

Die deutsche Version der CES-D als Instrument der gerontologischen Forschung¹

von M. Riediger, M. Linden und H.-U. Wilms

An Daten der Berliner Altersstudie wurde die Anwendbarkeit der deutschen Version der Center for Epidemiological Studies Depression Scale (CES-D) in der gerontologischen Forschung untersucht. Die Skala weist eine ausreichende Güte für den Einsatz als Forschungsinstrument auf. Sie verfügt über eine hohe interne Konsistenz. Die deutsche CES-D ist in gerontologischen Untersuchungen als Screening-Instrument zur Identifikation von Fällen mit „Depressiven Störungen“ und mit „Major Depression“ einsetzbar. Der bisher übliche Einsatz eines einheitlichen Schwellenwertes erscheint jedoch in verschiedenen Alters-Geschlechts-Gruppen der hochbetagten Bevölkerung nicht gerechtfertigt. Gütekennwerte verschiedener potentieller Schwellenwerte wurden für die Gesamtstichprobe und die nach Geschlecht unterteilten 70 bis 84jährigen sowie 85jährigen und älteren Studienteilnehmer dargestellt. Diese Werte können – da sie auf einer repräsentativen Stichprobe der mindestens 70jährigen Bevölkerung West-Berlins beruhen – als Orientierungshilfe bei der Auswahl geeigneter Schwellenwerte in gerontologischen Untersuchungen dienen. Eine konfirmatorische Faktorenanalyse erbrachte eine zufriedenstellende Anpassungsgüte des wiederholt