilität und eine mögliche Verfälschung seitens der Testperson.

Eine Antwort auf die Probleme im Zusammenhang mit der Ergebnisermittlung und der Beurteiler- Reliabilität bot die Methode des Selbstberichts. Eine frühe Form dieser Testtechnik wurde während des Ersten Weltkriegs von Woodworth entwickelt (DuBois 1970, Goldberg 1971, Symonds 1931). Der ursprüngliche Zweck war ein Test, um diejenigen zu identifizieren, die nicht für den Militäreinsatz geeignet waren. Der Krieg endete, bevor der Test einsatzbereit war; zivile Formen wurden jedoch für Erwachsene und Kinder entwickelt. Das Woodworth Personal Data Sheet diente als Prototyp und frühes Vorbild für viele spätere Inventare. Manche Entwicklungen erforschten konkrete Bereiche wie berufliche Ausrichtung, schulische Ausrichtung, Familie usw. Andere Tests erkundeten zwischenmenschliche Reaktionen in gesellschaftlichen Situationen, und später kamen Tests hinzu, die sich auf Interessen und Einstellungen konzentrierten. Das **INSIGHTS MDI®**-Instrument und seine Berichte sind in diesem Selbstbericht-Genre angesiedelt.

Eine weitere häufig verwendete Testmethode ist der Leistungs- oder Situationstest. Dabei wird die Testperson gebeten, eine Aufgabe auszuführen, die Messung erfolgt auf der Basis dieser Leistung. Der konkrete Zweck dieser Tests bleibt der Testperson zumeist verborgen. Eine frühe Anwendung dieses Modells wurde von Hartshorne, May u.a. (1928, 1929, 1930) entwickelt und für Schulkinder standardisiert. Situationstests für Erwachsene wurden während des Zweiten Weltkriegs vom Assessment Program des Office of Strategic Services entwickelt. Diese Tests waren für damalige Verhältnisse hochgradig komplex und erforderten eine detaillierte Vorbereitung und eine geschickte Ausführung. Dennoch gab es Probleme mit der Beurteiler-Reliabilität, und die Interpretation der Antworten erfolgte vergleichsweise subjektiv.

Eine andere mögliche Methode ist die der projektiven Testgestaltung. Bei dieser Methode erhält die Testperson eine Aufgabe ohne vorgegebenen Lösungsweg oder ein Bild einer Situation, das sie beschreiben soll. Auch hier wird der Zweck des Tests etwas verschleiert, um die Wahrscheinlichkeit zu reduzieren, dass die Testperson eine bevorzugte Reaktion erzeugt oder simuliert. Wie bei der freien Assoziation und einigen Situationstests bleibt auch hier Raum für Fehler im Bereich der Beurteiler-Reliabilität und demzufolge Schwankungen in der Ergebnisermittlung infolge der subjektiven Natur der Instrumentierung.

Das **INSIGHTS MDI®**-Instrument und die Reports verwenden die Methode der Selbsteinschätzung, die auf dem Wege einer objektiven Ergebnisermittlung die Probleme mit der Beurteiler-Reliabilität eliminiert. Mit der Methode der Selbstbeurteilung erhebt das Instrument die eigenen Selbstwahrnehmung und wertet die Ergebnisse der Selbsteinschätzung aus. Probleme gibt es jedoch hinsichtlich der Genauigkeit der gegebenen Selbstauskünfte und des Blickwinkels der Selbstwahrnehmung. Die Testperson wird deshalb ständig aufgefordert, ehrliche Antworten zu geben und sich dabei klar an die Instruktionen zu halten. Diese Methodik wird vielfach eingesetzt und in vielen akademischen und kommerziellen Instrumenten genutzt.



## 1.2. Verbindung zwischen DISC und den publizierten Instrumenten von Target Training International und Success Insights

In den Jahren 1983-84 erwarb Target Training International, Ltd., (TTI) im Rahmen einer Lizenzvereinbarung ein DISC-basiertes Instrument. Seither haben TTI und Success Insights in substantiellem Umfang Energie und Ressourcen in die kontinuierliche statistische Validierung des Instruments und der Reports investiert. In regelmäßigen Abständen wurden Veränderungen an den neueren Versionen des Instruments vorgenommen, um mit den heute gebräuchlichen Begriffen und Deskriptoren mitzuhalten, insbesondere um solche Begriffe und Deskriptoren zu ersetzen, die vor Jahrzehnten nützlich waren, ihre Aussagekraft aber im 21. Jahrhundert eingebüßt haben. TTI und Success Insights gehören zu den wenigen DISC-Anbietern, die für ihre statistische Validierung aktuelle Ergebnisse aus dem 21. Jahrhundert verwenden, die von jenen Sprach- und Kulturgruppen stammen, für die das Instrument gedacht ist. Das ermöglicht eine gesteigerte Reliabilität und Validität der Reports indem die Ergebnisse der einzelnen Testpersonen mit einer ausreichend großen, repräsentativen, aktuellen und auch kulturell passenden Datenbank abgeglichen werden.

## 2. Grundlagen

### 2.1. Theorie

Das **INSIGHTS MDI®**-Modell basiert auf dem weltweit vielfach verwendeten vierdimensionalen DISC-Modell und findet breite Akzeptanz in kommerziellen und nichtkommerziellen Unternehmen und Institutionen. Vierdimensionale Modelle haben eine lange Tradition in Philosophie und Psychologie. Empedokles (495 bis 435 v. Chr.) war der Gründer einer medizinischen Schule in Sizilien. Er gründete seine Lehre auf die Vorstellung, dass alle Dinge aus vier Elementen gemacht sind: Luft, Erde, Feuer und Wasser. Diese Elemente lassen sich auf unendlich viele Arten zu anderen Stoffen kombinieren. Hippokrates (400 v. Chr.) vertrat die These, dass vier verschiedene Arten von Klima und Landschaft das Verhalten und Auftreten der Menschen beeinflusst, die in diesen Umgebungen leben. Daraus entwickelte er die Theorie von den vier Temperamenten (sanguinisch, melancholisch, cholerisch und phlegmatisch) und assoziierte diese Temperamente mit vier Körperflüssigkeiten (Blut, schwarzer Galle, gelber Galle und Schleim). Galen von Pergamon (130-200 n. Chr.) assoziierte die vier Körpersäfte Blut, schwarze Galle, gelbe Galle und Schleim mit bestimmten Verhaltensweisen. Hinzu kam die Vorstellung von den vier externen Zuständen warm, kalt, trocken und feucht, die auf unseren Körper einwirken. Wichtige empirische Beiträge lieferte im Jahr 1921 Carl Gustav Jung in seinem Werk *Psychologische Typen*. Er identifizierte vier »Typen« psychologischer Funktionen: Denken, Fühlen, Empfinden und Intuition. Diese vier Typen werden weiter unterteilt durch die beiden Energiekräfte Introversion und Extraversion, die als Kategorien zu den vier Typen psychologischer Eigenschaften hinzukamen. Der Hauptentwickler des DISC-Modells ist Dr. William Moulton Marston von der Harvard University. Bekannt für sein Buch *The Lie Detector* aus dem Jahr 1938, bereicherte er das Gebiet der Typologien mit wichtigen Beiträgen. Das **INSIGHTS MDI®**-Instrument und die meisten anderen DISC-Instrumente basieren auf Marstons Originalentwurf. Marston wurde im Jahr 1893 geboren und erwarb an der Harvard University drei akademische Grade: B.A. (1915), L.L.B. (1918) und Ph.D. (1921). Marston arbeitete als Lehrer und psychologischer Berater und war Fakultätsmitglied an der American University, Tufts University, Columbia University und New York University. Er publizierte im *American Journal of Psychology*, in der *Encyclopedia Britannica* und in der *Encyclopedia of Psychology* und verfasste fünf Bücher.

Im Jahr 1928 erschien Marstons Buch *The Emotions of Normal People*, in welchem er die DISC-Theorie vorstellte, die noch heute verwendet wird. Seiner Vorstellung zufolge verhalten sich Menschen entlang zweier Achsen: Ihre Handlungen sind eher aktiv oder eher passiv, und sie nehmen die Umwelt eher als ihnen freundlich oder feindlich gesinnt wahr. Indem wir die Achsen rechtwinklig zueinander anordnen, ergeben sich vier Quadranten, die jeweils ein bestimmtes Verhaltensmuster beschreiben:

1. Dominanz erzeugt Aktivität in einem feindlichen Umfeld.
2. Ansporn erzeugt Aktivität in einem freundlichen Umfeld.
3. Festigkeit erzeugt Passivität in einem freundlichen Umfeld.
4. Folgsamkeit erzeugt Passivität in einem feindlichen Umfeld.

Marston war überzeugt, dass die Menschen in der Regel eine Selbstvorstellung entwickeln, die im Wesentlichen auf einen der vier Faktoren abgestimmt ist. Mit Marstons Theorie ist es möglich, die Kräfte der wissenschaftlichen Beobachtung auf das Verhalten anzuwenden und Objektivität und Beschreibung zu verstärken und Subjektivität und Bewertung zu verringern. Walter Clark konstruierte in den 1950er Jahren als Erster ein psychologisches Instrument auf der Basis der Marstonschen Theorie. Es trug den Namen »Activity Vector Analysis«. („... Manche von Clarks ursprünglichen Partnern verließen später sein Unternehmen und verfeinerten das Format, indem sie ihre eigenen Selbsteinschätzungs-Attributlisten für das Instrument erzeugten. Heute verwenden viele Anbieter die Marston-Theorie als Ausgangspunkt für die Verhaltensforschung mittels eines auf dem Prinzip der Selbsteinschätzung basierenden beschreibenden Verhaltensinstruments. **INSIGHTS MDI®** und seine Nebenformen erlauben uns, Verhaltensmuster in einer Weise zu identifizieren, dass damit die Marston-Theorie praktisch anwendbar wird.

## 2.2. Validität

Eine Vielzahl akademischer Studien bestätigt die theoretischen Konstrukte. Die wichtigsten von ihnen werden im Folgenden erwähnt. Weil das Modell von mehreren kommerziellen Anbietern verwendet wird, lässt sich die vollkommene Offenlegung ihrer Forschungsergebnisse aufgrund von Eigentumsrechten nur schwer bewerkstelligen. Success Insights und TTI Performance Systems sind führend in der Offenlegung bestimmter Ergebnisse, während andere geschützt bleiben, wie beispielsweise spezifische Ergebnisermittlungsschlüssel und Itemanalysen. Eine kurze Zusammenfassung der für die Weiterentwicklung der DISC-Theorie maßgeblichen Forschungsergebnisse finden Sie in der folgenden Aufzählung.

- 1967 – Dr. John G. Geier, University of Minnesota, »A Trait Approach to the Study of Leadership in Small Groups«, *The Journal of Communication*, Dezember 1967.
- 1977 – Dr. John G. Geier, University of Minnesota, *The Personal Profile System*, Performax Systems International, Minneapolis, MN.
- 1983 – Dr. Sylvan J. Kaplan erstellte eine Studie über das Personal Profile System und verglich es mit den folgenden psychologischen Instrumenten:
- Hamburg-Wechsler-Intelligenztest für Erwachsene
- Myers-Briggs-Typindikator

- 16-Persönlichkeits-Faktoren-Test, Schneewind, Kurt A., Schröder, G. & Cattell, R.B. (1994) Minnesota Multiphasic Personality Inventory
- Strong Interest Inventory
- 1989 – Dr. Russell J. Watson, Wheaton College, »A Statistical Comparison of the TTI Style Analysis and the Performax Personal Profile System«, Wheaton, IL. Diese Studie machte seinerzeit deutlich, dass Style Analysis und Personal Profile Systems keine statistisch signifikant unterschiedlichen Ergebnisse lieferten.
- 2002-2004 – Unter der Leitung von Dr. Peter T. Klassen, Professor Emeritus, College of DuPage, Glen Ellyn, IL, unterzogen Success Insights und TTI Performance Systems ihre Instrumente fundierten statistischen Analysen und nahmen subtile Veränderungen am Instrument und an der Ergebnisermittlung vor. Aufgrund dieser statistischen Analysen konnten alle in der anfänglichen Stichprobe gemessenen Variablen zur Konstruktreliabilität sowie die Augenscheinvalidität der Reports verbessert werden.

Im Laufe mehrerer Jahrzehnte haben zahlreiche Studien das **INSIGHTS MDI®-DISC-Modell** als Basis für Verhaltensforschungen in einer Vielzahl von Themenbereichen genutzt, darunter:

- »A case study which utilizes type indicators to analyze 360-degree performance assessments«, Dissertation in Pädagogischer Psychologie von George Landon Anderson, University of Louisville, Louisville, KY, USA, 1995.
- »Investigating the effects of behavior constructs on academic persistence in engineering, creativity, and risk-taking«, Dissertation in Psychologie von Viveca K. Deanes, Texas A & M University, USA, 2003.
- »Behavioral Style as a predictor of hearing aid return for credit«, Dissertation in Psychologie von Ateven A. Huart, Central Michigan University, USA, 2002.
- »Market segmentation: Exploring the need for further consumer behavior analysis«, Dissertation in Persönlichkeitspsychologie von James Joseph Kolacek, III, Nova Southeastern University, USA, 1999.
- »Jury deliberation style and just world belief«, Dissertation von Harry Naifach, Kent State University, USA, 2002.

## 2.3. Bestätigung im Zuge der Anwendung

Neben der formellen akademischen Forschung zum DISC-Modell oder der Verwendung des DISC-Modells als Basis für beschreibende oder experimentelle Forschung wurden auch zahlreiche informelle Studien auf der Team- oder Unternehmensebene durchgeführt. Während viele dieser Studien nicht veröffentlicht wurden oder Eigentum des Beratersunternehmens bleiben, gibt es auch Studien, die durch Artikel der Wirtschaftspresse oder durch Berater veröffentlicht wurden, um ihre Ergebnisse einem breiteren Interessentenkreis zugänglich zu machen. Manche dieser Ergebnisse werden wir im Folgenden erwähnen. Das DISC-Modell kommt im Unternehmensumfeld seit dreißig Jahren vielfältig zum Einsatz. Anstelle eines historischen Abrisses über diese Entwicklung wollen wir an dieser Stelle mehrere in jüngster Zeit veröffentlichte Quellen vorstellen. Die unten aufgeführten Quellen stammen aus frei zugänglichen Presseartikeln in den USA und andernorts.

Auch wenn es sich bei diesen Quellen nicht um begutachtete Fachartikel oder Dissertationen handelt, liefern sie doch ein Evidenzmuster und diverse Informationen zur Anwendung des DISC-Modells in einer Vielzahl von Unternehmensumgebungen:

- »What is Your Communication Style?«, Lisa Aldisert in Bank Marketing, Oktober 2000, Bd. 32, Nr. 10, S. 46. Präsentiert Kommunikationstipps für das Banken-Marketing. Betont die Vorteile eines verhaltensorientierten Kommunikationsstils. Liefert Informationen zu den verschiedenen Typen von Verhaltensstilen in der Kommunikation.
- »D.I.S.C. Drives«, Evan Cooper in Financial Planning, September 2000, Bd. 30, Nr. 9, S. 107. Erforscht die vier von Dr. Joseph Marshall identifizierten Verhaltensstile von Börsenhändlern. »Verhaltenstypen von Finanzplanern.« »Wie Börsenhändler von ihren Verhaltensstilen profitieren können.«
- »What's Your Client's Style?«, Susan Foster in Selling, Dezember 1998, Bd. 6, Nr. 5, S. 8. Stellt eine Klassifizierung individueller Verhaltenspräferenzen vor, mit deren Hilfe Verkäufer ihre Kunden besser verstehen können. Drei entscheidende Schritte, wie sich die Kenntnis von Verhaltensstilen auf Verkaufssituationen anwenden lässt.
- »Cracking the Communication Ice. What's Below the Surface?«, Kenneth R. Kramm und Deborah A. Kramm in Training & Development Journal, April 1989, Bd. 43, Nr. 4, S. 56. Präsentiert Aufwärmübungen, um in Seminaren und Workshops die Diskussionen anzuregen und die Beteiligung der Teilnehmer zu fördern. Verhaltensstile, die entscheiden helfen, wie sich gezielte Wortauswahl als Eisbrecher in der Kommunikation anwenden lässt und eine Evaluation dieser Übungen.
- »Preferences for Behavioural Style of minority and majority members who anticipate group interaction«, Alain Van Hiel und Ivan Mervielde in Social Behavior & Personality: An International Journal, 2001, Bd. 29, Nr. 7, S. 701. Die Studie untersucht, ob mögliche Minderheits- und Mehrheitsvertreter bestimmten Verhaltensstilen eine hohe Effektivität zuschreiben, um in einer anstehenden Gruppeninteraktion den eigenen Einfluss zu maximieren.

Diese Liste ist nur eine Auswahl, aber sie umfasst eine große Bandbreite von Anwendungen des DISC-Modells in den unterschiedlichsten Wirtschaftsbereichen.

## 3. Instrumentierung

### 3.1. Überblick über das Instrument

INSIGHTS MDI® baut auf Design und Entwicklung des bisherigen DISC-Instruments von TTI sowie auf weiteren DISC-Instrumenten auf. Die Testpersonen bringt die jeweils vier beschreibenden Adjektive einer Frage in einer Rangreihung, beginnend mit das »am besten zu mir passt«, bis zu demjenigen, das »am wenigsten zu mir passt«. Mit Hilfe dieser Rankings werden die vier DISC-Skalen gebildet. Die vier Skalen heißen »Dominanz« (D für dominance), »Initiative« (I für influence), »Stetigkeit« (S für steadiness) und »Gewissenhaftigkeit« (C für compliance).

Diese Skalenergebnisse werden mittels zweier Diagramme sowie erklärenden Texten vermittelt. Die Diagramme, eines für die adaptiven Verhaltensstile und eines für die natürlichen Verhaltensstile, veranschaulichen grafisch die relativen Werte der vier Skalenergebnisse. Als Grundlage dient der Vergleich der Häufigkeiten der von der Testperson ausgewählten Beschreibungen mit dem prozentualen Ranking einer ähnlichen Itemauswahl in der Referenzpopulation.

## 3.2. Skalenstruktur

Das INSIGHTS MDI®-Instrument umfasst 96 Adjektive die in 24 Fragen à vier Items gruppiert sind. In jeder Frage befindet sich zu jedem der vier Skalenkonstrukte jeweils ein zugehöriges beschreibendes Item. Je nach Testformat wählen die Testpersonen entweder selbst je ein »Passt am besten zu mir«- und ein »Passt am wenigsten zu mir«-Item aus, oder sie ordnen die vier Items entlang der Schiene »Passt am besten zu mir« bis »Passt am wenigsten zu mir«. Aus diesen Antworten werden zwei Dimensionen à vier Skalen gebildet. Die zwei Dimensionen heißen »adaptiver Stil« und »natürlicher Stil«, und jeder Stil enthält die vier Skalen D, I, S und C. Für die adaptiven Skalen werden die als »Passt zu am besten zu mir« ausgewählten oder in der Rangfolge an die erste Stelle gesetzten, den jeweiligen Skalen zugeordneten Adjektive aufaddiert. Für die natürlichen Skalen werden die als »Passt am wenigsten zu mir« ausgewählten oder in der Rangfolge an die letzte Stelle gesetzten, den jeweiligen Skalen zugeordneten Items aufaddiert, wobei das Skalenergebnis das Produkt der inversen Verrechnung der ausgewählten Items darstellt. Das bedeutet, dass die jeweilige Skala ein umso höheres prozentual dargestelltes Ergebnis erzielt, je weniger häufig Adjektive einer bestimmten Farbgruppe (bzw. D,I,S,G Gruppe) gewählt werden. Bedeutung und Nützlichkeit der zwei getrennten Skalendimensionen werden später diskutiert.

Peabody untersuchte die Bedeutung der Ergebnispolarität im Vergleich zur Zahl der Skalenpunkte. Er kommt zu dem Schluss: »... zusammengesetzte Ergebnisse spiegeln in erster Linie die Richtung der Antworten und weniger ihren Ausschlag wider. Für die Praxis folgt daraus, dass es gerechtfertigt ist, die Items getrennt nach der Richtung der Antwort aufzuaddieren« (Peabody, 1962, S. 73). Diese Schlussfolgerung wird von anderen Autoren geteilt (Komorita 1963, Komorita und Graham 1965, Jacoby und Matell 1971, Matell und Jacoby 1972).

### DISC-Skalen

Jede der vier DISC-Skalen ist als eine Konstellation von Verhaltensweisen zu verstehen, die kohärent mit dem psychologischen »Stil« einer Testperson zusammenhängen. Wie sowohl im theoretischen Abschnitt (Seite 9) als auch im Abschnitt »Situation und Umfeld« dargelegt, wird der konkrete Ausdruck der deskriptiven Verhaltensweisen unter Umständen von der Situation, der Kultur und anderen Umgebungsvariablen beeinflusst. Zu Kommunikationszwecken bezeichnen wir jede der vier Skalen mit dem ersten Buchstaben des Hauptquadranten (D,I,S,G), das dazu gedacht ist, die jeweilige zugehörige Verhaltenskonstellation zusammenzufassen. Je nach der Dominanz oder Rezessivität des Quellmerkmals (entspricht der zugrundeliegenden psychologischen Eigenschaft) zeigt das Verhalten einer Person die beschriebenen (Oberflächen-) Verhaltensweisen in stärkerem oder schwächerem Maße. An der Oberfläche sind die beschriebenen Verhaltensweisen deshalb je nach der Situation und den kulturellen Erwartungen und Randbedingungen entweder sichtbar oder nicht sichtbar.

Die folgende Tabelle listet einige der in früheren Versionen vorgeschlagenen theoretisch identifizierten nach außen sichtbaren Verhaltensweisen auf. Diese frühen Konstrukte wurden von Geier im Personal Profile System (1977) vorgeschlagen. Ihre Veröffentlichung ist nur ein Beispiel für die Bemühungen, die darauf abzielen die hohe Konstruktvalidität zu belegen.

Mit der Entwicklung statistischer Instrumente und Computer wurde es möglich, über einfache Korrelationen hinauszugehen. Eine systematische Überprüfung bestätigt, dass viele der aufgelisteten nach außen sichtbaren Verhaltensweisen mit anderen kovariieren, wodurch sich zugrundeliegende latente psychologische Merkmale identifizieren lassen. Dieselbe Untersuchung an aktuellen Stichproben hat zudem nach außen sichtbare Verhaltensweisen identifiziert, die wegen des Einflusses der sozialen Erwünschtheit (beziehungsweise Unerwünschtheit) nicht dazu geeignet sind, latente Merkmale zu identifizieren. Auf Basis dieser Grundlage werden überarbeitete aktuelle Instrumente entwickelt.

Dominanz antreibend, cholerisch	Ansporn expressiv, sanguin	Festigkeit liebenswert, phlegmatisch	Folgsamkeit analytisch, melancholisch
abenteuerlustig	gesellig	zuvorkommend	akkurat
aggressiv	bejahend	liebenswert	analytisch
ehrgeizig	fesselnd	fürsorglich	sorgfältig
streitbar	charmant	einfühlsam	vorsichtig
autoritär	fröhlich	willfährig	gewissenhaft
kühn	umgänglich	rücksichtsvoll	konservativ
dreist	zuversichtlich	konventionell	kontemplativ
anspruchsvoll	überzeugend	bedächtig	konventionell
wetteifernd	herzlich	hingebungsvoll	höflich
konfrontativ	angenehm	pflichtbewusst	abwägend
mutig	anschaulich	gelassen	diplomatisch
bestimmend	überschwänglich	treu	diszipliniert
fordernd	großherzig	freundlich	ausweichend
entschlossen	begeisterungsfähig	zuhörend	anspruchsvoll
direkt	ausdruckstark	inaktiv	pedantisch
dominierend	freundlich	loyal	sachlich
energisch	kontaktfreudig	bescheiden	pflichtbewusst
unternehmungslustig	gesprächig	ungezwungen	qualitätsbewusst
neugierig	impulsiv	unaufdringlich	logisch
zielorientiert	beeinflussend	verbindlich	ernsthaft
ungeduldig	inspirierend	passiv	methodisch
unabhängig	offen	abwartend	ordentlich
innovativ	optimistisch	vereinnahmend	organisiert
wissbegierig	extrovertiert	berechenbar	geduldig
beharrlich	verspielt	entspannt	perfektionistisch
vorangehend	ausgeglichen	verlässlich	präzise
kraftvoll	beliebt	zufrieden	unhektisch
verantwortlich	einnehmend	heiter	begreifend
ergebnisorientiert	ungezwungen	zurückhaltend	reserviert
risikofreudig	spontan	stabil	respektvoll
initiativ	stimulierend	beständig	zurückhaltend
egostark	gesprächig	teamorientiert	kontrolliert
willensstark	vertrauensvoll	umsichtig	strukturiert
unbeugsam		verständnisvoll	systematisch
unbeirrt		bereitwillig	taktvoll
unnachgiebig			

Tabelle 1: Repräsentative Merkmalsliste aus Geiers Personal Profile System, 1977



## Adaptive und natürliche Dimension

Das **INSIGHTS MDI®**-Instrument verwendet zwei miteinander verwandte Herangehensweisen zur Messung dieses DISC-Verhaltensstils (Merkmale). Nach dem Prinzip der reziproken Evaluationshandlung sind positiv und negativ gesetzte Aktivierungsfunktionen reziprok bestimmt. Mit anderen Worten, auf einer bipolaren Zustimmungsskala ist maximale Zustimmung reziprok zu minimalem Widerspruch und umgekehrt (Warburton, ohne Datum). Diese dichotomen Antworten resultieren in zwei Mengen von DISC-Skalen, die als adaptive und natürliche Dimension bezeichnet werden.

Die adaptiven und natürlichen Skalen der DISC-Messung werden konstruiert, indem die ausgewählten Adjektive aufaddiert werden, die jedes der vier Skalenkonzepte charakterisieren. Jede der 24 Fragen des Instruments enthält vier Begriffe oder Formulierungen. Jedes dieser Items ist eindeutig mit einer der Skalen assoziiert. Die Testpersonen wählen ein Item als »passt am besten zu mir« und ein Item als »passt am wenigsten zu mir« aus oder ordnen die vier Items nach »passt am besten zu mir« bis »passt am wenigsten zu mir«. So werden alle vier Skalen der beiden Stile (natürlicher und adaptierter) gemäß der Zuordnung von Items zur Selbstwahrnehmung von Merkmalen konstruiert.

Weil die vier adaptiven DISC-Skalen gemäß der Charakterisierung von Items als »Passt am besten zu mir« konstruiert werden, sind diese Skalen ein Hinweis auf das „öffentliche“ bewusst wahrgenommene Selbst, wie es sich in der gegebenen Situation präsentiert. Daraus folgt logischerweise, dass bei diesen Antworten die Gefahr absichtlicher Verfälschungen seitens der Testperson am größten ist.

Weil die vier natürlichen DISC-Skalen gemäß der Charakterisierung von Items als »Passt am wenigsten zu mir« konstruiert werden, haben diese Skalen eine umgekehrte Gewichtung. Je mehr Items als »Passt am wenigsten zu mir« eingestuft werden, desto geringer (niedriger Skalenwert) fällt das Ergebnis aus, und desto weniger wird die zugeordnete Eigenschaft als von der Testperson als eigene, zum Selbst gehörende Eigenschaft wahrgenommen. Sehr hohe Skalenergebnisse ergeben sich dort, wo ein einer Skala zugeordnetes beschreibendes Item NICHT ausgewählt wird. Daraus folgt logischerweise, und Vergleiche zwischen verschiedenen Gruppen von Testpersonen belegen dies, dass die natürlichen Skalen weniger anfällig sind für situationsbedingte oder absichtliche Verfälschungen seitens der Testperson.

## Interpretation von »Diskrepanzen« basierend auf einem Vergleich von adaptiver und natürlicher Dimension

Die Konstruktion der zwei Dimensionen (adaptiv/natürlich) verortet die adaptiven Skalen in der bekannten, öffentlichen Sphäre und die natürlichen Skalen in der verborgenen (privaten) Sphäre. Die adaptive Dimension ist situationsbezogen, das heißt, die Testperson bekennt sich offen zu einer Verhaltensweise. Die vier adaptiven Skalen sind folglich anfälliger für Erwartungen und/oder erwünschte Verhaltensweisen.

Auf der Basis dieser Konstruktion lässt sich ein Unterschied zwischen den Prozenträngen von adaptiver bzw. natürlicher Skala eine Spannung oder auch Differenz zwischen der erwünschten Eigenwahrnehmung oder Verhaltenserwartung und dem von der natürlichen Skala gemessenen unbewussten Verhaltensmerkmal interpretieren. Aufgrund psychologischer und gesellschaftlicher Faktoren weisen viele Testpersonen einen bestimmten Pegel an »normaler« Differenz auf.

Die Effekte dieser beiden Faktoren lassen sich jedoch minimieren, indem statistisch Kriterien als Anhaltspunkte dafür definiert werden, wann eine Differenz substantiell und wann sie normal und ohne Konsequenzen für den Probanden ist.

Wir beginnen unsere Überlegungen zu dieser Definition von Differenz (Spannung) mit der Definition von mindestens vier möglichen Situationen:

- In einer ersten Situation liegt die adaptive Skala über der natürlichen Skala. In diesem Fall sind die situationsbezogenen Erwartungen hinsichtlich der Eigenwahrnehmung stärker ausgeprägt als das innere Merkmal. Eine solche Situation kann als »Performancespannung« bezeichnet werden.
- In einer zweiten Situation liegt die natürliche Skala über der adaptiven Skala. In diesem Fall kommen die inneren Merkmale nicht voll zum Ausdruck. Ihr Ausdruck wird von situationsbezogenen Erwartungen in Schach gehalten oder kontrolliert. Diese Situation kann als »Kontrollspannung« bezeichnet werden.
- Eine dritte Situation ist ein Mittelding zwischen beidem, wo die Unterschiede zwischen den Dimensionen nicht substantiell sind.
- Eine vierte Situation besteht dann, wenn die Unterschiede keine Spannung für die Testperson, sondern lediglich eine künstliche (absichtliche oder unabsichtliche, bewusste oder unbewusste) Manipulation repräsentieren, die die Ergebnisse des Instruments ungültig machen

Die Konstruktion eines Spannungsmaßes gründet auf den Unterschieden zwischen den adaptiven und natürlichen Skalen für die vier DISC-Verhaltensmerkmale. Weil die Rohergebnisse dieser Merkmale einmalige und nichtstandardisierte Metriken verwenden, besteht der erste Schritt darin, sie in die eine oder andere bekannte Metrik zu übersetzen. Weil das **INSIGHTS MDI**<sup>®</sup>-Instrument Prozentränge für die Diagrammdarstellung der Skalen verwendet, bieten sich diese als logische Vergleichsmetrik an. Prozentränge gehen von 0 bis 100, mit einem Median bei 50. Die Spannungsmetrik verwendet folglich den Prozentrang auf einer adaptiven Skala abzüglich des Prozentrangs auf der zugehörigen natürlichen Skala. Das Ergebnis ist ein Spannungswert für jeden Einzelfall mit einem durchschnittlichen Erwartungswert über alle Fälle bei 0. Eine Untersuchung der beobachteten Spannungswerte ergibt, dass diese im Wesentlichen der Normalverteilung entsprechen.

Unter der Annahme einer Normalverteilung definiert eine Standardabweichung von  $\pm 1$  vom Mittelwert einen Bereich, in dem 68 Prozent der Fälle liegen. Er ist ein solider Bereich für das, was wir als normale oder übliche Spannungswerte bezeichnen können.

Unter der Annahme einer Normalverteilung lässt sich für jeden Wertebereich eine Fallprozentzahl schätzen. So liegen über 95 Prozent aller Fälle innerhalb des Bereichs von  $\pm 2$  Standardabweichungen. In 27 Prozent der Fälle, in denen der Wert im Bereich zwischen  $\pm 1$  und  $\pm 2$  Standardabweichung liegt, lässt sich eine Spannung beobachten. Wenn der Spannungswert positiv ist (adaptiv > natürlich), was auf 13,5 Prozent der Fälle zutrifft, können wir von einer »Performancespannung« sprechen. Wenn der Spannungswert negativ ist, bietet sich die Bezeichnung »verborgene Spannung« an.

Ferner liegen in 0,3 Prozent die Antworten von Testpersonen im Bereich jenseits  $\pm 3$  Standardabweichungen. Fälle mit einem derart hohen Spannungsbereich sollten sorgfältig auf absichtliche Manipulationen seitens der Testpersonen hin untersucht werden.

Bedenken Sie dabei, dass sich diese Kriterien auf die Verteilung der Spannungswerte in der Referenzpopulation und nicht auf die klinischen oder Verhaltensbeobachtungen hinsichtlich möglicher substanzieller Spannungsbereiche bei einzelnen Testpersonen beziehen. Sie sollten folglich lediglich als Quelle für Fragen und Diskussionen und nicht als definitive Diagnose Verwendung finden. Viele Testpersonen funktionieren sehr gut unter hohen Stress- und Spannungspegeln.

Die vier DISC-Spannungsvergleiche haben jeweils (laut Definition) einen Mittelwert von 0, aber verschiedene Standardabweichungen. Folglich muss jede Interpretation der Spannung die skalenspezifische Standardabweichung der Stresswerte für jede Referenzpopulation berücksichtigen. Die Statistiker müssen somit ein spezifisches Kriterium entwickeln, um zu beurteilen, welche Unterschiede bei den einzelnen Skalen signifikant und substanziell sind. Solche statistischen Vorgehensweisen helfen dem Praktiker in der Anwendung jedoch nur bedingt. Praktiker, die sich für Spannungsvergleiche interessieren, erhalten auf Anfrage von TTI spezifische Standardabweichungswerte für die Spannungswerte in der jeweils verwendeten Referenzpopulation. Zudem muss jede Interpretation den Unterschied zwischen der Verwendung statistischer Kriterien und den möglichen psychologischen Effekten (z.B. Verfälschungsabsichten) berücksichtigen.

### 3.3. Validität und Reliabilität

#### Antwortenbasierte Reliabilität und interne Struktur

Die erste Frage, die gestellt wird, wenn jemand wissen will, wie »gut« ein Instrument ist oder ob es überhaupt von Nutzen ist, betrifft die Reliabilität des Instruments. Das Wort »Reliabilität« meint im Zusammenhang mit Instrumenten und Tests stets »innere Konsistenz«. Es gibt mehrere Verfahren, die für diese statistischen Routineuntersuchungen üblicherweise zum Einsatz kommen.

**Retest-Reliabilität** bezeichnet die Konsistenz der Ergebnisse, die man erhält, wenn man die gleiche Versuchsperson mit dem identischen Instrument ein weiteres Mal, zu einem späteren Zeitpunkt testet.

Zum Nachweis der **Paralleltest-Reliabilität** werden derselben Versuchsperson zwei ähnliche Formen des Instruments vorgelegt. Bei der Retest- ebenso wie bei der Paralleltest-Reliabilität sollte die Dokumentation sowohl den Reliabilitätskoeffizienten als auch die zwischen erster und zweiter Testung verstrichene Zeit festhalten. Der Schwerpunkt beider Verfahren liegt auf der Konsistenz (Zuverlässigkeit) der Messung. Die Frage der Konsistenz sowie der »Lernvorteil« durch die Testwiederholung sind bei Geschicklichkeits- und Wissensmessungen ein entscheidendes Problem. INSIGHTS MDI® kennt keinen Vorteil durch wiederholte Vorgabe des Tests, weil eine Selbsteinschätzung abgefragt wird. Die Skalen des Instruments sind so stabil wie die Wahrnehmung der Situationsanforderungen durch die Versuchsperson, und die Selbstwahrnehmung ist relativ konstant.

**Split-Half-Reliabilität** erfordert lediglich die einmalige Anwendung des Instruments, wobei das Instrument jedoch in zwei Hälften geteilt wird, beispielsweise in die geraden und ungeraden Items, um anschließend die Korrelation zwischen den beiden Ergebnismessungen zu bestimmen. Diese Technik umgeht einige der vom zeitlichen Abstand zwischen den Tests verursachten Schwierigkeiten der Retest- und Paralleltest-Reliabilität.

Auch die **Kuder-Richardson-Reliabilität** basiert auf einer einzigen Form und einer einmaligen Anwendung des Instruments, indem sie die Konsistenz der Antworten auf alle Items des Tests misst. Die Kuder-Richardson-Formel ist das Mittel aus allen Split-Half-Koeffizienten auf der Basis unterschiedlicher Aufteilungen des Tests.

Die Formel der **Spearman-Brown-Reliabilität** ist ein weiteres statistisches Verfahren, das Reliabilitätskoeffizienten liefert und regelmäßig zusammen mit dem Split-Half-Verfahren zur Anwendung kommt. Die Spearman-Brown-Methode zeichnet sich durch die Verdoppelung der Zahl der Items im Instrument als Teil der Formel aus. Die Verdoppelung der Zahl der Items hat in der Regel eine gesteigerte Reliabilität zur Folge. Manche Kritiker der Spearman-Brown-Formel sagen, der Reliabilitätskoeffizient des Tests werde so künstlich angehoben. Alle bisher diskutierten Reliabilitätskoeffizienten lassen sich mit einem geeigneten Statistikprogramm oder, bei Kenntnis der Formeln mit einem einfachen Taschenrechner ermitteln.

Der Alpha-Koeffizient bringt die Reliabilität eines Instruments zum Ausdruck und rangiert von 0 bis +1,00. Ein Instrument mit perfekter Reliabilität hätte einen Alpha-Koeffizienten von +1,00, und kein Instrument für psychologische Diagnostik hat diesen Wert bislang erreicht. Überdies existiert kein Standard, kein anerkannter Wert, ab dem eine Korrelation für Testzwecke als gut oder schlecht zu gelten hat.

Man hat sich jedoch allgemein darauf verständigt, als Minimalstandard ein Alpha von mindestens 0,6 anzusetzen, wobei einige Experten für 0,7 oder einen höheren Wert plädieren. Je höher der Alpha-Koeffizient, desto stärker ist ganz offensichtlich die Kohärenz der Items. Cronbachs Alpha (Cronbach, 1951) ist nach Einschätzung vieler das bislang robusteste Reliabilitäts-Alpha (Anastazi, 1976; Reynolds, 1994). »Koeffizient Alpha ist der maximale Wahrscheinlichkeitsschätzwert des Reliabilitätskoeffizienten, wenn das Parallelmodell als wahr angenommen wird« (SPSS, S. 873). Für dichotome Daten »ist Cronbachs Alpha äquivalent mit dem Koeffizienten der Kuder-Richardson-Formel 20 (KR20)« (SPSS, S. 873).

Die Reliabilitätskoeffizienten des **INSIGHTS MDI®**-Instruments werden mittels Cronbachs Alpha bestimmt. Der Leser möge die in diesem Handbuch präsentierten Reliabilitätskoeffizienten mit den Reliabilitäten anderer Instrumente vergleichen und auch nachfragen, wie andere Anbieter ihre Alpha-Zahlen errechnen.

## Kontextbasierte Validität und Beziehungen zu anderen Variablen

Die Validität liefert eine Antwort auf die Frage: »Misst das Instrument, was es messen soll?« Sie stellt zudem eine tiefer qualitätsbezogene Frage: »Wie gut misst das Instrument diese Dinge?« Diese Fragen sind offensichtlich schwieriger zu beantworten und lassen unter Umständen Raum für Subjektivität. Bei allen Fragen zur Validität kommt es entscheidend auf die Beziehung zwischen der Performance im Instrument und anderen beobachtbaren Fakten bezüglich des betrachteten Verhaltens an. Wenn jemand sagt: »Der Test war nicht fair«, dann zielt der Kommentar in der Regel auf die Validität und nicht die Reliabilität des Tests. Zutreffender könnte man formulieren: »Der Test war nicht valide.« Es gibt drei wesentliche Formen der Validität: Inhaltliche, kriterienbezogene und Konstruktvalidität.

Die **inhaltliche Validität** untersucht den Inhalt des Instruments daraufhin, ob er das zu messende Verhalten abdeckt. Eine einfache Überprüfung der Items zum Beispiel in einem Biologie- oder Chemietest sollte Fragen zutage fördern, die in Bezug zum Lernstoff stehen. Im Zusammenhang mit der Entwicklung der DISC-Themen bedeutet inhaltliche Validität, dass alle vier Deskriptorkategorien in ungefähr gleichen Proportionen in der Auswahl von D-, I-, S-, und C-Deskriptoren repräsentiert sind. Darüber hinaus umfasst die inhaltliche Validität die Überprüfung der sozialen Erwünschtheit. Sobald ein Ungleichgewicht besteht zwischen den Deskriptoren, die gesellschaftlich akzeptiert, und solchen, die gesellschaftlich weniger akzeptiert sind, tangiert dies die inhaltliche Validität in unerwünschter Weise. Das Style-Insights(R)-Instrument wird kontinuierlich auf inhaltliche Validität hin überprüft, mit der Folge, dass seit der ersten Drucklegung einige Deskriptoren zugunsten der inhaltliche Validität und der Reliabilität des Instruments ersetzt wurden.

Die **kriterienbezogene Validität** bezeichnet die Fähigkeit eines Instruments, das Verhalten eines Teilnehmers in bestimmten Zukunftssituationen vorherzusagen. Die Ergebnisse des Teilnehmers im Instrument werden mit diversen externen »Kriterien« verglichen. Im Rahmen von **INSIGHTS MDI®**-Instrument und -Berichten liegen diverse Studien von Success Insights und TTI Performance Systems vor, die eine klare Beziehung zwischen bestimmten Verhaltensweisen oder Verhaltensmustern und beruflichem Erfolg in spezifischen, klar definierten Bereichen herstellen (Bonnstetter u.a., 1993).

**Kriterienbezogene Validität** hat zwei Formen: konkurrente Validität und prädiktive Validität. Die konkurrente Validität vergleicht die Ergebnisse eines Teilnehmers mit externen Kriterien zum Zeitpunkt der Verabreichung des Instruments. Die prädiktive Validität vergleicht die Ergebnisse im Instrument mit Kriterien nach einem genauer definierten Zeitintervall. Beide Methoden liefern robuste Unterstützung für das **INSIGHTS MDI®**-Instrument und seine Berichte (Bonnstetter et al., 1993).

**Konstruktvalidität** untersucht die Fähigkeit eines Instruments, ein theoretisches Konstrukt oder Merkmal zu messen. Die Konstruktvalidität basiert auf einem Evidenzmuster und diversen Maßen aus einer Vielzahl von Quellen. Zu den in der Verhaltensmerkmalanalyse betrachteten Konstrukten gehören beispielsweise Entwicklungsveränderungen von Teilnehmern, die in unterschiedlichen Lebensjahren, in unterschiedlichen Lebensphasen oder aus unterschiedlichen Perspektiven den Fragebogen beantworten. Die Korrelation mit anderen Tests ist ebenfalls eine Form der Konstruktvalidierung. Es gibt diverse Vergleiche zwischen dem **INSIGHTS MDI®**-Instrument und anderen Verhaltensinstrumenten wie dem MBTI (Myers Briggs Type Indicator), MMPI (Minnesota Multiphase Personality Inventory), 16-PF (16 Personality Factor) und anderen Instrumenten. Alle diese Vergleichsstudien bestätigen die Gesamtkonstrukte dieser Instrumente.

Eine sehr wichtige Technik innerhalb der Konstruktvalidierung ist die Faktorenanalyse. Diese Technik »verfeinert« ein Instrument, indem es seine Daten miteinander vergleicht und Beziehungen zwischen ihnen analysiert. Das Ziel ist es dabei, aus sämtlichen Anfangskombinationen eine kleinere Zahl von Faktoren oder gemeinsamen Merkmalen »herauszudestillieren«. Faktorenanalytische Untersuchungen unter Heranziehung anderer Instrumente haben ergeben, dass Instrumente einiger anderer Anbieter spezifische Deskriptoren haben, die sich auf Faktoren in anderen Kategorien auswirken als die, zu denen sie in dem Instrument gezählt werden (Golden, Sawacki und Franzen, 1990). Auch das **INSIGHTS MDI®**-Instrument ist im Zuge der Faktorenanalyse verfeinert worden, um die Gesamtvalidität und -reliabilität von Instrument und Berichten zu erhöhen (siehe »Untersuchung der theoretischen Kohärenz« Seite 24).

Eine Anmerkung zur **Augenscheinvalidität**: Augenscheinvalidität darf nicht mit den oben erwähnten Validitätsformen verwechselt werden. Augenscheinvalidität ist keine richtige Validität im technischen bzw. statistischen Sinne des Wortes. Die Augenscheinvalidität prüft, ob das Instrument in den Augen der Teilnehmer »valide erscheint«, und ob der Report für den Leser »valide klingt«. Während Augenscheinvalidität technisch keine Form der Validität ist, spielt sie dennoch eine wichtige Rolle sowohl bei der Instrumentenentwicklung als auch bei der Berichtabfassung. Einer der Autoren (Watson) hat das **INSIGHTS MDI®**-Instrument und seine Berichte für Untersuchungen an Tausenden von Teilnehmern verwendet. In Veranstaltungen mit Zuhörerschaften beliebiger Größe werden die Teilnehmer gefragt: »Zu wie viel Prozent scheint ihr Report für Sie zutreffend zu sein?«. Darauf antworteten die überwiegende Mehrzahl der Befragten: Zu 85 bis 90%. Diese Frage zielt auf die Augenscheinvalidität. Wenngleich diese Methode keine statistische Formel verwendet, hat sie dennoch das Konzept der Augenscheinvalidität im Visier. Diese Augenscheinvalidität weist auf die sehr hohe Akzeptanz des Reportergebnissen des Instruments hin.

### Konvergente und divergente Evidenz

Die Validitätsprüfung umfasst noch zwei weitere Punkte. Beide betreffen im Prinzip die Frage, ob eine Klassifizierung vermittels des Instruments ähnliche Testpersonen korrekt identifiziert (konvergent) und hinreichend zwischen Testpersonen unterscheidet, die zu unterschiedlichen Klassifikationen gehören (divergent). Auch hier beruht die Macht der Evidenz im Wesentlichen auf den fundierten Anwendungserfahrungen der das Instrument verwendenden Berater. Im Zuge der erstmaligen Überprüfung des **INSIGHTS MDI®**-Instruments im Jahr 2002 bot sich die Gelegenheit, Stichproben von Teilnehmern aus unterschiedlichen Beschäftigungsbereichen und Branchen zu vergleichen. Vier Datenmengen stammten von Testpersonen, die als Beschäftigte/mittleren Management, höheren Management, Verkäufer beziehungsweise Kundendienstmitarbeiter klassifiziert waren. Aus insgesamt 120898 Probanden wurde eine zufällige Stichprobe von 1076 Fällen ausgewählt. Unter der Annahme, dass es sich bei den acht Skalen um unabhängige Variablen handelt, wurde ein Diskriminanzmodell errechnet.

Die Anwendung dieses Klassifikationsmodells lieferte in 42 Prozent der Fälle Übereinstimmung mit der jeweiligen Datenquelle. Am oberen Ende der korrekten Klassifikation wurden 75 Prozent der Testpersonen aus dem Kundendienstbereich zutreffend klassifiziert. Am unteren Ende wurden 30 Prozent der höheren Manager zutreffend klassifiziert. In Anbetracht der Vielfalt innerhalb jeder Teilnehmerklassifikation und des Wechsels von Testpersonen zwischen unterschiedlichen Klassifikationen belegen diese Resultate eine substantielle Fähigkeit, Menschen anhand ihrer **INSIGHTS MDI®**-Skalen zu unterscheiden und zu klassifizieren.

## 3.4. Situation und Umfeld

Alle DISC-Instrumente konzentrieren sich auf Verhaltensweisen, und solche Verhaltensweisen spielen sich im öffentlichen Umfeld ab. Folglich kommen die zu messenden psychologischen Merkmale durch Interaktionen zum Ausdruck. Im Unterschied zur Messung rein interner Überzeugungen und Werte werden Verhaltensweisen durch diese Interaktionen gefördert oder zurückgedrängt. In diesem Zusammenhang tauchen drei Fragen hinsichtlich der Messung von Eigenschaften auf, die in dieser Weise Ausdruck finden. Eine Frage betrifft die Situationsanforderungen und die individuelle Wahrnehmung des Umfelds, in dem die Messung stattfindet. Eine zweite Frage dreht sich um die vom Teilnehmer wahrgenommene sozialen Erwünschtheit der als Indikatoren der vier DISC-Skalen verwendeten Deskriptoren. Die dritte Frage entsteht, sobald Deskriptoren übersetzt und das Instrument in einem anderen sprachlich-kulturellen Umfeld eingesetzt wird. Auf alle drei Fragen wollen wir im Folgenden eingehen.



## Situationsanforderungen

Situationsanforderungen ergeben sich aus dem Umfeld und seinen Bedingungen, wie sie der Einzelne wahrnimmt. Da unser Instrument auf Selbsteinschätzungen von Verhaltensweisen basiert, kann ein Proband seine Antwortentscheidungen je nach bewusster oder auch unbewusster Wahrnehmung der Situation, in der das Instrument eingesetzt wird, verfälschen. Wenn der Proband die Situation als risikoarm wahrnimmt, äußert er sich möglicherweise offener (im Sinne von ehrlicher) als in einem als bedrohlich empfundenen Umfeld. Wenn er hingegen die Situation als hochgradig bedrohlich empfindet, ist er möglicherweise weniger offen, um auf diese Weise ein positiveres Bild von sich zu vermitteln. Das ist ein Problem, das alle auf Selbsteinschätzungen basierenden (ipsativen) Instrumente gleichermaßen betrifft, siehe dazu auch Seite 22.

Die Wirkung solcher Situationswahrnehmungen fließt am ehesten in die Auswahl von Beschreibungen wie »passt am besten zu mir« ein. Wird hingegen nach dem gefragt, was »am wenigsten zu mir passt«, oder sollen Beschreibungen nach der geringsten Attraktivität geordnet werden, hat es der Proband schwerer, seine Antworten zu »korrigieren«.) Zum Vergleich der auf »Passt am besten zu mir« und »Passt am wenigsten zu mir« basierenden Skalen siehe Seite 13. Dennoch stehen auch diese »Passt-am-wenigsten-zu-mir« Antworten immer noch unter dem Einfluss der gesellschaftlich motivierten Vermeidungsstrategien, von denen der nächste Absatz handelt.

## Soziale Erwünschtheit

Nicht alle Beschreibungen oder Eigenschaften einer Skala sind als skalierbare Items in einem Instrument geeignet. Es gibt Beschreibungen, die gesellschaftlich als wünschenswert oder unerwünscht vorgelegt sind. Manche Beschreibungen einer Skala können also gesellschaftlich erwünscht sein. Wer eine Skaleneigenschaft in stärkerem Maße besitzt, handelt vielleicht häufiger in dieser gesellschaftlich gewünschten Art und Weise. Aber jeder möchte sich gern in diesem positiven Licht sehen. Ausgetauschtes Itembeispiel für sozial erwünscht: respektvoll, Respekt zeigend. In heutiger Version lautet diese Item anders. Wenn also diese Beschreibung zur Auswahl steht und viele Probanden sie wählen, erschwert das die Unterscheidung zwischen Probanden, bei denen dieses Skalenverhalten stark, mittel oder schwach ausgeprägt ist. Ebenso kann es Verhaltensbeschreibungen geben, die viele Menschen als »passt am wenigsten zu mir« bezeichnen würden, selbst wenn diese Beschreibung auf ihr Verhalten durchaus zutrifft. Beiderlei Arten von Beschreibungen eignen sich nicht zur Differenzierung der Probanden und werden deshalb im Instrument nicht verwendet. Diese Wertvorstellungen von erwünschten und unerwünschten Verhaltensweisen sind gesellschaftlich geprägt und werden von den meisten Mitgliedern einer Kultur getragen. Die Kultur beeinflusst Auswahl und Beschreibungen noch in anderer Weise, wie der folgende Abschnitt zeigt.

## Kultureller Einfluss

Auch wenn in einer Population viele Kulturen und Subkulturen vorhanden sein mögen, verwenden die Versionen des **INSIGHTS MDI®**-Instrumentes Sprachgruppen als Differenzierungsebene. Kulturen unterscheiden sich in der Art, wie bestimmte Verhaltensweisen definiert und gewertet werden. Wer eine andere Kultur besucht, bemerkt solche Unterschiede mitunter sofort. Lautes gleichzeitiges Sprechen mag in der einen Kultur für gute Freundschaft stehen und in einer anderen auf einen unmittelbar bevorstehenden Kampf hindeuten. Die Beschreibung einer Verhaltensweise mit ähnlichen Worten in zwei unterschiedlichen Sprachen kann sehr verschiedene Konnotationen haben. So können beispielsweise Solidarität und Mitleid unterschiedliche Konnotationen hinsichtlich der Rolle von Gleichheit und Sympathie transportieren.

Diese Unterschiede gilt es beim Einsatz eines Instruments in unterschiedlichen Kulturen zu berücksichtigen. Mit Blick auf diese Unterschiede werden für diverse Sprachgruppen je eigene Versionen von **INSIGHTS MDI®** entwickelt, validiert und getestet. Die im Instrument als Items verwendeten Beschreibungen werden für jede Sprachversion auf Reliabilität und Kohärenz mit den Skalenkonzepten hin getestet. Wenn das Instrument in hinreichendem Umfang zum Einsatz kommt und die Kunden es für wichtig erachten, können für jede definierbare Subpopulation je eigene Verteilungen und Normen errechnet werden.

### 3.5. Messbarkeitsgrenzen

Wird die Testperson aufgefordert, eine Selbsteinschätzung durch Auswahl aus einer begrenzten Zahl von Skalenoptionen abzugeben, spricht man von einem ipsativen Maß. Bei jedem solchen Maß gilt es, bestimmte Beschränkungen zu beachten. Drei Punkte sind beim vorliegenden Maß von besonderem Interesse. Sie betreffen die Art der Ergebnismessung, einen möglichen Selbsteinschätzungs-Bias und die Auswirkungen von Situationsanforderungen und Wahrnehmungen der Ergebnisse.

#### Häufigkeitszählungen und Ergebnis

Erstens resultiert aus der Aufsummierung der Häufigkeit von Antworten ein vergleichendes und kein quantitatives Maß. Ein Ergebnis (Messung) besteht aus der Zahl der vom Teilnehmer ausgewählten Beschreibungen. Diese Zahl wird mit den entsprechenden Zahlen weiterer Versuchspersonen aus der Referenzpopulation verglichen. Die Rohdaten von verschiedenen Skalen können nicht unmittelbar miteinander verglichen werden. Wählt jemand 10 x-Items und 5 y-Items, heißt das nicht, dass er mehr x ist. Wenn aber in der Referenzpopulation der Durchschnitt bei 5 x-Items und 7 y-Items liegt, kann man von dem Teilnehmer, der 10 x-Items auswählt, mit Grund behaupten, dass er sich mehr nach x-Art verhält, als dies in der Population allgemein zu erwarten wäre. Im vorliegenden Instrument wird der Vergleich angestellt zwischen den individuellen Rohdaten und einem Durchschnittswert der Normpopulation. Dabei gilt es zu beachten, dass die Skalen keine quantitativen Aussagen über die Eigenschaften zulassen.

Diese Vergleiche basieren darauf, dass die Referenzpopulation repräsentativ ist für Menschen wie jene, die sich des Instruments zwecks Feedbacks bedienen. Im vorliegenden Instrument sind die Vergleichsnormen repräsentativ für die gegenwärtigen Anwender des Instruments. Wo immer möglich, wurden für die einzelnen Sprach-/Kulturgruppen spezifische Normen entwickelt. Jede für eine Version des Instruments als Referenz verwendete Normverteilung wird klar angegeben.

#### Ipsative Messverfahren

Forced-Choice-Selbstaussagen als Quelle für die Verhaltensskalierung haben in der Psychometrie eine lange Tradition. Solche Messverfahren sind akzeptiert als Methode, um auf Selbstwahrnehmung basierende Einsichten zu gewinnen; dennoch haben sie ihre Grenzen. Ipsative Messverfahren haben statistische Grenzen, die den Einsatz bestimmter Verfahren beschränken. Ipsative Messverfahren unterliegen zudem einem möglichen bewussten oder unbewussten Bias.

Ein unbewusster Bias liegt beispielsweise dann vor, wenn wir uns selbst nicht so sehen, wie andere uns sehen. Unsere Selbstwahrnehmungen, mögen sie auch auf Feedback von anderen basieren, stimmen möglicherweise nicht mit dem überein, wie andere uns beschreiben würden. In der Teambildung und in der Personalentwicklung lohnt unter Umständen die Mühe, sich diese Nichtübereinstimmung bewusst zu machen und mit geeigneten Maßnahmen gegenzusteuern.

### Effekte situationsbezogener Erwartungen

Ein Beispiel für einen bewussten Bias liegt vor, wenn eine Testperson überzeugt ist, dass andere in einer konkreten Situation bestimmte Verhaltensmerkmale wertschätzen, und folglich versucht, sich in diesem Sinne von ihrer besten Seite zu zeigen. Situationsbezogene Erwartungen können auch zu Rollenspannungen führen. Anzeichen dafür liefert unter Umständen ein Vergleich der adaptiven und natürlichen Prozenträge.

## 4. Technische Informationen

### 4.1. Überprüfung und Überarbeitung

Target Training International (TTI) und Success Insights (SI) starteten im Frühjahr 2002 eine Überarbeitung ihres Style-Analysis-DISC-Instruments. Das Hauptaugenmerk lag bei dieser Überarbeitung auf der Skalierung und der Item-Reliabilität für die als Indikatoren für die Konstruktion der vier DISK-Skalen verwendeten 96 Beschreibungen.

Skalenreliabilität und Kohäsion der Items mit ihren zugehörigen Skalen wurden anhand von Stichproben aus den entsprechenden Subpopulationen für jede der fünf englischsprachigen Versionen, die deutsche und die ungarische Version untersucht. Dabei wurden die fünf englischsprachigen Subpopulationen getrennt und, wo sich dies anbot, gemeinsam untersucht. Das selbe Vorgehen wurde auf die deutsche Version angewandt. Die Überprüfung und nachfolgende Überarbeitung erfolgte für jede Sprachversion unabhängig voneinander. Im Folgenden werden die zwecks Überprüfung und Überarbeitung der Itemreliabilität und der Skalenkonstruktionen unternommenen Schritte beschrieben. Spezielle technische Informationen zu den einzelnen Sprachversionen finden Sie im Anhang ab Seite 28.

Alle untersuchten Beispiele stammten von Verhaltensstilanalyse-Fragebögen, die während des Jahres vor der Überprüfung (2001-2002) ausgefüllt wurden. In den meisten Stichproben überstieg die Zahl der verfügbaren Fälle die für die statistische Testung und Auswertung erforderliche Anzahl. Eine ausreichende Anzahl von Datensätzen wurde aus dieser größeren Datenmenge ausgewählt. Im Retest-Verfahren konnten die Schlussfolgerungen und Parameter dann bestätigt und bekräftigt werden. Derartige Vergleiche bestätigten allgemeine Muster psychologischer Eigenschaften mit signifikanten Unterschieden bezüglich der Art, wie bestimmte Indikatoren (Wörter, Ideen) in unterschiedlichen Sprach- und Kulturgruppen miteinander verbunden sind.

Die meisten statistischen Verfahren erfordern für die Überprüfung keine großen Fallzahlen. Für die meisten statistischen Evaluationen wurden folglich Zufallsstichproben aus den entsprechenden Subpopulationen ausgewählt. Mit Hilfe der Stichproben konnten Hypothesen entwickelt werden, die sich anschließend anhand von weiteren Stichproben, die unabhängig von den ersten waren, überprüfen ließen. Diese Testung diente häufig dazu, Überarbeitungsempfehlungen zu bestätigen.

Bei der Überprüfung der Kohärenz der DISC-Skalen wurden zwei Ansätze verfolgt. Der erste Ansatz bestand darin, nach Mustern gemeinsamer Varianz zu suchen. Dabei ging es um die Frage, ob Antworten Kohärenzmuster erkennen ließen, die die theoretische Konstruktion der Skalen rechtfertigen. Siehe dazu den folgenden Abschnitt »Überprüfung der theoretischen Kohärenz«. Ein zweiter Ansatz schaute mit Blick auf die Matrix der Skalenkonstruktion auf die Kohärenz der einzelnen Items mit den zugehörigen Skalen und auf die Gesamtreliabilität dieser Skalenkonstruktion.

## Untersuchung der theoretischen Kohärenz

Die Konstruktion einer Skala beginnt mit der Übersetzung theoretischer Konstrukte in praktische Messverfahren. Im Fall der DISC-Instrumente tauchte diese Konstruktvalidität bereits in der Veröffentlichung der frühen Instrumente auf. Zwecks Bestätigung der Kohärenz der einer Skala zugeordneten Beschreibungen wurde eine Stichprobe von Antworten mittels Hauptkomponentenanalyse (Principle Component Factor Analysis) untersucht. Bei diesem statistischen Verfahren wurden die 96 Items auf Muster ähnlicher Variation hin untersucht. Jeder Faktor ist ein latentes Konstrukt, ein nicht gemessenes Merkmal. Ergebnis des Verfahrens ist eine Liste von Faktoren zusammen mit einem Kovarianzmaß für jede Variable. Diese Koeffizienten können positiv, negativ oder neutral sein. Indem wir Items mit signifikanten positiven oder negativen Koeffizienten bezüglich eines Faktors wählen, identifizieren wir eine Itemkonstellation, die einen latenten Faktor beschreibt. Häufig spiegelt ein Faktor zwei kontrastierende Itemmengen wider. Eine Eigenschaft lässt sich unter den Items mit positiven Koeffizienten und eine zweite unter den Items mit negativen Koeffizienten finden. Wenn die Itemliste mit der Liste jener Items übereinstimmt, die einer Skala theoretisch zugeordnet sind, können wir darauf schließen, dass die Implementierung der Theorie als Skala wohlbegründet ist. Wenn ein Item einen starken positiven Koeffizienten mit anderen Items hat, die zu einer Skala gehören, der es nicht zugeordnet ist, müssen Theorie und/oder Item auf den Prüfstand gestellt werden. Die meisten Items harmonierten mit ihren zugehörigen Skalen. Die häufigste Anomalie sind jedoch Items, die mit keiner Skala einen starken positiven Koeffizienten aufweisen. In diesem Fall ist das Item kein brauchbarer Indikator für die Messung einer Eigenschaft, selbst wenn es sich um eine gute Beschreibung handelt. Dieses Paradox wurde im Abschnitt zur »Sozialen Erwünschtheit« erklärt.

Eine Hauptkomponentenanalyse der Items aus den unterschiedlichen Subpopulationen untersuchte die Kontinuität der konstruierten Skalen. Auf der Grundlage einer Analyse aller dieser Ergebnisse wurden alle von Items überprüft und eine begrenzte Zahl überarbeitet und anschließend im Feldversuch getestet. Im Zuge der Überarbeitung wurden neue Beschreibungen gefunden, die theoretische und linguistische Überlegungen mit Blick auf den aktuellen Sprachgebrauch und die Minimierung eines durch soziale Erwünschtheit bedingten Bias berücksichtigten.

Jede überarbeitete **INSIGHTS MDI®**-Version wurde vor ihrer Freigabe mehreren Runden des Feldversuchs, der erneuten Überarbeitung und der Bestätigung dieser Überarbeitungen unterworfen. Auch hier wurden wieder mehrere unterschiedliche Testpopulationen verwendet. Die aktuelle Version V3 weist eine gesteigerte Reliabilität in sämtlichen Skalen auf und verbesserte die Unabhängigkeit zwischen den S- und C-Skalen.

Die Skalenüberarbeitung ließ eine Überarbeitung der Populationsverteilung geboten erscheinen. Dieser Prozess fokussierte den Referenzpunkt für Normvergleiche der ursprünglichen Konstruktionsgrundlage des Instruments hin zum 21. Jahrhundert mit der Berücksichtigung sich verändernder Verhaltensweisen und neu hinzugekommener gesellschaftlicher Erwartungen.

## Item- und Skalenreliabilitäten

Skalenreliabilitäten wurden mittels des Cronbachschen Alphas berechnet. Cronbachs Alpha gilt als der geeignetste statistische Reliabilitätstest auf der Basis der Antworten, mit denen die Skalen konstruiert wurden. Für dichotome Daten ist Cronbachs Alpha äquivalent zum Koeffizienten der Kuder-Richardson-Formel 20 (KR20). Diese Kennzahlen modellieren die interne Konsistenz auf der Grundlage der durchschnittlichen Korrelation zwischen den Items. Dieses Evaluationsverfahren ist strenger als das traditionelle Split-Half-Verfahren.<sup>1</sup> Cronbachs Alpha ist eine Kenngröße zwischen 0 und 1. Im Allgemeinen gilt 0,6 als niedrigster akzeptabler Wert, wobei in Deutschland namhafte Testautoren für einen strengeren Standard von mindestens 0,7 plädieren.

Da dieses Handbuch für mehreren Sprachversionen von **INSIGHTS MDI®** erscheint, sind in den Zusammenfassungen der einzelnen Sprachversionen spezifische Reliabilitätskennzahlen aufgeführt. Die meisten der 96 Items dienen der Skalenkonstruktion für beide Dimensionen. Einige wenige Items sind nur einer der beiden Dimensionen zugeordnet. Spezielle Koeffizienten für einzelne Items (Indikatoren) mit ihrer zugehörigen Skala werden nicht angegeben, weil es sich dabei um proprietäre Informationen handelt, die den Schlüssel zur spezifischen Auswertung und Itemanalyse des Instruments darstellen.

## 4.2. Normen und Populationsparameter

Der Ursprung der gegenwärtigen Versionen von **INSIGHTS MDI®** basiert auf einer Vielzahl von Evaluationen mit diversen Datenquellen und -stichproben. Im Rahmen der im Jahr 2002 begonnenen Überprüfung früherer Versionen wurden über einhunderttausend Antwortbögen ausgewertet. Die gegenwärtige Item- und Skalenreliabilität ist das Resultat dieser wiederholten Evaluationen anhand verschiedener Stichproben. Der Ursprung des Instruments wird durch diese in regelmäßigen Abständen durchgeführten unabhängigen Evaluationen gestärkt. Die Stichproben stammen jeweils aktuellen (1 bis 2 Jahre alten) Datensätzen von Teilnehmern, die das Instrument genutzt haben. Diese Nutzer repräsentieren ein breites Spektrum. Die spezifischen Stichprobenbeschreibungen finden sich auf den technischen Informationsblättern zu jeder veröffentlichten Version im Anhang.

---

1 Wenn sich Varianz über eine Reihe von Items und genügend untersuchte Fälle erstreckt, ist Cronbachs Alpha mit einer traditionellen Split-Half-Statistik identisch. Ein Problem beim Vergleich von Cronbachs Alpha mit Split-Half-Berechnungen stellen die Korrekturen für eine größere Zahl von Items dar, die die Split-Half-Statistik aufblähen. Während solche Korrekturen im Rahmen der Messung von Fähigkeiten und Wissen gerechtfertigt sind, ist bei ipsativen Messungen Vorsicht geboten. Cronbachs Alpha lässt sich als Durchschnitt aller möglichen Resultate beliebiger Split-Half-Permutationen auffassen. Wir weisen ausdrücklich darauf hin, dass ein Vergleich korrigierter Split-Half-Statistiken mit Cronbachs Alpha statistisch ungültig und unzweckmäßig ist.

### 4.3. Geschlechterverteilung

Eine Thematik die jedes Fragebogen-Instrument betrifft, das im 21. Jahrhundert für Kunden im geschäftlichen und auch im individuellen Kontext entworfen wird, ist der Einfluss des Geschlechts auf die Antwortmuster. Bei der Evaluation des Instruments ging es unter anderem um Unterschiede in den Antwortmustern von Männern und Frauen. Wie nicht anders zu erwarten, gibt es gewisse Unterschiede in den durchschnittlichen Skalenresultaten für Männer und Frauen. Diese Unterschiede deuten auf vergleichsweise geringe Verschiebungen in der Dominanz bestimmter Verhaltensausrprägungen hin. Ob diese Unterschiede biologisch oder gesellschaftlich bedingt oder beides sind, ist für die Effektivität des Instruments nicht relevant. Wichtig ist, dass die Messverfahren des Instruments ein Feedback liefern das keinen Geschlechter-Bias beinhaltet. Die für die Zusammenstellung der Verteilungsnormen verwendeten Stichproben wurden daraufhin evaluiert. Wenn in einer Stichprobe Männer und Frauen proportional vertreten waren wurden keine Korrekturen vorgenommen. Wenn die Verteilung Männer zu Frauen in der Datenstichprobe disproportional war wurden entsprechende statistische Korrekturen so vorgenommen, dass die Auswirkung der unterschiedlichen Antwortmuster von Frauen und Männern ausgeglichen wurden. Diese Korrekturen sind in den technischen Informationsblättern zu den einzelnen Versionen.

### 4.4. Sprachversionen

INSIGHTS MDI® ist in mehreren Sprachversionen verfügbar. Technische Zusammenfassungen dieser als getrennt evaluierte Instrumente entwickelten Versionen finden sich im Anhang. Im Rahmen einer solchen Entwicklung werden die aus der englischen Version übersetzten Itembeschreibungen auf ihre Kohärenz mit ihrer zugehörigen Skala hin analysiert und die Reliabilitäten jener Skalen bewertet. Als Ergebnis dieses Prozesses wurden Items überarbeitet. Eine getrennte Auswertung und Überarbeitung der Skalen resultiert in einem Instrument, das reliabel und für die anvisierte Sprach-/Kulturgruppe geeignet ist. Spezifische Verteilungsnormen für die Sprachversion werden auf der Basis der Antworten auf diese Sprachversion errechnet, um den Nutzern eine Ergebnismessung in Reportform zu liefern, das für die Sprach-/Kulturgruppe, die das Instrument nutzt, relevant ist.

## 5. Instrumentenprotokolle und Verwendung

### Administrierung und Anwendung

Eine Schulung in der Anwendung und Interpretation wird für die Anwender angeboten und für die Interpretation der Reports vorausgesetzt.



## Skalierung & Diagramme

Die unmittelbare grafische Darstellung der Skalenergebnisse erfolgt mittels zweier Diagramme. Ein Diagramm zeigt die adaptiven, das zweite die natürlichen DISC-Ergebnisse. Die Diagrammdarstellung der Skalenergebnisse erfolgt unter Bezugnahme auf eine in einer Normierungspopulation beobachtete Prozentrangverteilung der Skalenergebnisse. Zwei wichtige Themen gilt es bei dieser Art von Eigenberichten und Vergleichen zu beachten.

Erstens sind die vier Skalen relativ zueinander. Höhere Eigeneinschätzungen bei Beschreibungen, die einer bestimmten Skala zugeordnet sind, vermindern mögliche Zuordnungen zu den anderen drei Skalen. Die Skalenergebnisse gehen also in das relative Zuordnungsniveau der vier Verhaltensstile und nicht in eine Quantität oder Häufigkeit dieses Stils ein. Zweitens betreffen die Prozentvergleiche die Häufigkeit der Zuordnung der Testpersonen zu dem beschriebenen Verhalten und die kumulative Häufigkeit solcher Eigeneinschätzungen in der Normstichprobe. Der Vergleich der einzelnen Testperson zur Normpopulation erfolgt also mit Bezug auf die Eigeneinschätzungen von relativen Zuordnungsniveaus.

Während die Skalendiagramme anhand der neuesten Stichproben aktualisiert wurden, ist das Verfahren der Prozentrangverteilung noch immer dasselbe wie im ursprünglichen Instrument. Die Prozentrangverteilung ist eine »universelle« und keine »normale« Verteilung. Das heißt, die Fälle sind nicht gemäß einer »normalen« Glockenkurve verteilt, sondern bilden eine flachere Verteilung mit annähernd gleichen Zahlen bei jedem Skalenpunkt. Aus diesem Grund tauchen die selten erscheinenden Rohergebnisse geballt im oberen oder unteren Bereich der Diagramme auf.

Diese Diagramme stellen Vergleiche zwischen den vier Skalen dar. Die Skalen sind keine unabhängigen Maße. Parallel zur Zunahme der Eigeneinschätzungen (Auswahl von beschreibenden Items) bezüglich einer Skala reduzieren sich dadurch eine oder mehrere andere Skalen. Im Zuge der Überarbeitungen der aktuellen Version wurden diese Vergleiche die ursprünglich aus einer Normpopulation aus Englisch sprechenden, überwiegend weißen Männern der 1960er und 1970er Jahre stammen auf die für aktuelle Nutzer repräsentative Vergleiche aktualisiert.

Die einzelnen Skalen spiegeln keine Quantität jenes Verhaltenstyps wider, sondern einen prozentualen Anteil von Menschen in der Normgruppe, deren Antworten ein geringeres Ergebnis liefern als die der Testperson. Ein im Diagramm dargestelltes Ergebnis von 90 Prozent D bedeutet also, dass rund 90 Prozent der Normierungspopulation ein relativ zu den übrigen Verhaltensbeschreibungen niedrigeres Niveau des Verhaltens vom Typ D zu Protokoll gaben. Es heißt nicht, dass 90 Prozent des Verhaltens der Testperson vom Typ D ist. Nach derselben Logik bedeutet ein Skalenergebnis von 10 Prozent S, dass die Zuordnung der Testperson zu Verhaltensbeschreibungen vom Typ S (oder ihre Wahl auf Rang 4 oder »Passt am wenigsten zu mir«) höher war als bei 10 Prozent der Normstichprobe.

Unter Berücksichtigung von Ähnlichkeiten der Testpersonen in den Ergebnissen bedeutet dies auch, dass rund 78 Prozent der Normstichprobe eine höhere Zuordnung zum S-Verhalten zu Protokoll gab und 12 Prozent der Normstichprobe ein ähnliches Niveau von S-Beschreibungen wählte wie die Testperson. Die letzten zwei Zahlen lassen sich dem Diagramm nicht so einfach entnehmen, aber es ist wichtig zu wissen, dass eine signifikante Zahl von Personen ähnliche Ergebnisse erzielt.

## Report

Dieses technische Handbuch behandelt die DISC-Skalenkonstruktion, jüngste Überarbeitungen und die Darstellung der Skalenergebnisse vermittels zweier Skalendiagramme mit jeweils vier Säulen. Die vom System produzierte Reports enthalten Textabschnitte und die diagrammbasierte Interpretationen. Wer mehr Informationen zu den diversen Textabschnitten und zum Aufbau des **INSIGHTS MDI®** Reports wünscht, kann sich an Target Training International, Success Insights oder ihre autorisierten Lizenznehmer wenden, um spezifische Informationen über Reportvarianten und Feedback-Formate zu erhalten.

## Feedback und Interpretation

Ein Training in der Anwendung und für die Interpretation wird bei Bedarf durch **INSIGHTS MDI®** Deutschland GmbH und durch TTI (für Nutzer in den USA oder ihre autorisierten Lizenzhalter) vermittelt. Weitere Interpretationsrelevante Informationen finden sich in dem Buch: *The Universal Language DISC – A Reference Manual* von »Bill« J. Bonnstetter, Judy I. Suiter und Randy J. Widrick, erscheinen bei Target Training International, Ltd.

Weitere Informationen zu Zertifizierungsmöglichkeiten sind erhältlich für Interessierte, die sich mit der Anwendung von **INSIGHTS MDI®** und den anderen Instrumenten vertraut machen möchten.

# 6. Anhang

## 6.1. INSIGHTS MDI®: Deutsche Version 2021

### Hintergrund

Die regelmäßig vorgenommene Aktualisierung der Normen, zuletzt im Januar 2014 (siehe auch nachfolgende Seiten Beschreibung **INSIGHTS MDI®**: Deutsche Version 2014) ist ein wichtiger Bestandteil des laufenden kontinuierlichen Verbesserungsprozesses der TTI SUCCESS INSIGHTS Analysen. So wurden auch im Mai 2021 von TTI unter Verwendung der deutschsprachigen Datensätze erneut Reliabilitätsstudien durchgeführt. Die Auswertung erfolgte auf Basis der statistischen Berechnungen von Dr. Eric Gehrig, Ph. D. in Mathematics und VP of Research & Development bei TTI SUCCESS INSIGHTS. Diese Auswertungen wurden von TTI für alle unterschiedlichen Sprachversionen des Verhaltensstilanalyse Instruments darunter auch das deutschsprachige Instrument **INSIGHTS MDI®** vorgenommen.

Die Daten der Befragten wurden aus dem TTI SI Internet Delivery System (IDS)<sup>®</sup> ausgewählt und bestehen hauptsächlich aus der allgemeinen Geschäfts- oder berufstätigen Erwachsenenbevölkerung der einzelnen Sprachen. Für die Studie wird eine Zufallsstichprobe von Männern und Frauen aus jeder Sprache gesammelt. Die aktuelle Studie berücksichtigt Daten aus den Jahren 2016 bis 2020.

Basierend auf den Ergebnissen dieser Studie kann man schließen, dass die Einschätzungen der **INSIGHTS MDI®** Auswertungen als intern konsistent und zuverlässig bestätigt werden.

## Normierungsstichprobe

Die verwendete Stichprobe umfasste 49.176 Fälle, wobei die Geschlechterverteilung unter den Probanden genau gleich war. Die Anzahl der weiblichen Testpersonen betrug 24.588, die der

Frauen	Adaptiver Stil				Basisstil			
	D	I	S	G	D	I	S	G
Mittelwert	39	57	60	52	40	61	61	49
Anzahl	24.588	24.588	24.588	24.588	24.588	24.588	24.588	24.588
Standardabweichung	10,0	8,4	9,2	8,0	10,2	7,9	7,3	8,3

männlichen Testpersonen 24.588.

Männer	Adaptiver Stil				Basisstil			
	D	I	S	G	D	I	S	G
Mittelwert	49	50	53	55	51	53	54	50
Anzahl	24.588	24.588	24.588	24.588	24.588	24.588	24.588	24.588
Standardabweichung	10,7	7,9	9,0	8,2	9,6	8,1	7,5	8,1

## Skalenreliabilität

Zur Überprüfung der Skalenzuverlässigkeit wurde die Cronbach Alpha Methode angewandt. Es stellt die geeignetste Maßzahl für die interne Konsistenz einer Skala dar und bezeichnet das Ausmaß, in dem die Items bzw. Fragen einer Skala miteinander in Beziehung stehen. Deshalb wird Cronbachs Alpha auch als Maß der internen Konsistenz einer Skala bezeichnet.

Die durchgeführte Überprüfung ergab robuste Cronbach Reliabilitätskennzahlen der einzelnen Skalen. Die angestrebten Werte von alpha 0,7 werden von jeder Skala erreicht, sie liegen in sehr zufriedenstellenden Bereichen zwischen 0,85 bis 0,91.

Somit ist gewährleistet, dass die Skalen des **INSIGHTS MDI®** Instruments sehr zuverlässig messen.

D Skala	0,91
I Skala	0,85
S Skala	0,87
G Skala	0,86

## 6.2. INSIGHTS MDI®: Deutsche Version 2014

### Hintergrund

Die bis dato durchgeführten Überprüfungen der Reliabilitäten der beiden Dimensionen (Adaptierter und natürlicher Verhaltensstil) wurden von TTI SUCCESS INSIGHTS unter Verwendung der deutschsprachigen Datensätze durchgeführt. Die regelmäßig vorgenommene Aktualisierung der Normen, zuletzt im September 2009 (siehe auch nachfolgende Seiten Beschreibung DISC: Deutsche Version 2009.h) erfolgte auf Basis der statistischen Berechnungen von Peter T. Klassen, Ph. D. Dieser Auftrag wurde von TTI für alle unterschiedlichen Sprachversionen des Verhaltensstilanalyse Instruments darunter auch das deutschsprachige Instrument INSIGHTS MDI® vergeben.

Im Zuge der Qualitätsinitiative der **INSIGHTS MDI International® Deutschland GmbH** wurde die Berechnung und Überprüfung der aktuellen Reliabilitätskennzahlen für INSIGHTS MDI® als Auftrag extern vergeben. Dies sollte eine unabhängige und neutrale Überprüfung durch einen Statistikexperten sicherstellen.

Der Auftrag wurde von Prof. Dr. Karlheinz Zwerenz in München ausgeführt. Im Januar 2014 wurden die benötigten Datensätze in anonymisierter Form als Rohdaten zur Verfügung gestellt. Hauptanliegen war die von TTI bereitgestellten Reliabilitätskennzahlen zu überprüfen (siehe Stellungnahme Prof. Dr. Zwerenz, Seite 30) und aktuelle Zahlen zur Nutzung zur Verfügung zu haben.

### Normierungsstichprobe

Die verwendete Stichprobe umfasste 31.994 Fälle, wobei die Geschlechterverteilung unter den Probanden nahezu ausgeglichen war. Die Anzahl der weiblichen Testpersonen betrug 15.686, die der

Frauen	Adaptiver Stil				Basisstil			
	D	I	S	G	D	I	S	G
Mittelwert	16,35	18,9	21,63	15,12	16,35	16,35	16,35	16,35
Anzahl	15.668	15.668	15.668	15.668	15.668	15.668	15.668	15.668
Standardabweichung	10,158	8,633	9,267	7,845	10,096	7,904	7,484	8,32

männlichen Testpersonen 16.308.

Männer	Adaptiver Stil				Basisstil			
	D	I	S	G	D	I	S	G
Mittelwert	20,9	16,21	18,65	16,24	14,98	18,45	16,96	21,61
Anzahl	16.308	16.308	16.308	16.308	16.308	16.308	16.308	16.308
Standardabweichung	10,759	8,466	9,272	8,05	10,131	8,204	7,683	8,299

## Skalenreliabilität

Zur Überprüfung der Skalenzuverlässigkeit wurde die Cronbach Alpha Methode angewandt. Es stellt die geeignetste Maßzahl für die interne Konsistenz einer Skala dar und bezeichnet das Ausmaß, in dem die Items bzw. Fragen einer Skala miteinander in Beziehung stehen. Deshalb wird Cronbachs Alpha auch als Maß der internen Konsistenz einer Skala bezeichnet.

Die durchgeführte unabhängige Überprüfung durch Prof. Dr. Karlheinz Zwerenz ergab robuste Cronbach Reliabilitätskennzahlen der einzelnen Skalen. Die angestrebten Werte von alpha 0,7 werden von jeder Skala erreicht, sie liegen in sehr zufriedenstellenden Bereichen zwischen 0,850 bis 0,916.

Somit ist gewährleistet, dass die Skalen des INSIGHTS MDI® Instruments sehr zuverlässig messen.

D Skala	0,916
I Skala	0,850
S Skala	0,876
G Skala	0,854

## Stellungnahme Prof. Dr. Karlheinz Zwerenz

AMW Institut, Perhamer Str. 18, 80687 München

Herrn  
Frank Scheelen  
Vorstandsvorsitzender  
Scheelen AG  
Klettgaustraße 21  
79761 Waldshut-Tiengen



München, den 17.1.2014

**Berechnungen von Cronbachs Alpha für die Modelle R4 und PIAV2 analog TTI-SUCCESS**

**STELLUNGNAHME**

Cronbachs Alpha ist eine Maßzahl für die interne Konsistenz einer Skala. Der Wertebereich von Alpha reicht von minus unendlich bis 1. Skalen sollten nur dann Verwendung finden, wenn Alpha mindestens den Wert von 0,65 erreicht.

Bei allen 10 überprüften Skalen konnten mit SPSS die Berechnungsergebnisse von TTI-SUCCESS (bis auf geringfügige Differenzen von maximal +/-0,003) bestätigt werden. Sämtliche Alphas liegen über 0,65 und bestätigen daher die Konsistenz der Skalen.

Diese Aussagen gelten sowohl für die Berechnungen auf Basis der Rohdaten als auch für die Ergebnisse, die auf Basis standardisierter Werte ermittelt wurden.

Für Rückfragen – auch wegen Einsatzmöglichkeiten der Faktorenanalyse – stehe ich Ihnen gerne zur Verfügung.

Viele Grüße



Prof. Dr. Karlheinz Zwerenz  
AMW Institut München

Anlage: Excel-Tabelle mit Berechnungsergebnissen

---

AMW Institut, Perhamer Str. 18, 80687 München,  
Prof. Dr. Karlheinz Zwerenz, Tel.: 089 / 54644285 Fax: 089 / 58927660

## 6.3. INSIGHTS MDI®: Deutsche Version 2009.h

### Zusammenfassung

Diese Auswertung setzt einen im Jahr 2002 begonnenen Qualitätsverbesserungsprozess fort. Diese Überprüfung von 2009 bestätigte anhand einer Reihe von Auswertungen von Skalenkohärenz und Itemreliabilitäten an diversen Testpopulationen die Reliabilität des deutschsprachigen INSIGHTS MDI®-2009.h-Instruments. Die folgende Auswertung von INSIGHTS MDI® basiert auf Stichproben aus einer Population von 18 856 deutschsprachigen Antworten. Diese Datensätze stammen von Testpersonen mit unterschiedlichen beruflichen Hintergrund, die dieses Instrument beantwortet haben.

Die Ergebnisse der Auswertung des 2009.h-Instruments belegen die Reliabilität für die zwei Dimensionen (adaptiv und natürlich) der vier parallelen Skalen (D, I, S, C) mit Cronbachschen Alpha-Werten zwischen 0,819 und 0,97.

Die Ergebnisse der Skalen zeigen eine Normalverteilung über einen breiten Bereich von Skalenpunkten und stützen damit Vergleiche zwischen Testpersonen und den selbstberichteten Verhaltensweisen in der deutschsprachigen Stichprobe.

Die Korrelationen zwischen adaptiven und natürlichen Skalen belegen, dass diese zwei Dimensionen paralleler Skalen, wie zu erwarten, hochgradig korreliert sind, aber auch, dass die Skalen hinreichend unabhängige Maße sind, um separate Interpretationen und Vergleiche zu rechtfertigen.

Diese Version des Instruments umfasst eine Aktualisierung der Populationsverteilungen. Diese Normen aktualisieren die Referenzpopulation, so dass sie repräsentativ ist für deutschsprachige Nutzer im beruflichen Kontext des 21. Jahrhunderts. Das deutschsprachige Instrument INSIGHTS MDI® 2009.h ist ein starkes, zuverlässig messendes Instrument, das sich auf eine Vielzahl von Berufsgruppen anwenden lässt.

### Hintergrund

Das Instrument INSIGHTS MDI® Deutsche Version 2009.h von Target Training International baut auf dem Design und der Entwicklung seiner DISC-Vorgängerversionen auf. Im Jahr 2002 startete Target Training International, Ltd., einen fortlaufenden Prozess der kontinuierlichen Überprüfung der Reliabilität der vom Unternehmen angebotenen Instrumente.

Es folgt eine Zusammenfassung der Ergebnisse einer Auswertung der deutschsprachigen INSIGHTS MDI®-2008-Daten. Diese Daten umfassen Informationen von 18 856 Teilnehmern. 37 Prozent der Antworten stammen von Frauen, und 63 Prozent von Männern. Dieser Unterschied in den Teilnahmeraten wird für die Berechnung der Referenznormen und -verteilungen in der weiter unten beschriebenen Art und Weise korrigiert.

Diese Daten wurden in 2 Datensätzen übermittelt. Die Gruppen wurden verglichen. Die Durchschnittswerte dieser beiden Gruppen wurden auf signifikante Unterschiede hin getestet. Wie zu erwarten, führt die Verwendung kompletter Datengruppen aufgrund der großen Zahl der Datensätze zu statistisch signifikanten Unterschieden bei allen Skalen. Ein anschließend durchgeführter t-Test für unabhängige Stichproben, angewendet auf zwei kleinere

Stichproben ( $\approx 1000$ ), zeigte, dass lediglich die (adaptiven und natürlichen) I-Skalen sich auf dem 0,001-Niveau signifikant unterschieden. Aufgrund dieses Ergebnisses wurde entschieden, dass die beiden Gruppen ähnlich genug waren, um kombiniert zu werden. Von den kombinierten Datensätzen darf angenommen werden, dass sie für eine größere Population repräsentativ sind als die separate Verwendung einer der beiden Subpopulationen.

Das **INSIGHTS MDI®**-Instrument umfasst 96 Adjektive die in 24 Rahmen (Fragen) à vier Items organisierte Beschreibungen sind. In jedem Rahmen findet sich für jedes der vier Skalenkonstrukte ein zugehöriges beschreibendes Item (Aussage). Die Testpersonen ordnen diese Items entlang der Schiene »Passt am besten zu mir« bis »Passt am wenigsten zu mir«. Aus diesen Antworten werden je zwei Dimensionen der vier Skalen konstruiert. Dabei handelt es sich um die adaptive und die natürliche Dimension. Die Bedeutung und Nützlichkeit der zwei Skalendimensionen wird weiter unten diskutiert. Die Entwicklung des deutschsprachigen **INSIGHTS MDI®**-Instruments erfolgte unabhängig von der Entwicklung der übrigen Sprachversionen. Die deutschsprachigen Items wurden auf der Basis des gegenwärtigen deutschsprachigen kulturellen und gesellschaftlichen Kontextes überprüft und evaluiert.

Die Skalenreliabilitäten und die Kohäsion der Items zu ihren zugehörigen Skalen wurde anhand zufällig gezogener Stichproben aus den Testpopulationen überprüft. Im Zuge früherer Auswertungen bestätigte eine Faktorenanalyse der Items die Konsistenz der Skalen. Auf der Basis einer Analyse all dieser Indikatoren wurde eine begrenzte Zahl von Items überprüft, überarbeitet und im Feldversuch getestet. Die Itemüberprüfungen basierten auf der theoretischen Konstruktion von Items in Kombination mit linguistischen Überlegungen, die sich auf den modernen Sprachgebrauch und die Minimierung eines sozialen Erwünschtheits-Bias konzentrierten. Dieser Prozess fokussierte den Referenzpunkt für Normvergleiche der ursprünglichen Konstruktionsgrundlage des Instruments hin zum 21. Jahrhundert mit der Berücksichtigung sich verändernder Verhaltensweisen und neu hinzugekommener gesellschaftlicher Erwartungen.

## Referenzpopulation

Der Ursprung des deutschsprachigen Instruments umfasst eine Vielzahl von Datenquellen und Stichproben. Die aktuelle Itemkohärenz und Skalenreliabilität ist das Ergebnis mehrerer Evaluationsrunden unter Verwendung unterschiedlicher unabhängiger Stichproben. Die für die Deutsche Version 2009.h **INSIGHTS MDI®** verwendeten Normen basieren auf 18.856 Antworten von Nutzern des Instruments. Die Stichprobe repräsentiert die unterschiedlichen Testpersonen, die das Instrument in einer Vielzahl von Situationen nutzen.

37 Prozent dieser Antworten stammen von Frauen und 67 Prozent von Männern. Bei der Berechnung dieser Referenznormen und -verteilungen wird der Unterschied in den Teilnehmeraten mittels etablierter statistischer Verfahren zur Stichprobengewichtung ausgeglichen.



	adapt. D	natürl. D	adapt. I	natürl. I	adapt. S	natürl. S	adapt. C	natürl. C
Erwartungswert	18,24	-17,92	18,59	-16,03	20,29	-15,59	14,88	-22,46
Std.-Fehler des EW	0,076	0,072	0,061	0,058	0,066	0,045	0,057	0,060
Median	17	-17	18	-15	20	-15	14	-22
Standardabw.	10,433	9,901	8,382	7,987	9,011	7,458	7,850	8,234
<b>Geschlecht: weiblich</b>								
Erwartungswert	16,03	-20,17	20,01	-14,57	21,76	-14,46	14,20	-22,81
Std.-Fehler des EW	0,102	0,102	0,087	0,078	0,094	0,076	0,078	0,085
Median	15	-20	20	-14	22	-14	14	-23
Standardabw.	9,883	9,924	8,432	7,612	9,093	7,361	7,563	8,227
<b>Geschlecht: männlich</b>								
Erwartungswert	20,44	-15,68	17,16	-17,49	18,83	-16,73	15,57	-22,10
Std.-Fehler des EW	0,108	0,096	0,083	0,083	0,089	0,076	0,083	0,085
Median	20	-15	17	-17	18	-17	15	-22
Standardabw.	10,504	9,356	8,085	8,085	8,686	7,382	8,070	8,227

INSIGHTS MDI®, German 2009: N = 18.856

## Skalenreliabilität

Die Skalenreliabilitäten wurden mittels Cronbachs Alpha berechnet. Cronbachs Alpha gilt als der am besten geeignete statistische Reliabilitätstest. Diese Statistiken modellieren die interne Konsistenz auf der Basis der durchschnittlichen Korrelation zwischen den Items. Cronbachs Alpha ist ein statistischer Wert zwischen 0 und 1. Im Allgemeinen gilt ein Alpha von 0,6 als minimal akzeptables Niveau, auch wenn einige deutsche Testautoren für 0,7 oder einen höheren Wert plädieren.

Die folgende Tabelle fasst die Reliabilitäten gemäß Cronbachschem Alpha für die 2009.f-Daten zusammen. Die Skalenreliabilitäten haben sich gegenüber früheren Versionen verbessert. Diese Ergebnisse bestätigen die Überarbeitungen und dokumentieren die kohärente Skalenkonstruktion und die Reliabilität des deutschsprachigen Instruments.

adapt. D	natürl. D	adapt. I	natürl. I	adapt. S	natürl. S	adapt. C	natürl. C
0,897	0,820	0,857	0,825	0,884	0,825	0,819	0,831

Cronbachs alpha, German 2009.h N = 18.856

## Skalenbeziehungen – Korrelationen

Eine Untersuchung der Beziehung zwischen den Skalen konzentriert sich auf zwei Punkte. Erstens bestätigt die Beziehung zwischen den adaptiven und natürlichen Skalen ein theoretisches Fundament. Während manch einer argumentieren mag, dass die DISC-Skalen dadurch gestärkt werden, dass diese beiden Dimensionen schlicht kombiniert werden, stützt eine Untersuchung der folgenden Korrelationstabelle die These, dass diese beiden Dimensionen einen feinen aber wichtigen Unterschied messen. Die Korrelationen zwischen gleichskaligen adaptiven und natürlichen Werten schwanken um 0,9. Aufgrund von in sämtlichen Testpopulationen gemachten Beobachtungen und Vergleichen zwischen diversen unabhängigen Versionen von DISC-Instrumenten bestehen starke Argumente für die These, dass die natürlichen Skalen weniger anfällig sind für einen durch soziale Erwünschtheit bedingten Bias und Variationen aufgrund von Situation, Umfeld und Erwartungen der Testpersonen.

Ein zweiter Punkt, der Beachtung verdient, ist die starke Unabhängigkeit zwischen den S- und C-Skalen. Aufgrund von Beobachtung und Logik lässt sich die These aufstellen, dass die adaptiven Skalen ein Maß für die Situation sind und unter dem Einfluss des Einschätzungs-Bias stehen. Das heißt, die Testperson wird tendenziell eher Merkmale an sich erkennen, die ihren Wünschen und Hoffnungen entspricht, so sehr sie sich auch bemüht, eine solche Verzerrung zu vermeiden. Die natürlichen Skalen sind konstruktionsbedingt tendenziell weniger anfällig für diesen Selbsteinschätzungs-Bias. Mit anderen Worten, es fällt der Testperson schwerer, die natürlichen Skalen zu manipulieren.

	adapt. D	natürl. D	adapt. I	natürl. I	adapt. S	natürl. S	adapt. C
natürl. D	0,939	1					
adapt. I	-0,060	0,012	1				
natürl. I	-0,013	-0,013	0,928	1			
adapt. S	-0,829	-0,813	-0,206	-0,178	1		
natürl. S	-0,811	-0,812	-0,215	-0,233	0,941	1	
adapt. C	-0,326	-0,339	-0,740	-0,758	0,186	0,237	1
natürl. C	-0,372	-0,444	-0,726	-0,750	0,291	0,289	0,929

Leere Zellen verweisen auf nichtsignifikante Korrelationen

## Schlussfolgerungen

Ich bin in Bezug auf das deutschsprachige **INSIGHTS MDI® 2009.h** anhand einer Analyse von Antwortmustern aus einer vielgestaltigen Population von Testpersonen zu folgenden Schlussfolgerungen gekommen:

- Die aktuellen Daten bestätigen Itemkohärenz und Skalenreliabilität. Die acht Skalen werden ausweislich eines Cronbachschen Alphas größer 0,8 mit einer soliden Reliabilität konstruiert.
- Die Items sind kohärent rund um konsistente Konstruktindikatoren der von den Instrumentbeschreibungen repräsentierten theoretischen Konzepte.
- Die Skalenergebnisse sind normalverteilt mit ausreichend Varianz über alle Skalenpunkte, um Interpretationen und Vergleiche zwischen Testpersonen zuzulassen, wenn sie als Vergleiche von Einzelpersonen mit Verteilungen dieser selbstberichteten Verhaltensweisen in einer Population interpretiert werden.
- Die adaptiven und natürlichen Dimensionen umfassen parallele Skalen, die miteinander konsistent sind, aber sie repräsentieren auch ein Potential für eine sinnvolle komplementäre Interpretation.
- Das Instrument wird in aktuellen Populationen referenziert und die Messungen erfolgen auf Basis von aktuellen Normgruppen-Vergleichen somit im 21. Jahrhundert.

Diese Auswertungen des **INSIGHTS MDI®**-Instruments bestätigen das Bemühen von TTI um eine kontinuierliche Qualitätsverbesserung.

### Eingereicht von:

Peter T. Klassen, Ph. D.

Principal, DocumentingExcellence.com

Professor Emeritus, College of DuPage

20. September 2009

## 6.4. INSIGHTS MDI®: Deutsche Version 2006.f10,.f24

### Hintergrund

Am 23. Juni 2005 stellte Success Insights, Inc., sieben Datensätze (Stichproben) mit den Datensätzen von 16.986 Testpersonen vom deutschsprachigen **INSIGHTS MDI®**-Instrument bereit. Die Hauptaufgabe bestand in der Bestätigung der Normen und Verteilungen des deutschsprachigen Instruments. Dazu war es erforderlich, die Daten und die Itemreliabilitäten vergleichend auszuwerten. Während die 2004d-Normen bestätigt wurden, mündeten Analyse und Diskussionen in dem Vorhaben zwei auf die spezialisierte Nutzung fokussierte Anwendungen bereitzustellen. Die Anwendungen verwendeten eine unterschiedliche Sammelmethodik und konzentrierten sich auf unterschiedliche spezifische Populationen von identifizierten Nutzern. Diese Faktoren führten zur Entwicklung und Implementierung der deutschsprachigen Normen 2006.F10 und 2006.F24. Beide verwenden unterschiedliche Skalenkonstruktionsverfahren und Normen von vollkommen unterschiedlichen Referenzpopulationen. Die beiden Anwendungen setzen einen aus der Ursprungsversion von DISC Style Analysis 2 abgeleiteten Qualitätsverbesserungsprozess fort.

## Reliabilität

Auswertungen beider Anwendungen offenbaren, dass die meisten INSIGHTS MDI®-Alpha-Werte im Bereich zwischen 0,7 und 0,8 liegen, mit Ausnahme eines Alpha-Wertes um 0,6.<sup>2</sup> Die Anwendungen 2006.f.10 und 2006.f.24 verwenden unterschiedliche Sammel- und Skalenstrukturierungsmethoden.

Cronbachs $\alpha$	adaptiv				natürlich			
	D	I	S	C	D	I	S	C
2006.f.10	0,817	0,737	0,769	0,719	0,799	0,728	0,643	0,764
2006.f.24	0,870	0,792	0,828	0,789	0,852	0,776	0,795	0,790

Daraus kann man schließen, dass die acht Skalen in jeder der zwei Anwendungen als reliable Implementierung der in früheren INSIGHTS MDI®-Instrumenten identifizierten theoretischen Konstrukte betrachtet werden können.

## Beziehung zwischen den Anwendungen

Die vergleichbaren Skalenkonstruktionen jeder der beiden Anwendungen sind hochgradig korreliert mit Werten zwischen 0,91 und 0,96. Das deutet darauf hin, dass beide Anwendungen mit der Gewissheit eingesetzt werden können, dass die Merkmale sowohl zwischen den beiden Versionen als auch mit früheren Versionen des Instruments konsistent sind.

Spearman-Korrelationen	adaptiv				natürlich			
	f.24 AD	f.24 AI	f.24 AS	f.24 AC	f.24 ND	f.24 NI	f.24 NS	f.24 NC
f.24 AD	0,96	-0,08	-0,76	-0,25	0,84	0,06	-0,69	-0,42
f.24 AI	-0,05	0,94	-0,24	-0,678	0,06	0,76	-0,15	-0,61
f.24 AS	-0,77	-0,19	0,95	0,15	-0,72	-0,25	0,76	0,39
f.24 AC	-0,26	-0,66	0,17	0,94	-0,32	-0,62	0,16	0,76
f.24 ND	0,85	-0,01	-0,72	-0,26	0,95	0,04	-0,71	-0,48
f.24 NI	0,04	0,79	-0,26	-0,66	0,06	0,91	-0,2	-0,66
f.24 NS	-0,74	-0,16	0,82	0,17	-0,75	-0,25	0,91	0,33
f.24 NC	-0,33	-0,61	0,29	0,78	-0,43	-0,65	0,19	0,94

<sup>2</sup> Ein INSIGHTS MDI®-alpha von 0,6 gilt als minimaler Wert, um eine Skala als reliabel einzustufen, wengleich manche Experten einen minimalen Wert von 0,7 ansetzen.

Die Konstruktion der adaptiven und natürlichen Dimensionen der vier DISC-Skalen wird in beiden Anwendungen beibehalten. In der f.24-Anwendung werden kleinere Kontraste zwischen den beiden Dimensionen beobachtet. Das kann Einfluss auf die Interpretationen einer Testperson haben, wenn man die adaptiven und natürlichen Diagramme vergleicht.

Die zwei Anwendungen sind für den Gebrauch mit unterschiedlichen spezifischen Populationen gedacht. Beide verwenden unterschiedliche Antwortverfahren. Die jeweiligen Skalen und Normen basieren auf unterschiedlichen Referenzstichproben. Deshalb sollten die Ergebnisse nicht zwischen den Anwendungen verglichen werden. Allgemeine Muster von Höhen und Tiefen auf den Diagrammkurven ihrer jeweiligen Profile dürfen aber dennoch verglichen werden.

### Schlussfolgerung

Die Veränderungen in den Normen der deutschsprachigen INSIGHTS MDI®-Versionen 2006.f10 und f24 sind kleinerer Art. Die Implementierung von zwei getrennten Anwendungen mit je einer eigenen Referenznorm empfiehlt sich vor allem deshalb, weil jedes Instrument sein eigenes Messverfahren verwendet und mit einer separaten Population verwendet wird. Die Validität einer jeden Anwendung beruht auf ihrer jeweils eigenen Entwicklung. Als separate Anwendungen sind beide solide Instrumente, solange sie mit ihrer jeweiligen zugehörigen Referenzpopulation genutzt werden. Weil aber beide Anwendungen jeweils eine fokussierte Referenznorm verwenden, kann weder die eine noch die andere als repräsentativ für eine allgemeine Population angesehen werden. Zudem sollten Vergleich zwischen den beiden Versionen nur mit größter Zurückhaltung vorgenommen werden.

#### Eingereicht von:

Peter T. Klassen, Ph. D.

Principal, DocumentingExcellence.com

Professor Emeritus, College of DuPage

10. Februar 2006

## 7. Literaturliste

Anastazi, A. (1976). *Psychological Testing*. New York: Macmillan Publishing Co.

Bonnstetter, W. J., Suiter, J. I., & Widrick, R. J. (1993). *The Universal Language DISC: A Reference Manual*. Scottsdale, AZ: Target Training International, Ltd.

Clark, W. (1954). *Activity Vector Analyses*. Self-published.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297-34.

DuBois, P. H. (1970). *A history of psychological testing*. Boston: Allyn and Bacon.

Goldberg, L. R. (1971). A historical survey of personality scales and inventories. In P. McReynolds (Ed.), *Advances in psychological assessment*. Vol. 2, 293-336. Palo Alto, CA: Science and Behavior Books.

Golden, C.J., Sawicki, R.F., & Franzen, M.D. (1990). Test construction. In G. Goldstein & M. Hersen (Eds.) *Handbook of Psychological Assessment*. New York: Pergamon Press.

- Geier, J. G. (1967, December). "A trait approach to the study of leadership in small Groups." *The Journal of Communication*, 17, (4), 316-323.
- Geier, J. G. (1977). *The Personal Profile System*. Minneapolis, MN: Performax Systems, Intl.
- Hartshorne, H. & May, M. A. (1928) *Studies in Deceit*. New York: Macmillan.
- Hartshorne, H., May, M. A., & Maller, J. B. (1929) *Studies in Service and Self-control*. New York: Macmillan.
- Hartshorne, H., May, M. A., & Shuttleworth, F. K. (1930) *Studies in Organization of Character*. New York: Macmillan.
- Jacoby, J. & Matell, M. (1971). Three point Likert scales are good enough. *Journal of Marketing Research*, 8,495-500.
- Kaplan, S. J. (1983). *A Study of the Validity of the Personal Profile System*. Minneapolis, MN: Performax Systems Intl.
- Komorita, S. S. (1963). Attitude content, intensity, and the neutral point on a Likert scale. *Journal of Social Psychology*, 61, 327-334.
- Komorita, S.S. & Graham, K. (1965). Number of scale points and the reliability of scales. *Journal of Educational and Psychological Measurements*, 25,987-995.
- Kraepelin, E. (1892). *Über die Beeinflussung einfacher psychischer Vorgänge durch einige Arzneimittel*. Jenna: Fischer.
- Matell, M. & Jacoby, J. (1972). Is there an optimal number of alternatives for Likert items? *Journal of Applied Psychology*, 56, 506-509.
- Styles Insights® 28 1 July 2011 Instrument Validation Manual Marston, W. M. (1928). *The Emotions of Normal People*. London: Harcourt, Brace & Co.
- Peabody, D. (1962). Two components in bipolar scales: direction and extremeness. *Psychological Review*, 69, 65-73.
- Reynolds, C.R. (1994). Reliability. In R.J. Sternberg (Ed.), *Encyclopedia of Human Intelligence*. New York: Macmillan Publishing Co.
- SPSS. (1988). *SPSS-X User's Guide*, 3rd Ed. Chicago: SPSS Inc.
- Standards for Educational and Psychological Testing, Committee to Develop (c/o American Psychological Association). (1999). *Standards for educational and psychological testing*. Washington D.C.: American Educational Research Association.
- Symonds, P. M. (1931). *Diagnosing personality and conduct*. New York, NY: Century.
- Warburton, D. M. (n.d.). *The Importance of Finding the Real Person at Work*. Reprints available at <http://enlightenedpartners.com/profiles/ValidityStudies.pdf> or [http://www.cgroupinc.com/files/Validity\\_Study.pdf](http://www.cgroupinc.com/files/Validity_Study.pdf) (Retrieved: 3 March 2004).
- Watson, R. J. (1989). *A statistical comparison of the TTI Style Analysis and the Performax Personal Profile System*. Unpublished research report, Department of Education, Wheaton College, Wheaton, IL.





# Driving Forces

## 1. Einführung und Hintergrund

Menschen entwickeln auf der Grundlage ihrer Überzeugungen, Wertvorstellungen und Erfahrungen bestimmte Verhaltensmuster. Diese Muster verändern sich in der Regel kaum mit der Zeit. Sie sind Teil der Persönlichkeit. Es haben sich verschiedene Ansätze entwickelt, wiederkehrende Persönlichkeitsmuster zu messen und zu beschreiben. Daraus entstanden unterschiedliche Theorien. Keine der bestehenden Methoden, Persönlichkeitsmuster zu klassifizieren, kann als die einzig wahre gelten, auch wenn einige Ideen mehr Zuspruch finden als andere.

Verschiedene Kräfte haben zur Entwicklung von Persönlichkeitstheorien beigetragen. Viele Menschen möchten sowohl ihr eigenes Verhalten verstehen als auch das Verhalten anderer Menschen vorhersagen können. Beobachtungen lassen darauf schließen, dass bestimmte Persönlichkeitsmerkmale Rückschlüsse auf das zu erwartende Verhalten zulassen. Ein Persönlichkeitsaspekt, von dem es regelmäßig heißt, dass er unser Verhalten beeinflusst, ist die Motivation. Über einige Motivationsfaktoren, wie beispielsweise Geld, Dienst am Mitmenschen oder utilitaristischer Pragmatismus, besteht weitgehend Einigkeit.

Während die Diskussion zwischen den Verfechtern existierender Persönlichkeitstheorien in vollem Gange ist und lebhafte Debatten geführt werden, wie sich Persönlichkeitsmerkmale messen lassen, werden weiterhin Instrumente zur Persönlichkeitsklassifizierung entwickelt und eingesetzt. Es hat sich eine pragmatische Herangehensweise herausgebildet, die den sorgfältigen Einsatz von Messverfahren unterstützt, welche nicht nur in der Individualpsychologie sondern auch von vielen Arbeitgebern im Rahmen ihrer Einstellungs- und Personalentwicklungsprozesse als eine nützliche Hilfe empfunden werden. Daraus hat sich ein breites Interesse an der Erforschung der für Arbeitsteams sowie Pre-Screening-, Einstellungs- und Auswahlverfahren relevanten Motivationsfaktoren und Wertvorstellungen entwickelt. Dass Werte individuelle Verhaltensweisen anstoßen und fördern, wird von kaum jemandem bestritten. Indem wir die »treibenden Faktoren« hinter den jeweiligen Verhaltensweisen untersuchen, erhalten wir folglich Einblick sowohl in das, was einen Menschen motiviert (seine Wertvorstellungen), als auch in die Mechanismen, wie sich diese Motivation in beobachtbarem Verhalten manifestiert (Verhaltensweisen).

## 2. Grundlagen

Die Themen des Instruments PIW – Persönliche Interessen und Werte® sind den Arbeiten und Schriften Eduard Sprangers entlehnt, der in seinem Hauptwerk Lebensformen (1914/1921) sechs Grundantriebe im Menschen identifizierte: theoretische, ökonomische, ästhetische, soziale, politische und spirituelle.

Das heutige Modell behält einige der Sprangerschen Bezeichnungen bei, während andere erweitert und eine ersetzt wurde. »Theoretisch«, »ästhetisch« und »sozial« wurden übernommen, »ökonomisch« durch »utilitaristisch« sowie »politisch« durch »individualistisch« ergänzt und »spirituell« durch »traditionell« ersetzt, wobei zugleich der Anteil der Antwortmöglichkeiten mit religiösem Bezug reduziert und um solche mit traditionellem oder sinnstiftendem Bezug ergänzt wurde.

Das PIW – Persönliche Interessen und Werte® Modell verwendet in Übereinstimmung mit Sprangers Originalarbeit sechs Wertethemen. Andere Modelle verwenden sieben, acht oder gar bis zu achtzehn solcher Themen. Unter der Annahme, dass Werte in direktem Bezug zu Trieben und Bedürfnissen stehen, liefert die Literatur diverse Kategorisierungen. Freud (1922) kennt zwei, Maslow (1954) schon fünf und Murray (1938) gar 28 solcher Bedürfnisse. Daraus ergibt sich die Frage: Wer hat Recht? Die Antwort lautet: Es gibt keine richtigen und falschen, sondern nur unterschiedliche Theorien. Wissenschaft funktioniert nach dem Negationsprinzip. Das bedeutet nicht, dass Wissenschaft negativ ist, sondern lediglich, dass eine Theorie solange als potenziell gültige Erklärung aufrechterhalten wird, bis sie mit den Mitteln des wissenschaftlichen Diskurses widerlegt wurde. Weil aber die Theorien von Freud, Maslow, Murray und Spranger nicht widerlegt wurden und somit auch heute noch Gültigkeit besitzen, kommen sie alle als mögliche Erklärung der unterschiedlichen Facetten von menschlichem Verhalten in Frage. Alle Wissenschaft, ob Gesellschafts- oder Naturwissenschaft, funktioniert nach diesem Prinzip.

Aufgrund spezifischer Strukturen in den einzelnen Theorien ist es schwierig, Elemente mehrerer Theorien miteinander zu kombinieren. Eine sorgfältig durchgeführte Untersuchung führte zu dem Ergebnis, dass die Orientierung am Sprangerschen Originalmodell mehrere Vorteile mit sich bringt.

- Erstens ist das Modell, abgesehen von einer behutsamen Aktualisierung durch die Umbenennung einzelner Wertethemen, nach wie vor zutreffend.
- Zweitens unterstützt es eine der grundlegenden und am häufigsten angewendeten Theorien aus dem Wertekontext.
- Drittens entsprechen die sechs Wertethemen dieses Modells der Arbeitswirklichkeit in einem breiten Spektrum von Unternehmen, Institutionen und Branchen sowohl aus dem kommerziellindustriellen Sektor als auch aus dem Bereich der Non-Profit-Organisationen, der konfessionellen und Bildungseinrichtungen und der staatlichen Behörden.

Die Entscheidung zugunsten der Sprangerschen Theorie bietet folglich die tragfähigste und flexibelste Grundlage für die Entwicklung dieses Instruments.

Die Textdateien für die Berichte nehmen zudem Bezug auf A Study of Values(1960) von Allport, Vernon und Lindzey, auf Gordon Allports Pattern and Growth in Personality (1961) und auf Milton Rokeachs The Nature of Human Values (1973). Bill Bonnstetter verfasst seit 1984 Texte und Artikel auf der Grundlage des Sprangerschen Modells. Dr. Watson arbeitete über zehn Jahre lang mit großen und kleinen Fokusgruppen, um auf diese Weise die für die Arbeitsstellenanalyse®-Reports erstellten Textbausteine mit individueller und kollektiver Augenscheinvalidität zu versehen. Mit Hilfe dieser Fokusgruppen konnten die Aussagen in den Reports verfeinert und so spezifisch wie möglich an die einzelnen Bewertungssegmente angepasst werden. Auch stärkten sie die Aussagekraft der Textdateien in der Fertigstellungsphase des Instruments.

Der Titel PIW – Persönliche Interessen und Werte® wurde gewählt, weil Werte gelegentlich als die versteckten Motivatoren bezeichnet werden. Unsere Motivatoren treten nur dort zu Tage, wo sie sich in unserem Verhalten manifestieren. Ohne beobachtbares Verhalten oder der Möglichkeit, nach den Gründen für das Tun eines Menschen zu fragen, bleiben unsere Werte größtenteils verborgen. Das in der Sprache des DISC-Modells beschriebene Verhaltensmuster eines Menschen zeigt auf, wie er etwas tut.

Seine Werte aber betreffen die Frage, warum er es tut. Indem wir das Wie und das Warum des Verhaltens eines Menschen und seiner inneren Motivatoren verstehen, sind wir viel besser in der Lage, das Aktivitätsmuster eines Menschen oder eines Teams zu erkennen und zu durchdringen, als wenn wir uns lediglich mit einer dieser Facetten beschäftigen würden.

## 3. Instrumentierung

Die Entwicklung der Struktur eines Instruments erfolgt in drei Phasen. Zuerst entwerfen und operationalisieren die Testentwickler ihre Idee. Als zweites wird das vorläufige Instrument in der Praxis getestet und überarbeitet. Schließlich wird das Instrument freigegeben, wobei die Beobachtungsdaten aus dem Einsatz dazu genutzt werden, um seine Qualität fortlaufend zu verbessern. Die folgenden Abschnitte erläutern die allgemeinen Prozesse der Instrumentenentwicklung. Die konkreten Prozesse für dieses Instrument werden in Abschnitt 4 »Überblick über das Instrument« weiter hinten vorgestellt. Weil sich dieses Handbuch auf diverse Sprach- und Kontextversionen des Instruments bezieht, finden Sie die technischen Informationen zu den für den deutschsprachigen Raum eingesetzten Versionen im Anhang.

### 3.1. Beginn der Entwicklung – Theoretische Validität

Die Entwicklung eines Instruments beginnt mit Ideen, Konzept, vorhandener Theorie und Wissen. Die Entwickler bestimmen zuerst einen oder mehrere Interessenbereiche, indem sie beispielsweise Nischen, ungelöste Probleme oder neue konzeptionelle Denkweisen identifizieren. Daraus ergeben sich ein oder mehrere Zielbereiche. Diese Anfangsideen werden anschließend weiterentwickelt. Dieser Entwicklungsprozess bildet das Fundament der THEORETISCHEN VALIDITÄT. Im Zuge der Ausgestaltung und Implementierung nutzen die Entwickler vorhandene Forschungsergebnisse und Expertenmeinungen, um diese Zielkonzepte weiter zu konkretisieren und zu verfeinern.

Im nächsten Schritt werden diese Zielkonzepte in messbare Größen operationalisiert. In einem Brainstormingprozess werden zum Zielkonzept passende Indikatoren ermittelt. Parallel dazu werden verschiedene psychometrische Strukturen zwecks Messung des Zielkonzepts untersucht.

In dieser Phase der Entwicklung werden Items (Fragen) skizziert, die im Instrument Verwendung finden könnten. Viele Möglichkeiten werden in Betracht gezogen, und häufig werden viel mehr Items skizziert, als für das endgültige Instrument benötigt werden. Ob einzelne Items letztlich zum Ziel und zu den theoretischen Konzepten passen, kann abschließend erst im Rahmen der statistischen Datenanalyse geklärt werden, siehe der Abschnitt »Validierung« auf Seite 10.

### 3.2. Bestimmung einer Messstruktur

Die Bestimmung einer Messstruktur beginnt mit Erwägungen zu den Charakteristiken des Zielkonzepts. Manche Ziele beinhalten Wissensfragen, zu denen es korrekte Antworten und Expertenwissen gibt. Manche Ziele erfordern Fähigkeiten beispielsweise im Bereich des Lernens oder des Lösens von Problemen. Im Zentrum des PIW (Persönliche Interessen und Werte) -Instruments stehen, wie der Name schon sagt, die persönlichen

Antriebskräfte eines Menschen. In dieser Anwendung gilt kein Motivator als grundsätzlich besser oder schlechter. In konkreten Situationen (Arbeit/berufliche Funktion) können jedoch einzelne Motivatoren effektiver oder nachhaltiger sein als andere. Zur Klassifikation der in diesem Test verwendeten Motivatoren siehe der Abschnitt »Skalenstruktur« auf Seite 11.

Die in einem Instrument verwendeten Items können aus Bewertungsfragen oder Sortierlisten bestehen. Bewertungsfragen können eine Likert-Skala mit Antwortmöglichkeiten wie »trifft zu«, »trifft eher zu«, »trifft eher nicht zu« und »trifft nicht zu« enthalten. Ein solches Bewertungsmuster liefert unabhängige Items, kann aber zu Pattsituationen führen, wenn Items gleichbewertet werden.

Alternativ können die Teilnehmer gebeten werden, Items auf einer Skala von 1 bis 10 zu bewerten. Diese Skalenform lässt Rückschlüsse auf relative Werte und auf die Stärke des Bewertungsgefälles zu, da angenommen werden darf, dass zwischen 1 und 10 ein stärkeres Gefälle besteht als zwischen 1 und 3.

Eine dritte Variante erlaubt den Teilnehmern, Items explizit gegeneinander in Relation zu setzen mit Bewertungen wie »mir am ähnlichsten« bis »mir am unähnlichsten«. Alternativ können die Teilnehmer gebeten werden, Einträge einer Liste nach Präferenz oder Attraktivität zu sortieren. Hier wird eine Entscheidung erzwungen. Hier geht es ausschließlich um die Reihenfolge und nicht um absolute Präferenz oder Attraktivität.

Sobald der Entwurf eines Instruments steht, kann es getestet und überarbeitet werden.

### 3.3. Konstruktvalidierung

Die Validierung beginnt mit Feldversuchen. Dabei kann der gesamte Text oder nur einzelne Abschnitte im Blickpunkt stehen. Manchmal werden bei anfänglichen Feldtests die Teilnehmer gebeten, die Fragen durchzulesen, zu beantworten und sich anschließend dazu zu äußern. Früher oder später muss der Teilnehmerkreis hinreichend vergrößert werden, um statistische Rückschlüsse zu ermöglichen. Die Teilnehmer sollten repräsentativ für den Bevölkerungsteil sein, für den das Instrument vorgesehen ist.

Die Evaluation eines Instruments erfolgt auf zwei Ebenen. Dabei werden die Antworten der Teilnehmer auf die einzelnen Items zu Skalen aggregiert, die als Maß für die Zielkonzepte dienen. So werden auf der einen Ebene die einzelnen Items bewertet, während auf der zweiten Ebene die Eignung der Messskalen getestet wird. Beide Ebenen sind gleichzeitig aktiv.

#### Analyse ohne konzeptionelle Beschränkungen – Faktorenanalyse

Ein erstes statistisches Verfahren untersucht Antworten auf alle Items ohne strukturelle Annahmen. Das heißt, die Skalenzuordnungen werden vorläufig außer Acht gelassen. Die Faktorenanalyse liefert mögliche Muster gemeinsamer Kohäsion oder Variation zwischen den Items, wenn also Teilnehmer, die auf bestimmte Items eindeutig positiv reagieren, auch auf andere Items eindeutig positiv (oder negativ) reagieren. Daraus ergeben sich erste Anhaltspunkte, ob die Vorstellungen der Entwickler hinsichtlich der Clusterung bestimmter Indikatoren um ihr Zielkonzept dem Praxistest standhalten.

Auch wenn die Faktorenanalyse kein Allheilmittel ist, liefert sie einige erstaunliche Resultate. Die Faktoren deuten mitunter auf die Existenz von Kohärenzmustern hin, mit denen die Entwickler nicht gerechnet haben. Ein einzelner Faktor spiegelt sich möglicherweise in mehreren, zueinander komplementären Items wider. Im Verlauf der weiteren Evaluationsschritte kann immer wieder auf diese nichtbeschränkten Muster zurückgegriffen werden, um mögliche Probleme mit bestimmten Items zu identifizieren.

### **Analyse zwecks Bestätigung einer vorgegebenen Struktur – Skalenreliabilität**

Die Bestätigung der Kohärenz zwischen den Items einer Skala gehört in den Bereich der Reliabilitätsprüfung. Die Bestimmung der Skalenreliabilität wurde in der Vergangenheit auf unterschiedliche Weise durchgeführt. Eine vergleichende Zusammenstellung dieser Verfahren finden Sie im Abschnitt »Antwortenbasierte Reliabilität und interne Struktur« auf Seite 12. Weil Cronbachs Alpha ( $\alpha$ ) Feedback dazu liefert, wie sich die Gesamtreliabilität einer Skala verändert, wenn ein Item aus der Skala entfernt wird, ist es möglich, Items zu identifizieren, die möglicherweise Korrekturen erfordern. Auf diese Weise können Entwickler die Reliabilität einer Skala maximieren.

### **Analyse der Korrelationen zwischen Skalen – Korrelationen zwischen Skalen**

Mittels Untersuchung der Korrelationen zwischen Skalen können Entwickler beurteilen, ob Skalen voneinander unabhängig oder stark miteinander verbunden sind. Im Idealfall sollten die Skalen sich gegenseitig ausschließen und folglich unabhängig sein. Auf viele Konzepte in unserer realen Welt trifft das jedoch nicht zu. Die Untersuchung von Korrelationen kann auch zeigen, ob Skalen entgegengesetzt sind. Dieser Fall kommt sehr viel häufiger vor. Diese breite Beziehung in einer Referenzpopulation liefert Einsichten in konzeptionelle und theoretische Interpretationen, die die nachträgliche Befragung der Teilnehmer erleichtern kann.

### **Entwicklung von Skalenwerten und Referenznormen**

Die vorangegangenen drei Evaluationsprozesse liefern den Entwicklern Hinweise auf die Gesamtqualität des Instruments und helfen ihnen, Items ausfindig zu machen, die nicht funktionieren. Die Entwicklung eines Instruments erfordert häufig den mehrmaligen Durchlauf dieser Schritte, bis die Standards der Entwickler erfüllt sind. Diese Ergebnisse dienen den Entwicklern als Richtschnur für die Überarbeitung und Korrektur der Items oder für die Entscheidung, dass das Instrument für den nächsten Schritt bereit ist.

Sobald die Items eines Instruments gemäß den Wünschen der Entwickler funktionieren, können die Skalenstrukturen fertiggestellt werden. Wenn die Skalen stehen, muss zuletzt noch festgelegt werden, wie die Skalenwerte zu interpretieren sind. Häufig müssen dazu die rohen Skalenwerte in eine standardisierte oder normalisierte Skala übersetzt werden. Diese normalisierten Skalen beziehen sich stets auf eine Population und nicht auf eine Stichprobenmenge.

## **3.4. Freigabe und Follow-Up – Praxisbestätigung**

Mit der Freigabe des Instruments ist die Arbeit der Entwickler nicht zu Ende. Die Anwendung des Instruments liefert sowohl quantitatives als auch qualitatives Feedback. Dieses Feedback umfasst Einzelfallbeschreibungen zur Effektivität des Instruments. Die regelmäßige Durchsicht der Daten der Teilnehmer ermöglicht die kontinuierliche Überprüfung von Itemkohärenz, Skalenreliabilität und Referenznormen auf der Basis einer sehr viel größeren Population (im Unterschied zur Feldversuchstichprobe).

## 4. Überblick über das Instrument

PIW – Persönliche Interessen und Werte basiert in Design und Entwicklung auf dem früheren TTI-Instrument Personal Interests, Attitudes and Values (PIAV). Im Jahr 2002 startete TTI eine Überprüfung der Reliabilität der sechs Skalen und der zugehörigen Items. Im Rahmen einer Qualitätsüberprüfung wurde die englischsprachige Ausgabe des Instruments mehreren Auswertungszyklen unterworfen. Unabhängige Überprüfungen wurden zudem für die deutsch-, niederländisch- und ungarischsprachigen Ausgaben sowie für eine speziell für britische Jugendliche entworfene Ausgabe des Instruments durchgeführt. Zusammenfassungen dieser Bewertungen finden Sie im Anhang.

### 4.1. Skalenstruktur

Das Instrument PIW – Persönliche Interessen und Werte umfasst zwölf Rahmen mit je sechs Sätzen oder Aussagen. Jeder Satz oder jede Aussage ist ein Indikator für einen von sechs latenten Motivatoren. Die Teilnehmer werden aufgefordert, die Aussagen in eine Reihenfolge von 1 bis 6 zu bringen, wobei 1 für den besten und 6 für den schlechtesten Rang steht. Die Skalen werden konstruiert, indem die Rangfolge umgekehrt, die Ränge zusammenhängender Items aufaddiert und die Punktzahlen nach oben korrigiert werden, um mögliche Nullen zu vermeiden.

Die Skalen haben die Bezeichnungen:

THEORETISCH, ÖKONOMISCH, ÄSTHETISCH, SOZIAL, INDIVIDUALITISCH und TRADITIONELL.

### 4.2. Validität und Reliabilität

Häufig hört man Fragen und Kommentare zur Validität und Reliabilität von Instrumenten. Im Grundsatz bezieht sich Validität auf die Frage, ob ein Instrument oder Item das misst, was es zu messen vorgibt. Für die Testung und Untermauerung der Validität werden diverse Methoden verwendet. Reliabilität bezieht sich auf die Frage, ob ein Instrument oder Item auf zuverlässige Weise misst. Manchmal wird auch darüber diskutiert, ob ein Instrument, das nicht reliabel ist, überhaupt valide sein kann. Wir wollen uns an dieser Diskussion an dieser Stelle nicht beteiligen.

#### Antwortenbasierte Reliabilität und interne Struktur

Die erste Frage, die gestellt wird, wenn jemand wissen will, wie gut ein Instrument ist oder ob es überhaupt von Nutzen ist, betrifft die Reliabilität des Instruments. Das Wort »Reliabilität« meint im Zusammenhang mit Instrumenten und Tests stets »Konsistenz«. Validität basiert auf Kontext und Beziehungen zu anderen Variablen. Es gibt mehrere Verfahren, die für diese statistische Routinebehandlung gemeinhin zum Einsatz kommen.

Retest-Reliabilität bezeichnet die Konsistenz der Ergebnisse, die man erhält, wenn man die gleiche Versuchsperson mit dem identischen Instrument ein weiteres Mal testet.

Zum Nachweis der Paralleltest-Reliabilität werden derselben Versuchsperson zwei ähnliche Formen des Instruments vorgelegt. Bei der Retest- ebenso wie bei der Paralleltest-Reliabilität sollte die Dokumentation sowohl den Reliabilitätskoeffizienten als auch die zwischen erster und zweiter Testung verstrichene Zeit festhalten.

Der Schwerpunkt beider Verfahren liegt auf der Konsistenz der Messung. Die Konsistenzfrage sowie der »Lernvorteil« durch die Testwiederholung sind bei Geschicklichkeits- und Wissensmessungen ein entscheidendes Problem. PIW – Persönliche Interessen und Werte kennt keinen Vorteil durch wiederholte Vorgabe des Tests, weil eine Selbsteinschätzung abgefragt wird. Die Skalen des Instruments sind so stabil wie die Wahrnehmung der Situationsanforderungen durch die Versuchsperson, und die Selbstvorstellung ist relativ konstant. Eine gewisse Abweichung stellen auch wir bei der wiederholten Testung fest, aber die beobachteten Abweichungen sind so klein, dass sie die Gültigkeit des Gesamtergebnisses nicht beeinträchtigen.

Split-Half-Reliabilität erfordert lediglich die einmalige Anwendung des Instruments, wobei das Instrument jedoch in zwei Hälften geteilt wird, beispielsweise in die geraden und ungeraden Items, um anschließend die Korrelation zwischen den beiden Ergebnismengen zu bestimmen. Diese Technik umgeht einige der vom zeitlichen Abstand zwischen den Tests verursachten Probleme der Retest- und Paralleltest-Reliabilität.

Auch die Kuder-Richardson-Reliabilität basiert auf einer einzigen Form und einer einmaligen Anwendung des Instruments, indem sie die Konsistenz der Antworten auf alle Items des Tests misst. Die Kuder-Richardson-Formel ist das Mittel aus allen Split-Half-Koeffizienten auf der Basis unterschiedlicher Aufteilungen des Tests.

Die Formel der Spearman-Brown-Reliabilität ist ein weiteres statistisches Verfahren, das Reliabilitätskoeffizienten liefert und regelmäßig zusammen mit den Split-Half-Verfahren zur Anwendung kommt. Die Spearman-Brown-Methode zeichnet sich durch die Verdoppelung der Zahl der Items im Instrument als Teil der Formel aus. Die Verdoppelung der Zahl der Items hat in der Regel eine gesteigerte Reliabilität zur Folge. Manche Kritiker der Spearman-Brown-Formel sagen, der Reliabilitätskoeffizient des Tests werde so künstlich angehoben. Alle bisher diskutierten Reliabilitätskoeffizienten lassen sich per Hand oder mit einem einfachen Taschenrechner ermitteln.

Cronbachs Alpha ( $\alpha$ ) (Cronbach, 1951) ist nach Einschätzung vieler das bislang robusteste Reliabilitätskoeffizient (Anastazi, 1976; Reynolds, 1994). »Koeffizient  $\alpha$  ist der maximale Wahrscheinlichkeitsschätzwert des Reliabilitätskoeffizienten, wenn das Parallelmodell als wahr angenommen wird« (SPSS, S. 873). Für dichotome Daten »ist Cronbachs Alpha äquivalent mit dem Koeffizienten der Kuder-Richardson-Formel 20 (KR20)« (SPSS, S. 873). Der Alpha-Koeffizient bringt die Reliabilität eines Instruments zum Ausdruck und rangiert von 0 bis +1,00. Ein Instrument mit perfekter Reliabilität hätte einen Alpha-Koeffizienten von +1,00, und kein psychologisch-diagnostisches Instrument hat diesen Wert bislang erreicht. Überdies existiert kein Standard, kein anerkannter Wert, ab dem eine Korrelation für Testzwecke als gut oder schlecht zu gelten hat. Man hat sich jedoch allgemein darauf verständigt, als Minimalstandard ein Alpha von mindestens 0,6 anzusetzen, wobei einige Testautoren für 0,7 oder einen höheren Wert plädieren. Je höher der Alpha-Koeffizient, desto stärker ist ganz offensichtlich die Kohärenz der Items.

Die Reliabilitätskoeffizienten des PIW (Persönliche Interessen und Werte) Instruments werden mittels Cronbachs Alpha bestimmt. Der Leser möge die in diesem Handbuch präsentierten Reliabilitätskoeffizienten mit den Reliabilitäten anderer Instrumente vergleichen und auch darauf achten, wie andere Anbieter ihre Reliabilitätszahlen errechnen.



## Kontextbasierte Validität und Beziehungen zu anderen Variablen

Die Validität liefert eine Antwort auf die Frage: »Misst das Instrument, was es messen soll? Sie stellt zudem eine tiefer qualitätsbezogene Frage: Wie gut misst das Instrument diese Dinge? Diese Fragen sind offensichtlich schwieriger zu beantworten als die Frage nach der Reliabilität und lassen unter Umständen Raum für Subjektivität. Bei allen Fragen zur Validität kommt es entscheidend auf die Beziehung zwischen der Ergebnismessung des Instruments und anderen beobachtbaren Fakten bezüglich des betrachteten Verhaltens an. Wenn jemand sagt: »Der Test war nicht fair«, dann zielt der Kommentar in der Regel auf die Validität und nicht die Reliabilität des Tests. Zutreffender könnte man formulieren: »Der Test war nicht valide.«

Es gibt drei wesentliche Formen der Validität: Inhaltliche, kriterienbezogene und Konstruktvalidität.

Die inhaltliche Validität untersucht den Inhalt des Instruments daraufhin, ob er das zu messende Verhalten abdeckt. Eine einfache Überprüfung der Items in einem Biologie- oder Chemietest sollte Fragen zutage fördern, die in Bezug zum Lernstoff stehen. Im Zusammenhang mit der Entwicklung der PIW – Persönliche Interessen und Werte -Themen bedeutet inhaltliche Validität, dass alle sechs Merkmalskategorien in gleichen Proportionen repräsentiert sind. Darüber hinaus umfasst die inhaltliche Validität die Überprüfung der sozialen Erwünschtheit. Sobald ein Ungleichgewicht besteht zwischen den Deskriptoren, die gesellschaftlich akzeptiert, und solchen, die gesellschaftlich weniger akzeptiert sind, tangiert dies die inhaltliche Validität. Das PIW – Persönliche Interessen und Werte -Instrument wird auf inhaltliche Validität hin überprüft, mit der Folge, dass seit der ursprünglichen Testfreigabe einige Deskriptoren zugunsten der inhaltlichen Validität und der Reliabilität des Instruments ersetzt wurden.

Die kriterienbezogene Validität bezeichnet die Fähigkeit eines Instruments, das Verhalten eines Teilnehmers in bestimmten Zukunftssituationen vorherzusagen. Die Ergebnisse des Teilnehmers im Instrument werden mit diversen externen Kriterien verglichen. Im Rahmen von PIW – Persönliche Interessen und Werte-Instrument und -Berichten stehen diverse Studien von Success Insights und TTI Performance Systems bereit, die eine klare Beziehung zwischen bestimmten Ergebnissen oder Ergebnismustern und beruflichem Erfolg in spezifischen, klar definierten Bereichen herstellen (Bonnstetter u.a., 1993). Kriterienbezogene Validität hat zwei Formen: konkurrente Validität und prädiktive Validität. Die konkurrente Validität vergleicht die Ergebnisse eines Teilnehmers mit externen Kriterien zum Zeitpunkt der Verabreichung des Instruments. Die prädiktive Validität vergleicht die Ergebnisse des Instruments mit Kriterien nach einem genauer definierten Zeitintervall.

Konstruktvalidität untersucht die Fähigkeit eines Instruments, ein theoretisches Konstrukt oder Merkmal zu messen. Die Konstruktvalidität basiert auf einem Evidenzmuster und diversen Angaben aus einer Vielzahl von Quellen. Zu den in der Verhaltensmerkmalanalyse betrachteten Konstrukten gehören beispielsweise Entwicklungsveränderungen von Teilnehmern, die in unterschiedlichen Lebensjahren, in unterschiedlichen Lebensphasen oder mit unterschiedlichen Aufmerksamkeitsschwerpunkten auf das Instrument antworten. Korrelationen mit anderen Tests ermöglichen ebenfalls ebenfalls eine Form der Konstruktvalidierung. Es gibt diverse Vergleiche zwischen dem PIW – Persönliche Interessen und Werte Instrument und anderen Verhaltensinstrumenten wie dem MBTI (Myers Briggs Type Indicator), MMPI (Minnesota Multiphase Personality Inventory), 16-PF (16 Personality Factor) und anderen Instrumenten.

Alle diese Vergleichsstudien bestätigen die Gesamtkonstrukte des PIW – Persönliche Interessen und Werte Instruments.

Eine sehr wichtige Technik innerhalb der Konstruktvalidierung ist die Faktorenanalyse. Diese Technik verfeinert ein Instrument, indem es seine Itemstruktur miteinander vergleicht und Beziehungen zwischen den einzelnen Items analysiert. Das Ziel ist es dabei, aus sämtlichen Antwortenalternativen eine kleinere Zahl von Faktoren oder gemeinsamen Merkmalen herauszudestillieren. Auch das PIW – Persönliche Interessen und Werte® Insights® Instrument ist im Zuge der Faktorenanalyse verfeinert worden, um die Gesamtvalidität und -reliabilität von Instrument und den Reports zu erhöhen (siehe »Überprüfung der theoretischen Kohärenz« ab Seite 18).

Eine Anmerkung zur Augenscheinvalidität: Augenscheinvalidität darf nicht mit den oben erwähnten Validitätsformen verwechselt werden. Augenscheinvalidität ist keine richtige Validität im technischen Sinne des Wortes. Die Augenscheinvalidität schaut, ob das Instrument in den Augen der Teilnehmer »valide erscheint«, und ob der Bericht für den Leser »valide klingt«. Während Augenscheinvalidität statistisch betrachtet keine Form der Validität ist, spielt sie dennoch eine wichtige Rolle sowohl bei der Instrumentenentwicklung als auch bei der Reportabfassung.

### **Konvergente und divergente Evidenz**

Die Validitätsprüfung umfasst noch zwei weitere Punkte. Beide betreffen im Prinzip die Frage, ob eine Klassifizierung durch das Instrument ähnliche Testpersonen korrekt identifiziert (konvergent) und hinreichend zwischen Testpersonen unterscheidet, die zu unterschiedlichen Klassifikationen gehören (divergent). Auch hier beruht die Macht der Evidenz im Wesentlichen auf den fundierten Anwendungserfahrungen der das Instrument verwendenden geschulten Berater.

## **4.3. Situation und Umfeld**

Menschen funktionieren und agieren in öffentlichen Situationen. Folglich kommen die zu messenden psychologischen Eigenschaften in soziopsychologischen Interaktionen sowie in inneren Merkmalen zum Ausdruck. Im Unterschied zur Messung rein innerer Überzeugungen und Werte werden Verhaltensweisen durch diese Interaktionen gefördert oder zurückgedrängt. In diesem Zusammenhang tauchen drei Fragen hinsichtlich der Messung von Eigenschaften auf, die in dieser Weise Ausdruck finden. Eine Frage betrifft die Situationsanforderungen und die individuelle Wahrnehmung des Umfelds, in dem die Messung stattfindet. Eine zweite Frage dreht sich um die vom Teilnehmer wahrgenommene sozialen Erwünschtheit der als Indikatoren der sechs PIW – Persönliche Interessen und Werte-Skalen verwendeten Deskriptoren (Fragen). Die dritte Frage entsteht, sobald Deskriptoren übersetzt und das Instrument in einem anderen sprachlich-kulturellen Umfeld eingesetzt wird. Auf alle drei Themen wollen wir im Folgenden eingehen.

### **Situationsanforderungen**

Situationsanforderungen entstehen aus dem Umfeld und seinen Bedingungen heraus, wie sie der Proband wahrnimmt. Da unser Instrument auf Selbsteinschätzungen beruht, können die Antwortentscheidungen je nach bewusster oder auch unbewusster Wahrnehmung der Situation, in der das Instrument eingesetzt wird, verfälscht sein. Wenn der Proband die Situation als risikoarm wahrnimmt, äußert er sich möglicherweise offener

(im Sinne von ehrlicher) als in einem als bedrohlich empfundenen Umfeld. Wenn er hingegen die Situation als hochgradig bedrohlich empfindet, ist er möglicherweise weniger offen, um auf diese Weise ein positiveres Bild von sich zu vermitteln. Das ist ein Problem, das alle auf Selbsteinschätzungen basierenden (ipsativen) Instrumente gleichermaßen betrifft, und das es zu berücksichtigen gilt, wenn dieses Instrument im Rahmen des Screenings und der Auswahl von Einstellungskandidaten eingesetzt wird.

Wird ein Empfehlungsschreiben im Bewerbungsportfolio verlangt, bitten die Kandidaten stets jemanden um ein solches Schreiben, der ein sehr positives Bild vermitteln kann. Ebenso präsentieren sie sich im Einstellungsgespräch von ihrer allerbesten Seite. Deshalb besteht Grund zu der Annahme, dass auch die Antworten in einem im Zuge des Kandidatenscreenings verwendeten Instruments einer gewissen »Eigenkorrektur« unterworfen sind. Das betrifft alle entsprechenden Instrumente, egal von welchem Anbieter. Das vorliegende Instrument versucht, mittels der Vielzahl von Fragenkomplexen das Muster der Eigenkorrektur in den gegebenen Antworten zu minimieren.

### Soziale Erwünschtheit

Nicht alle Beschreibungen oder Eigenschaften einer Skala sind als skalierbare Items in einem Instrument geeignet. Es gibt Beschreibungen, die gesellschaftlich als wünschenswert oder unerwünscht vorbelegt sind. Manche Beschreibungen einer Skala können also gesellschaftlich erwünscht sein. Wer eine Skaleneigenschaft in stärkerem Maße besitzt, handelt vielleicht häufiger in dieser gesellschaftlich gewünschten Art und Weise. Aber jeder möchte sich gern in diesem positiven Licht sehen. Wenn also diese Beschreibung zur Auswahl steht und viele Probanden sie wählen, erschwert das die Unterscheidung zwischen Probanden, bei denen dieses Skalenverhalten stark, mittel oder schwach ausgeprägt ist. Ebenso kann es Verhaltensbeschreibungen geben, die viele Menschen als »passt am wenigsten zu mir« bezeichnen würden, selbst wenn diese Beschreibung auf ihr Verhalten durchaus zutrifft.

Beiderlei Arten von Beschreibungen eignen sich nicht zur Differenzierung der Probanden und werden deshalb im Instrument nicht verwendet. Diese Wertvorstellungen von erwünschten und unerwünschten Verhaltensweisen sind gesellschaftlich geprägt und werden von den meisten Mitgliedern einer Kultur getragen. Die Kultur beeinflusst Auswahl und Beschreibungen noch in anderer Weise, wie der folgende Abschnitt zeigt.

### Kultureller Einfluss

Auch wenn in einer Population viele Kulturen und Subkulturen vorhanden sein mögen, verwenden die Versionen des PIW – Persönliche Interessen und Werte Instruments, Sprachgruppen als Differenzierungsmöglichkeit. Kulturen unterscheiden sich deutlich in der Art, wie bestimmte Wertvorstellungen definiert und gewertet werden. Wer eine andere Kultur besucht, bemerkt solche Unterschiede mitunter sofort. Am Beispiel von Verhaltensweisen in denen Wertvorstellungen häufig ihren Ausdruck finden lässt sich dies am deutlichsten erkennen. Lautes gleichzeitiges Sprechen mag in der einen Kultur für gute Freundschaft stehen und in einer anderen auf einen unmittelbar bevorstehenden Kampf hindeuten. Die Beschreibung einer Verhaltensweise mit ähnlichen Worten in zwei unterschiedlichen Sprachen kann sehr verschiedene Konnotationen haben. So können beispielsweise Solidarität und Mitgefühl unterschiedliche Konnotationen hinsichtlich der Rolle von Gleichheit und Sympathie transportieren. Diese Unterschiede gilt es beim Einsatz eines Instruments in unterschiedlichen Kulturen zu berücksichtigen.

Mit Blick auf diese Unterschiede werden für diverse Sprachgruppen je eigene Normierungsversionen von PIW – Persönliche Interessen und Werte entwickelt, validiert und getestet. Die im Instrument als Items verwendeten Beschreibungen werden für jede Sprachversion auf Reliabilität und Kohärenz mit den Skalenkonzepten hin getestet. Wenn das Instrument in hinreichendem Umfang zum Einsatz kommt und die Kunden es für wichtig erachten, können für jede definierbare Subpopulation je eigene Verteilungen und Normen errechnet werden.

## 4.4. Messbarkeitsgrenzen

Im Zusammenhang mit der Messung tauchen drei Probleme auf. Erstens setzt, wie in den Abschnitten »Situationsanforderungen« und »Soziale Erwünschtheit« bereits erwähnt, der durch Selbsteinschätzung gewonnenen Messung gewisse Grenzen. Zweitens liefern Skalen, die nach dem Prinzip der Gleichgewichtungsarbeiten, lediglich Rangfolgen, aber keine quantitativen Daten, siehe der nachfolgende Abschnitt »Itemgewichtung und Skalenkonstruktion«. Drittens sind Forced-Choice-Befragungen, bei denen der Teilnehmer eine Rangfolge bilden muss, mit einigen statistischen Schwierigkeiten verbunden, auf die wir im Abschnitt »Ipsative Messverfahren« weiter unten zu sprechen kommen. Diese Einschränkungen betreffen alle Persönlichkeitsmessverfahren von diesem Typ.

### Itemgewichtung und Skalenkonstruktion

Erstens resultiert aus der Aufsummierung der Häufigkeit von Antworten ein vergleichendes und kein quantitatives Maß. Ein Ergebnis besteht aus der Zahl der vom Probanden ausgewählten Beschreibungen. Diese Zahl wird mit den entsprechenden Zahlen weiterer Versuchspersonen aus der Referenzpopulation verglichen. Die Rohdaten von verschiedenen Skalen können nicht unmittelbar miteinander verglichen werden. Wählt jemand 10 x-Items und 5 y-Items, heißt das nicht, dass er mehr x ist. Wenn aber in der Referenzpopulation der Durchschnitt bei 5 x-Items und 7 y-Items liegt, kann man von dem Teilnehmer, der 10 x-Items auswählt, mit Grund behaupten, dass er sich selbst mehr durch x motiviert sieht, als dies in der Population allgemein zu erwarten wäre. Solange sich die Interpretation auf diese Art von Rangfolgenvergleichen beschränkt, stellt die Beobachtung, dass ein x möglicherweise eine stärkere Verbindung zu einer Eigenschaft hat als ein anderes x, kein Problem dar.

Im vorliegenden Instrument wird der Vergleich angestellt zwischen den individuellen Rohdaten und einem Durchschnittswert der Referenzpopulation. Dabei gilt es zu beachten, dass die Skalen keine quantitativen Aussagen über die Eigenschaften zulassen.

Diese Vergleiche basieren darauf, dass die Referenzpopulation repräsentativ ist für die Probanden die das Instrument nutzen, Im vorliegenden Instrument sind die Vergleichsnormen repräsentativ für die gegenwärtigen Nutzer des Instruments. Wo immer möglich, wurden für die einzelnen Sprach-/Kulturgruppen spezifische Normen entwickelt. Jede für eine Version des Instruments als Referenz verwendete Normverteilung wird klar angegeben. Weitere Ausführungen zum Thema der Referenzpopulation finden Sie auf Seite 19 und in den technischen Zusammenfassungen im Anhang.

## Ipsative Messverfahren

Forced-Choice-Selbstaussagen als Quelle für die Verhaltensskalierung haben in der Psychometrie eine lange Tradition. Solche Messverfahren sind akzeptiert als Methode, um auf Selbstwahrnehmung basierende Einsichten zu gewinnen; dennoch haben sie ihre Grenzen. Ipsative Messverfahren haben statistische Grenzen, die den Einsatz bestimmter Verfahren beschränken. Ipsative Messverfahren unterliegen zudem einem möglichen bewussten oder unbewussten Bias.

Ein unbewusster Bias liegt beispielsweise dann vor, wenn wir uns selbst nicht so sehen, wie andere uns sehen. Unsere Selbstwahrnehmungen, mögen sie auch auf Feedback von anderen basieren, stimmen möglicherweise nicht mit dem überein, wie andere uns beschreiben würden. In der Teambildung und in der Personalentwicklung lohnt unter Umständen die Mühe, sich diese Nichtübereinstimmung bewusst zu machen und mit geeigneten Maßnahmen gegenzusteuern. Ein bewusster Bias liegt beispielsweise vor, wenn ein Teilnehmer mit Absicht versucht, sich in einer Weise zu präsentieren, die nicht mit seinen typischen Wertvorstellungen oder seinen Verhaltensmustern übereinstimmt. Wie bereits oben erwähnt, kommt diese Situation in Bewertungen die auf Basis von einer Selbsteinschätzung entstanden ist etwas häufiger vor, besonders dann, wenn der Teilnehmer überzeugt ist, dass bestimmte Eigenschaften positiv oder negativ belegt sind und er seine Antworten folglich so abgibt, dass er nach seiner Auffassung besonders gut abschneidet. Über diese Problematik sprachen wir ausführlicher im Abschnitt Situationsanforderungen.

# 5. Technische Informationen

## 5.1. Überprüfung und Überarbeitung

Target Training International (TTI) und Success Insights (SI) starteten im Frühjahr 2002 eine Überarbeitung ihres Instruments Personal Interests, Attitudes and Values(TM). Das PIW – Persönliche Interessen und Werte® Insights®-Instrument ist in zwei Reportvarianten erhältlich: Arbeitsstellenanalysen® und dem PIW – Persönliche Interessen und Werte. Das Hauptaugenmerk lag bei dieser Überarbeitung auf der Skalierung und der Item-Reliabilität für die für die Konstruktion der sechs Skalen verwendeten zwölf Gruppen à sechs Sätzen oder 72 Deskriptoren.

Skalenreliabilität und Kohäsion der Items mit ihren zugehörigen Skalen wurden anhand von Stichproben untersucht. Im Folgenden werden die zwecks Überprüfung und Überarbeitung der Itemreliabilität und der Skalenkonstruktionen unternommenen Schritte beschrieben. Spezielle technische Informationen zu den einzelnen Sprachversionen finden Sie im Anhang ab Seite 26.

Alle untersuchten Beispiele stammten von PIW – Persönliche Interessen und Werte-Fragebögen, die während des Jahres vor der Überprüfung ausgefüllt wurden. In den meisten Bereichen überstieg die Zahl der verfügbaren Fälle die für die statistische Testung und Auswertung erforderliche Anzahl. Die benötigte Anzahl an Datensätzen wurde aus dieser größeren Datenmenge ausgewählt. Im durchgeführten statistischen Verfahren konnten die Schlussfolgerungen und Parameter dann bestätigt und bekräftigt werden.

Die meisten statistischen Verfahren erfordern für die Überprüfung keine großen Fallzahlen. Für die meisten statistischen Evaluationen wurden folglich Zufallsstichproben aus den entsprechenden Subpopulationen

ausgewählt. Mit Hilfe der Stichproben konnten Hypothesen entwickelt werden, die sich anschließend anhand von weiteren Stichproben, die unabhängig von den ersten waren, überprüfen ließen. Diese Testung diente häufig dazu, Überarbeitungsempfehlungen zu bestätigen. Derartige Vergleiche bestätigten allgemeine Muster psychologischer Eigenschaften mit signifikanten Unterschieden bezüglich der Art, wie bestimmte Indikatoren (Wörter, Ideen) in unterschiedlichen Sprach- und Kulturgruppen miteinander verbunden sind.

Bei der Überprüfung der Kohärenz der PIW – Persönliche Interessen und Werte-Skalen wurden zwei Ansätze verfolgt. Der erste Ansatz bestand darin, nach Mustern gemeinsamer Varianz zu suchen (Faktorenanalyse). Dabei ging es um die Frage, ob Antworten Kohärenzmuster erkennen ließen, die die theoretische Konstruktion der Skalen rechtfertigen. Siehe dazu den folgenden Abschnitt »Überprüfung der theoretischen Kohärenz«.

Ein zweiter Ansatz schaute mit Blick auf die Matrix der Skalenkonstruktion auf die Kohärenz der einzelnen Items mit den zugehörigen Skalen und auf die Gesamtreliabilität dieser Skalenkonstruktion. Über diese Überprüfungsmethode, die von Cronbachs Alpha ( $\alpha$ ) Gebrauch macht, sprachen wir im Abschnitt »Kontextbasierte Validität und Beziehungen zu anderen Variablen« auf Seite 13. Die jeweiligen Cronbach-Alphas für die Instrumente finden Sie in den zusammenfassenden Berichten im Anhang.

### Untersuchung der theoretischen Kohärenz

Die Konstruktion einer Skala beginnt mit der Übersetzung theoretischer Konstrukte in praktische Messverfahren. Zwecks Bestätigung der Kohärenz der einer Skala zugeordneten Beschreibungen wurde eine Stichprobe von Antworten mittels Hauptkomponentenanalyse untersucht. Bei diesem statistischen Verfahren wurden die 72 Items auf Muster ähnlicher Variation hin untersucht. Jeder Faktor ist ein latentes Konstrukt, eine nicht gemessene Eigenschaft. Ergebnis des Verfahrens ist eine Liste von Faktoren zusammen mit einem Kovarianzmaß für jede Variable. Diese Koeffizienten können positiv, negativ oder neutral sein. Indem wir Items mit substantziellen positiven oder negativen Koeffizienten bezüglich eines Faktors wählen, identifizieren wir eine Itemkonstellation, die einen latenten Faktor beschreibt.

Häufig spiegelt ein Faktor zwei kontrastierende Itemmengen wider. Eine Eigenschaft lässt sich unter den Items mit positiven Koeffizienten und eine zweite unter den Items mit negativen Koeffizienten finden. Wenn die Itemliste mit der Liste jener Items übereinstimmt, die einer Skala theoretisch zugeordnet sind, können wir darauf schließen, dass die Implementierung der Theorie als Skala hinreichend fundiert ist. Wenn ein Item einen starken positiven Koeffizienten mit anderen Items hat, die zu einer Skala gehören, der es nicht zugeordnet ist, sind Fragen zur Theorie und/oder zum Item erlaubt. Die meisten Items harmonierten mit ihren zugehörigen Skalen. Die häufigste Abweichung sind jedoch Items, die mit keiner Skala einen starken positiven Koeffizienten aufweisen. In diesem Fall ist das Item kein brauchbarer Indikator für die Messung einer Eigenschaft, selbst wenn es sich um eine gute Beschreibung handelt. Dieses Paradox wurde im Abschnitt zur »Sozialen Erwünschtheit« auf Seite 15 erklärt.

### Item- und Skalenreliabilitäten

Skalenreliabilitäten wurden mittels des Cronbachschen Alphas ( $\alpha$ ) berechnet. Cronbachs  $\alpha$  gilt als der geeignetste statistische Reliabilitätstest auf der Basis der Antworten, mit denen die Skalen konstruiert wurden. Für dichotome Daten ist Cronbachs  $\alpha$  äquivalent zum Koeffizienten der Kuder-Richardson-Formel 20 (KR20). Diese Kennzahlen modellieren die interne Konsistenz auf der Grundlage der durchschnittlichen Korrelation zwischen den Items.

Dieses Evaluationsverfahren ist strenger als das traditionelle Split-Half-Verfahren. Cronbachs  $\alpha$  ist eine Kenngröße zwischen 0 und 1. Im Allgemeinen gilt 0,6 als niedrigster akzeptabler Wert, auch wenn einige namhafte Testautoren für einen strengeren Standard von mindestens 0,7 plädieren.

Da dieses Handbuch für mehrere Sprachversionen von PIW – Persönliche Interessen und Werte erscheint, sind die Reliabilitätskennzahlen in den Zusammenfassungen der einzelnen Sprachversionen im Anhang aufgeführt. Spezielle Koeffizienten für einzelne Items (Indikatoren) mit ihrer zugehörigen Skala werden nicht angegeben, weil es sich dabei um proprietäre Informationen handelt, die den Schlüssel zur spezifischen Auswertung und Itemanalyse des Instruments darstellen.

## 5.2. Normen und Populationsparameter

Der Stammbaum der gegenwärtigen Versionen von PIW – Persönliche Interessen und Werte basiert auf einer Vielzahl von Evaluationen mit diversen Datenquellen und -stichproben. Im Rahmen der im Jahr 2002 begonnenen Überprüfung früherer Versionen wurden über einhunderttausend Antwortbögen ausgewertet. Die gegenwärtige Item- und Skalenreliabilität ist das Resultat dieser wiederholten Evaluationen anhand verschiedener Stichproben. Der Ursprung des Instruments wird in regelmäßigen Abständen durchgeführten unabhängigen Evaluationen gestärkt. Die Stichproben stammen jeweils von den aktuellen Nutzern des Instruments. Diese Nutzer repräsentieren ein breites Spektrum. Die spezifischen Stichprobenbeschreibungen finden sich auf den technischen Informationsblättern zu jeder veröffentlichten Version im Anhang.

Dadurch verlagerte sich der Bezugspunkt für den Stilvergleich von seinem historischen Entwicklungspunkt ins 21. Jahrhundert mit seinen veränderten Verhaltensweisen und gesellschaftlichen Erwartungen.

### Geschlechterverteilung

Eine Thematik die jedes Fragebogen-Instrument betrifft, das im 21. Jahrhundert für Kunden im kommerziellen und auch im individuellen Kontext entworfen wird, ist der Einfluss des Geschlechts auf die Antwortmuster. Bei der Evaluation des Instruments ging es unter anderem um Unterschiede in den Antwortmustern von Männern und Frauen. Wie nicht anders zu erwarten, gibt es gewisse Unterschiede in den durchschnittlichen Skalenresultaten für Männer und Frauen. Diese Unterschiede deuten auf vergleichsweise geringe Verschiebungen in der Dominanz bestimmter Verhaltensausrprägungen hin. Ob diese Unterschiede biologisch oder gesellschaftlich bedingt oder beides sind, ist für die Effektivität des Instruments nicht relevant. Wichtig ist, dass die Messverfahren des Instruments ein Feedback liefern, das keinen Geschlechter-Bias beinhaltet.

Die für die Aufstellung der Verteilungsnormen verwendeten Stichproben wurden daraufhin evaluiert. Wenn in einer Stichprobe Männer und Frauen proportional vertreten waren, wurden keine Korrekturen vorgenommen. Wenn die Verteilung Männer zu Frauen in der Datenstichprobe disproportional war wurden entsprechenden und notwendigen statistischen Korrekturen so vorgenommen, dass die Auswirkung der unterschiedlichen Antwortmuster von Frauen und Männern ausgeglichen wurden. Diese Korrekturen sind in den technischen Informationsblättern zu den einzelnen Versionen im Anhang vermerkt.



## Sprachversionen

Das PIW – Persönliche Interessen und Werte-Instrument ist in mehreren Sprachversionen verfügbar. Mit der Freigabe der durchgeführten Überarbeitungen wurden viele dieser Versionen separat evaluiert und als unabhängige Instrumente entwickelt. Im Rahmen einer solchen Entwicklung werden die aus der englischen Version übersetzten Itembeschreibungen auf ihre Kohärenz mit ihrer zugehörigen Skala hin analysiert und die Reliabilitäten jener Skalen bewertet. Als Ergebnis dieses Prozesses werden Items editiert und, wenn erforderlich, Skalen überarbeitet, um so zu einem Instrument zu gelangen, das zuverlässig misst und für die anvisierte Sprach-/Kulturgruppe geeignet ist. Spezifische Verteilungsnormen für die Sprachversion wurden auf Basis der Antworten auf diese Sprachversion errechnet, um den Nutzern ein Feedback zu liefern, das für die Sprach-/Kulturgruppe, die das Instrument nutzt, relevant ist. Für jede Sprachversion werden anschließend gesonderte detaillierte Informationsblätter herausgegeben, siehe Anhang.

# 6. Instrumentenprotokolle und Einsatz

## Verwaltung und Einsatz

Die Ergebnisse des PIW – Persönliche Interessen und Werte Instruments lassen sich in zwei konkreten, gesonderten Formaten dokumentieren. Beide Instrumente, PIW – Persönliche Interessen und Werte – und die Arbeitsstellenanalyse liefern Feedback zu den sechs Skalen. In beiden Reports sind allgemeine Informationen grafisch und textlich dargestellt, indem sie die Testpersonen mit einer Referenzpopulation vergleichen.

Berater, die den Einsatz dieser Instrumente in Erwägung ziehen, können selbst entscheiden, welche Reportvariante im Hinblick auf ihre konkreten Bedürfnisse und Kundenwünsche besser geeignet erscheint.

Eine Einschulung für die Nutzung und Interpretation wird von der Insights Deutschland GmbH angeboten und ist für alle Berater und Anwender des Instruments verpflichtend vorgesehen. Es finden Jahreskonferenzen, Webseminare und Telefonberatungen zum Zwecke des Supports und der vertiefenden Schulung statt.

## Report

Diese Reports werden so abgefasst, dass sie den praktischen Nutzen der Reportergebnisse maximieren. Die Probanden können mit ihrer Hilfe ihre eigene Motivation und damit sich selbst besser verstehen lernen. Teams können damit ihre Zusammenarbeit verbessern und ihr gegenseitiges Verständnis erhöhen, aber auch die Stärken der einzelnen (Team-)Mitarbeiter besser zur Geltung bringen. Dem Management dienen sie zur Kandidatenauswahl für bestimmte Aufgaben oder Spezialeinsätze auf der Grundlage individueller Antriebs- und Motivationsfaktoren. Und sie können in einer Vielzahl von Unternehmen und Branchen zur Einschätzung im Rahmen der Kandidatenauswahl für die Stellenbesetzung verwendet werden.

Im Report der Arbeitsstellenanalyse ist der Teil zum PIW – Persönliche Interessen und Werte Instrument so gestaltet, dass das höchste Wertethema des Kandidaten zuerst – meistens auf Seite 7 des Berichts – und das am geringstem ausgeprägte Wertethema zuletzt – in der Regel ab Seite 15 – erscheint. Jedes Thema beginnt mit einer allgemeinen Definition des Themas, gefolgt von Einträgen zu fünf Kategorien:

1. Allgemeine Eigenschaften
2. Wert für das Unternehmen
3. Wie sich der Betreffende führen und motivieren lässt
4. Training, berufliche Entwicklung und Lernerfolge
5. Verbesserungsfähige Bereiche

## Skalierung und Diagramme

Der PIW – Persönliche Interessen und Werte Fragebogen dient sowohl dazu, die relative Bedeutung der sechs Motivationsskalen für den Einzelnen zu ermitteln, als auch den Einzelnen mit einer breiteren Referenzpopulation zu vergleichen.

Die Reports liefern Feedback in Textform zur jeweiligen Bedeutung der sechs Motivationsskalen für den Einzelnen sowie dazu, wie diese Motivation ausgedrückt und interpretiert werden kann. Umfang und Schwerpunkt dieses Feedbacks kann je nach gewählter Reportvariante unterschiedlich ausgestaltet sein.

Ein Vergleich mit einer Normgruppe wird in Form einer Tabelle gegeben, wobei der Mittelwert als Bereich markiert ist, der eine Standardabweichung von  $\pm 1$  repräsentiert. In diesen Bereich fallen 68 Prozent der Ergebnisse der Referenzpopulation. Der Vergleich des Einzelnen mit der Referenznorm liefert dann die expliziten Ergebnisse.

- Im Durchschnitt, falls der Wert in den  $\pm 1$ -Bereich fällt,
- Indifferent, falls er darunter liegt (kleiner oder gleich -1),
- Stark ausgeprägt, falls er darüber liegt (größer oder gleich +1),
- Extrem, falls er mit Abstand darüber liegt (größer oder gleich +3).

Ein zweites graphisches Feedback wird in Form eines Balkendiagramms gegeben, das die Ergebnisse des Einzelnen in den sechs Skalen zusammen mit den nationalen Durchschnitten zeigt. Daraus geht nicht nur hervor, was den Einzelnen motiviert, sondern auch, wo er damit im Vergleich zur (nationalen) Referenzpopulation steht. Bei der Verwendung des Graphen gilt es zu beachten, dass die nationalen Durchschnittswerte für die sechs Skalen nicht auf einem Niveau liegen. Die Interpretation sollte sich folglich auf die relative Höhe der Einzelbalken im Vergleich zum jeweiligen nationalen Mittel und nicht auf die absolute Höhe der Einzelergebnisse beziehen.

Das in diesen Berichten präsentierte Balkendiagramm spiegelt zwei Arten von Feedback wider. Erstens wird jedes Thema im Instrument von einem der sechs vertikalen Balken des Graphen vertreten. Die Skalierung wird als Rohdatenskala angegeben, mit Anker beim Wert von 12 Punkten und Maximum beim Wert von 72 Punkten. Zweitens wird zur Eindeutigkeit der Interpretation die individuelle Rangfolge der Themen unter den Rohdaten vermerkt.

Auf diese Weise wird bei der raschen Interpretation des Diagramm auf den ersten Blick deutlich, welches die erste, zwei, dritte usw. Präferenz ist.

Numerische Unterschiede in der Reihenfolge aufgrund von nahezu gleichhohen Punktwerten sind

nicht signifikant und sollten als Gleichstände betrachtet und interpretiert werden. Ein Vergleich der Rangordnung des Einzelnen mit den Ergebnissen der Referenzpopulation liefert in solchen Fällen möglicherweise ein informativeres Feedback. Auch gilt es zu bedenken, dass eine wiederholte Beantwortung des PIW Fragebogens bei solch ähnlichen Punktwerten leicht zu einer Veränderungen in der Rangordnung führen kann. Das ist typisch für Instrumente, die auf Selbsteinschätzungen basieren.

Ein Kreisdiagramm liefert eine dritte Visualisierung der Rohdaten und der relativen Rangliste der Testperson.

### Training und Interpretation

Eine Einführung in die Anwendung und Interpretation kann durch die **INSIGHTS MDI® Deutschland GmbH** vermittelt werden. Trainings- und Zertifizierungsprogramme sind erhältlich für diejenigen, die sich mit dem Einsatz von PIW – Persönliche Interessen und Werte und den anderen **INSIGHTS MDI®** Instrumenten vertraut machen wollen.

## 7. Literaturverzeichnis

Allport, G.W., Vernon, P.E., & Lindzey, G. (1960). Boston: Houghton Mifflin.

Allport, G.W. (1961). Pattern and Growth in Personality. New York: Holt, Rinehart, & Winston.

Anastazi, A. (1976). Psychological Testing. New York: Macmillan Publishing Co.

Bonnstetter, „Bill“ J., J. I. Suiter, R. J. Widrick. (1993). The Universal Language™ DISC: A Reference Manual, 6th ed. Target Training International, Ltd., Scottsdale, Arizona.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika 16, 297-334.

Reynolds, C.R. (1994). Reliability. In R.J. Sternberg (Ed.), Encyclopedia of Human Intelligence. New York: Macmillan Publishing Co.

Rokeach, M. (1973). The Nature of Human Values. New York: Macmillan.

Spranger, E. (1928). Types of Men. Berlin: Max Niemeyer Verlag, Halle

Spranger, E. (1966). Lebensformen. Max Niemeyer Verlag GmbH, Tübingen

SPSS. (1988). SPSS-X User's Guide, 3rd Ed. Chicago: SPSS Inc

## 8. Anhang

### 8.1. PIW / Motive (Driving Forces), Deutsche Version 2021

#### Hintergrund

Die regelmäßig vorgenommene Aktualisierung der Normen, zuletzt im Januar 2014 (siehe auch nachfolgende Seiten Beschreibung **INSIGHTS MDI®**: Deutsche Version 2014) ist ein wichtiger Bestandteil des laufenden kontinuierlichen Verbesserungsprozesses der TTI SUCCESS INSIGHTS Analysen. So wurden auch im Mai 2021 von TTI unter Verwendung der deutschsprachigen Datensätze erneut Reliabilitätsstudien durchgeführt. Die Auswertung erfolgte auf Basis der statistischen Berechnungen von Dr. Eric Gehrig, Ph. D. in Mathematics und VP of Research & Development bei TTI SUCCESS INSIGHTS. Diese Auswertungen wurden von TTI für alle unterschiedlichen Sprachversionen des Motivationsanalyse Instruments darunter auch das deutschsprachige Instrument **INSIGHTS MDI®** vorgenommen.

Die Daten der Befragten wurden aus dem TTI SI Internet Delivery System (IDS)<sup>®</sup> ausgewählt und bestehen hauptsächlich aus der allgemeinen Geschäfts- oder berufstätigen Erwachsenenbevölkerung der einzelnen Sprachen. Für die Studie wird eine Zufallsstichprobe von Männern und Frauen aus jeder Sprache gesammelt. Die aktuelle Studie berücksichtigt Daten aus den Jahren 2016 bis 2020.

Basierend auf den Ergebnissen dieser Studie kann man schließen, dass die Einschätzungen der **INSIGHTS MDI®** Auswertungen als intern konsistent und zuverlässig bestätigt werden.

#### Normierungsstichprobe

Die verwendete Stichprobe umfasste 33.802 Fälle, wobei die Geschlechterverteilung unter den Probanden genau gleich war. Die Anzahl der weiblichen Testpersonen betrug 16.901, die der männlichen Testpersonen auch 16.901.

	Total (n = 33.802)		Männer (nm = 16.901)		Frauen (nf = 16.901)	
	Mittelwert	Abweichung	Mittelwert	Abweichung	Mittelwert	Abweichung
Intellektuell	58	15,0	57	15,0	58	15,1
Effizienzgetrieben	44	13,4	49	13,1	40	12,9
Harmonisch	40	16,8	33	15,5	47	16,6
Altruistisch	28	14,0	25	12,9	31	14,8
Machtorientiert	50	17,8	57	16,5	42	17,4
Prinzipientreu	29	12,4	28	12,2	31	12,6
Instinktiv	25	12,8	25	12,9	24	12,7
Idealistisch	38	13,3	32	12,6	42	13,1
Objektiv	44	15,7	50	15,0	38	15,1
Eigennützig	56	15,6	58	14,6	51	16,1
Kooperativ	35	17,6	26	15,3	42	18,0
Aufgeschlossen	54	13,3	56	13,0	53	13,6

## Skalenreliabilität

Zur Überprüfung der Skalenzuverlässigkeit wurde die Cronbach Alpha ( $\alpha$ ) Methode angewandt. Es stellt die geeignetste Maßzahl für die interne Konsistenz einer Skala dar und bezeichnet das Ausmaß, in dem die Items bzw. Fragen einer Skala miteinander in Beziehung stehen. Deshalb wird Cronbachs Alpha auch als Maß der internen Konsistenz einer Skala bezeichnet.

Die durchgeführte Überprüfung ergab robuste Cronbach Alpha Reliabilitätskennzahlen der einzelnen Skalen. Die angestrebten Werte von  $\alpha$  0,7 werden von allen der sechs Skalen erreicht, sie liegen insgesamt in sehr zufriedenstellenden Bereichen zwischen 0,70 bis 0,87.

Somit ist gewährleistet, dass die sechs Skalen des PIW – Persönliche Interessen und Werte Instruments sehr zuverlässig messen.

Theoretischer Wert	0,80
Ökonomischer Wert	0,71
Ästhetischer Wert	0,84
Sozialer Wert	0,84
Individualistischer Wert	0,87
Traditioneller Wert	0,70

## 8.2. PIW – Persönliche Interessen und Werte, Deutsche Version 2014

### Hintergrund

Die bis dato durchgeführten Überprüfungen der Reliabilitäten der sechs Werte-Dimensionen wurden von TTI SUCCESS INSIGHTS unter Verwendung der deutschsprachig Datensätze durchgeführten. Die regelmäßig vorgenommene Aktualisierung der Normen, zuletzt im September 2009 (siehe auch nachfolgende Seiten Beschreibung) erfolgte auf Basis der statistischen Berechnungen von Peter T. Klassen, Ph. D. Dieser Auftrag wurde von TTI für alle unterschiedlichen Sprachversionen des PIW – Persönliche Interessen und Werte darunter auch das deutschsprachige Instrument **INSIGHTS MDI®** vergeben.

Im Zuge der Qualitätsinitiative der **INSIGHTS MDI International® Deutschland GmbH** wurde die Berechnung und Überprüfung der aktuellen Reliabilitätskennzahlen für das PIW – Persönliche Interessen und Werte Instrument als Auftrag extern vergeben. Dies sollte eine unabhängige und neutrale Überprüfung durch einen Statistikexperten sicherstellen.

Der Auftrag wurde von Prof. Dr. Karlheinz Zwerenz in München ausgeführt. Im Januar 2014 wurden die benötigten Datensätze in anonymisierter Form als Rohdaten zur Verfügung gestellt. Hauptanliegen war die von TTI bereitgestellten Reliabilitätskennzahlen zu überprüfen (siehe Anhang Stellungnahme Prof. Dr. Zwerenz) und aktuelle Zahlen zur Nutzung zur Verfügung zu haben.

## Normierungsstichprobe

Die verwendete Stichprobe umfasste 31.785 Fälle, wobei die Geschlechterverteilung unter den Probanden nahezu ausgeglichen war. Die Anzahl der weiblichen Testpersonen betrug 15.618, die der männlichen Testpersonen 16.167.

Stichprobe gesamt						
Mittelwert (x)	49,46	49,72	32,36	44,10	42,00	34,37
Anzahl (n)	31.785	31.785	31.785	31.785	31.785	31.785
Standardabweichung (s)	8,958	9,883	8,871	10,461	8,580	7,654

Stichprobe Frauen						
Mittelwert (x)	47,80	47,70	34,51	46,89	39,11	35,98
Anzahl (n)	15.618	15.618	15.618	15.618	15.618	15.618
Standardabweichung (s)	8,835	9,692	9,111	10,458	8,569	7,809

Stichprobe Männer						
Mittelwert (x)	51,05	51,67	30,28	41,40	44,79	32,82
Anzahl (n)	16.167	16.167	16.167	16.167	16.167	16.167
Standardabweichung (s)	8,785	9,672	8,106	9,729	7,614	7,168

## Skalenreliabilität

Zur Überprüfung der Skalenzuverlässigkeit wurde die Cronbach Alpha ( $\alpha$ ) Methode angewandt. Es stellt die geeignetste Maßzahl für die interne Konsistenz einer Skala dar und bezeichnet das Ausmaß, in dem die Items bzw. Fragen einer Skala miteinander in Beziehung stehen. Deshalb wird Cronbachs Alpha auch als Maß der internen Konsistenz einer Skala bezeichnet.

Die durchgeführte unabhängige Überprüfung durch Prof. Dr. Karlheinz Zwerenz ergab robuste Cronbach Alpha Reliabilitätskennzahlen der einzelnen Skalen. Die angestrebten Werte von  $\alpha$  0,7 werden von fünf der sechs Skala erreicht, sie liegen insgesamt in sehr zufriedenstellenden Bereichen zwischen 0,615 bis 0,866.

Somit ist gewährleistet, dass die sechs Skalen des PIW – Persönliche Interessen und Werte Instruments sehr zuverlässig messen.

Theoretischer Wert	0,797
Ökonomischer Wert	0,812
Ästhetischer Wert	0,789
Sozialer Wert	0,866
Individualistischer Wert	0,721
Traditioneller Wert	0,615

## 8.3. PIW – Persönliche Interessen und Werte: Deutsch 2008

### Schlussfolgerungen für den Einsatz:

- Die Struktur des Instruments wird durch diese Daten von 2008 bestätigt.
- Die sechs Skalen zeigen bei der Untersuchung mittels Cronbachs Alpha eine akzeptable Reliabilität ( $\alpha > 0,7$ ).
- Die sechs Skalen erweisen sich als unabhängig zueinander.
- Die Skalen zeigen eine Normalverteilung mit geeigneten Bereichen für die Interpretation.
- PIW – Persönliche Interessen und Werte: Die deutschsprachige Version zeigt Reliabilität und Konstruktvalidität.

### Daten und Methodik

Das Instrument PIW – Persönliche Interessen und Werte von Success Insight baut auf dem Design und der Entwicklung früherer PIW-Instrumente auf. Die aktuelle im Einsatz befindliche Version des Instruments basiert auf der Version die im Jahr 2003 freigegeben wurde. Mittels der im Jahr 2008 gesammelten Daten starteten TTI und Success Insights eine Follow-Up-Bewertung.

### Beschreibung der Stichprobe

Die am 19. März 2009 eingegangenen Daten umfassten insgesamt 17.807 Antworten. Diese Antworten stammten von zwei Teilgruppen. In Summe waren 63 Prozent der Probanden Männer. und 37 Prozent Frauen mit unterschiedlichem beruflichem Hintergrund. Auf Basis einer sorgfältigen Untersuchung der zwei Untergruppen und der Unterschiede in den von Männern und Frauen stammenden Antworten wurde bei der Berechnung der Referenznormen eine Faktor-2-geschichtete Stichprobenkorrektur angewandt.

In Anbetracht der Stichprobengröße darf mit guten Grund postuliert werden, dass diese Datenzahl als repräsentative Referenzstichprobe für die Evaluation des Instruments dienen kann. Der folgende Bericht konzentriert sich auf drei Punkte.

- Reliabilität und Itemkohärenz
- Beziehungen zwischen den Skalen
- Referenznormen

### Reliabilität und Itemkohärenz

Die Skalen wurden konstruiert, indem ein psychologisches Konzept theoretisch definiert und anschließend Indikatoren entworfen werden, die geeignet sind, die Bedeutung zu erfassen, die die Testperson dem Konstrukt beimisst. Diese Items werden sodann zum Nachweis der Reliabilität auf ihre Kohärenz hin evaluiert. Cronbachs Alpha stellt ein anerkanntes Reliabilitätsmaß dar. Vereinfacht ausgedrückt ist  $\alpha$  der Durchschnitt aller Korrelationen zwischen den ausgewählten Items. Somit ist es stärker als das Split-Half- und andere Verfahren, die früher in der Computersoftwareentwicklung zur Anwendung kamen. Tabelle 5 listet die früheren und die jetzigen  $\alpha$ -Koeffizienten auf.



	Theoretisch	Ökonomisch	Ästhetisch	Sozial	Individualistisch	Traditionell
PIAV1.G	0,59	0,71	0,76	0,71	0,57	0,37
PIAV2.G 2003 (N=691)	0,78	0,83	0,80	0,85	0,72	0,68
PIAV2.G 2008 (N=17807)	0,77	0,81	0,78	0,86	0,73	0,77

Tabelle 5: Cronbachs Alpha ( $\alpha$ )

Laut Bewertung mittels Cronbachs Alpha erscheinen alle Skalen kohärent und messen zuverlässig. Die Cronbachs  $\alpha$ -Koeffizienten liegt im Bereich zwischen 0,856 und 0,729.<sup>2</sup>

## Beziehung zwischen den Skalen

Tabelle 6 listet die Korrelationen zwischen den sechs Skalen auf.

	Theoretisch	Ökonomisch	Ästhetisch	Sozial	Individualistisch	Traditionell
Theoretisch	1					
Ökonomisch	-0,077	1				
Ästhetisch	-0,223	-0,349	1			
Sozial	-0,372	-0,428	-0,111	1		
Individualistisch	0,081	0,229	-0,462	-0,407	1	
Traditionell	-0,370	-0,475	0,219	0,207	-0,435	1

Tabelle 6: Korrelationen zwischen den Skalen

Den stärksten positiven Koeffizienten haben Ästhetisch und Traditionell, eng gefolgt von Sozial und Traditionell. Das kann entweder eine wahre psychologische Beobachtung widerspiegeln oder, was wahrscheinlicher ist, eine Überlappung im Fokus der für die Konstruktion der Skalen verwendeten Items.

Die negativen Koeffizienten zwischen den meisten der Zahlen zeigen an, dass diese Skalen zueinander in Opposition stehen. Höhere Werte auf einer dieser Skalen gehen mit niedrigeren Werten auf mehreren anderen Skalen einher. Ein wesentlicher Grund dafür ist die ipsative Natur der Instrumentenkonstruktion.

2 Cronbachs Alpha ( $\alpha$ ) (Cronbach, 1951) ist nach Einschätzung vieler das bislang robusteste Reliabilitäts-Alpha (Anastazi, 1976; Reynolds, 1994). »Koeffizient  $\alpha$  ist der maximale Wahrscheinlichkeitsschätzwert des Reliabilitätskoeffizienten, wenn das Parallelmodell als wahr angenommen wird« (SPSS, S. 873). Für dichotome Daten »ist Cronbachs Alpha äquivalent mit dem Koeffizienten der Kuder-Richardson-Formel 20 (KR20)« (SPSS, S. 873). Ein Instrument mit perfekter Reliabilität hätte einen Alpha-Koeffizienten von +1,00, und kein Instrument hat diesen Wert bislang erreicht. Überdies existiert kein Standard, kein anerkannter Wert, ab dem eine Korrelation für Testzwecke als gut oder schlecht zu gelten hat. Man hat sich jedoch allgemein darauf verständigt, als Minimalstandard ein Alpha von mindestens 0,6 anzusetzen, wobei einige Experten für 0,7 oder einen höheren Wert plädieren. Sämtliche Skalenreliabilitätskoeffizienten für das PIW – Persönliche Interessen und Werte-Instrument werden mittels Cronbachs Alpha ermittelt.

## Referenznormen

Die für die Festlegung der Referenznormen verwendeten Daten wurden so korrigiert, dass die Antworten von Männern und Frauen in den beiden Stichprobenuntergruppen gleichgewichtig vertreten sind. Die stratifizierte Stichprobenkorrektur, bei der jeder Fall gewichtet und die Summen ganzzahlig gerundet wurden, resultierte in einem Gesamtstichprobenumfang von 17 804 statt der ursprünglichen 17 807. Mittels dieser Korrektur wurden die Antworten der Männer und Frauen in ein gleichgewichtiges Verhältnis gesetzt, ohne die relative Größe der beiden Untergruppen zu verändern. Das bedeutet, dass die Referenznormen die in der Population zu erwartenden männlich-weiblichen Unterschiede und nicht die Unterschiede in den Rohdaten widerspiegeln.

Wenngleich die statistischen Maße für Schiefe und Wölbung statistische Signifikanz signalisieren, ist dies eine Folge der großen Zahl der für die Berechnung dieser Koeffizienten verwendeten Antworten. Eine Untersuchung zeigte, dass alle sechs Skalen relativ nah bei der Normalverteilung liegen.

### Eingereicht von:

Peter T. Klassen, Ph. D.

Principal, DocumentingExcellence.com

Professor Emeritus, College of DuPage

20. April 2009

## 8.4. PIW – Persönliche Interessen und Werte – Deutsche Version: 2003

### Zusammenfassung

Die Reliabilität der sechs Skalen wurde gegenüber der Vorläuferversion verbessert. Drei der sechs Skalen weisen ein Cronbachsches Alpha von 0,8 oder mehr auf, und eine weitere Skala kommt diesem Wert nahe; zwei Skalen haben Alphas zwischen 0,68 und 0,72. Die zwischen den Versionen 2.b und 2.c vorgenommenen Änderungen bewirkten eine deutliche Verbesserung der betreffenden Items. Die Überprüfung der Korrelationen zwischen den Skalen belegt die relative Unabhängigkeit der gemessenen Konzepte.

### Hintergrund

Das PIW – Persönliche Interessen und Werte Instrument baut auf dem Design und der Entwicklung des früheren PIAV-Instruments von TTI auf. Im Spätsommer 2002 starteten TTI und Success Insights eine Überprüfung der Reliabilität der sechs Skalen und der zugehörigen Items.

Das PIW – Persönliche Interessen und Werte Instrument umfasst zwölf Kästen à sechs Aussagen. Jede Aussage ist ein Indikator für einen der sechs latenten Werte. Die Testpersonen ordnen die sechs Items von 1 bis 6, wobei 1 für die höchste und 6 für die niedrigste Bewertung steht. Die Skalen werden konstruiert, indem die Werte invertiert, zusammengehörige Itemränge aufaddiert und das Ergebnis nach oben korrigiert wird, um Nullen zu vermeiden. Die Skalen tragen die Bezeichnungen Theoretisch, Ökonomisch, Ästhetisch, Sozial, Individualistisch und Traditionell.

Skalenreliabilitäten und Kohäsion der Items zu den zugehörigen Skalen werden wie später beschrieben überprüft. Basierend auf der Analyse dieser Indikatoren wurde eine begrenzte Zahl von Items überprüft, editiert und in Feldversuchen getestet. Die Itemüberprüfung erfolgte auf der Grundlage der theoretischen Konstruktion der Items in Kombination mit linguistischen Überlegungen insbesondere mit Blick auf den aktuellen Sprachgebrauch und die Minimierung des Bias im Bereich der sozialen Erwünschtheit bei der Verwendung der deutschsprachigen Versionen. Die deutsche Version ist mehr als eine Übersetzung; sie wurde eigenständig getestet und auf ihre Anwendbarkeit hin untersucht.

Das PIW – Persönliche Interessen und Werte® Insights® Instrument wurde vor seiner Freigabe in mehreren Durchläufen in Feldversuchen getestet, weiter editiert und erneut überprüft. Die gegenwärtige Fassung weist in allen Skalen eine gesteigerte oder hohe Reliabilität auf. Die aktuelle Analyse basiert auf Daten, die am 7. September 2003 eingingen. Diese Daten umfassen 691 Antworten auf die deutsche Version des PIW – Persönliche Interessen und Werte® Insights® Instruments und die errechneten Skalenergebnisse, die Anzahl und Zusammensetzung der Daten kann als repräsentativ betrachtet werden. 42,7 Prozent der Befragten waren Frauen, 57,3 Prozent waren Männer. Die für die Analyse verwendete Skalenkonstruktionsmatrix wurde mittels der bereitgestellten errechneten Ergebnisse bestätigt. Anschließend wurden für jede Skala das Cronbachsche Alpha und Item/Skalen-Korrelationen errechnet. Im Anschluss an die Bewertung der Skalenkonstruktion wurden Korrelationen zwischen den sechs Skalen berechnet.

## Reliabilität

Skalenreliabilitäten wurden mit Hilfe von Cronbachs Alpha errechnet. Cronbachs Alpha gilt als der geeignetste statistische Test für Reliabilität. Diese Statistiken modellieren interne Widerspruchsfreiheit auf der Grundlage der durchschnittlichen Korrelation zwischen den Items. Diese Evaluationen bieten einen robusteren Ansatz als eine traditionelle Split-Half-Statistik. Cronbachs Alpha liefert Werte zwischen 0 und 1. Im Allgemeinen gilt ein  $\alpha$  von 0,6 als minimaler akzeptabler Wert, auch wenn einige Experten für eine strengere Norm von mindestens 0,7 plädieren.

	Deutsche PIW 1 (N = 1145)	Deutsche PIW 2.c (N = 691)
Theoretisch	0,59	0,78
Ökonomisch	0,71	0,83
Ästhetisch	0,76	0,80
Sozial	0,71	0,85
Individualistisch	0,51	0,72
Traditionell	0,37	0,68

Tabelle 7: Skalenreliabilität

Drei der sechs Skalen haben ein Cronbachsches Alpha von 0,8 oder höher, und eine weitere Skala kommt dem Wert sehr nahe; zwei Skalen haben Alphas von 0,68 beziehungsweise 0,72.

## Korrelationen

Eine Überprüfung der Korrelationen zwischen den Skalen belegt eine weitgehende Unabhängigkeit der gemessenen Konzepte. Die meisten Korrelationen sind klein oder negativ. Die stärkste Korrelation ist die zwischen Sozial und Ökonomisch mit -0,49, woraus folgt, dass sie ungefähr 24 Prozent Varianz in einer inversen Beziehung teilen. Die stärkste positive Korrelation ist die zwischen Ökonomisch und Individualistisch mit 0,246, entsprechend einer geteilten Varianz von rund sechs Prozent.

	Theoretisch	Ökonomisch	Ästhetisch	Sozial	Individualistisch	Traditionell
Theoretisch	1					
Ökonomisch	-0,120	1				
Ästhetisch	-0,190	-0,243	1			
Sozial	-0,270	-0,493	-0,171	1		
Individualistisch	0,046	0,246	-0,380	-0,431	1	
Traditionell	-0,340	-0,418	0,070	0,228	-0,374	1

Tabelle 8: Korrelationen zwischen den Skalen  
Spearman-Rangordnungskorrelationen nicht signifikant bei 0,000

## Schlussfolgerungen

Ich (Peter T. Klassen, Ph. D.) bin in Bezug auf das Instrument PIW – Persönliche Interessen und Werte von Success Insights auf der Basis einer Analyse der Antwortmuster einer deutschen Stichprobe von Testpersonen zu folgenden Schlussfolgerungen kommen:

- Die Ergebnisse auf den Skalen bilden zwar keine statistische Normalverteilung, sind aber dennoch hinreichend über die Skalen verteilt, um Interpretationen vorzunehmen und Testpersonen mit den Verteilungen dieser selbstberichteten Interessen und Werte zu vergleichen. Wenngleich die Skalen mit jeweils gleich vielen Items konstruiert wurden, gilt es zu beachten, dass aus den Häufigkeitsreaktionsmustern in der Indikatorenauswahl durch die Testpersonen Ergebnisse mit substantiell unterschiedlichen statistischen Verteilungen resultieren.
- Tendenziell kovariieren die Items mit konsistenten Konstruktindikatoren der von den Instrumentbeschreibungen repräsentierten theoretischen Konzepte.
- Die sechs Skalen werden aus kohärenten Items mit solider Reliabilität konstruiert, ausgewiesen durch ein Cronbachsches Alpha zwischen 0,66 und 0,85.
- Die in der deutschen Version von PIW2 vorgenommenen Überarbeitungen verbessern signifikant die Skalenreliabilitäten und die Korrelationen der Items mit ihren zugehörigen Skalen.

Diese Evaluation und Überprüfung der von Success Insights bereitgestellten deutschen Version des PIW – Persönliche Interessen und Werte Instruments stellt den Beginn eines kontinuierlich fortgeführten Qualitätsverbesserungsprozesses dar.

Peter T. Klassen, Ph. D.  
Principal, DocumentingExcellence.com  
Professor Emeritus, College of DuPage

## Stellungnahme Prof. Dr. Karlheinz Zwerenz

AMW Institut, Perhamer Str. 18, 80687 München

Herrn  
Frank Scheelen  
Vorstandsvorsitzender  
Scheelen AG  
Klettgaustraße 21  
79761 Waldshut-Tiengen



München, den 17.1.2014

**Berechnungen von Cronbachs Alpha für die Modelle R4 und PIAV2  
analog TTI-SUCCESS**

**STELLUNGNAHME**

Cronbachs Alpha ist eine Maßzahl für die interne Konsistenz einer Skala. Der Wertebereich von Alpha reicht von minus unendlich bis 1. Skalen sollten nur dann Verwendung finden, wenn Alpha mindestens den Wert von 0,65 erreicht.

Bei allen 10 überprüften Skalen konnten mit SPSS die Berechnungsergebnisse von TTI-SUCCESS (bis auf geringfügige Differenzen von maximal +/-0,003) bestätigt werden. Sämtliche Alphas liegen über 0,65 und bestätigen daher die Konsistenz der Skalen.

Diese Aussagen gelten sowohl für die Berechnungen auf Basis der Rohdaten als auch für die Ergebnisse, die auf Basis standardisierter Werte ermittelt wurden.

Für Rückfragen – auch wegen Einsatzmöglichkeiten der Faktorenanalyse – stehe ich Ihnen gerne zur Verfügung.

Viele Grüße



Prof. Dr. Karlheinz Zwerenz  
AMW Institut München

Anlage: Excel-Tabelle mit Berechnungsergebnissen

---

AMW Institut, Perhamer Str. 18, 80687 München,  
Prof. Dr. Karlheinz Zwerenz, Tel.: 089 / 54644265 Fax: 089 / 58927660

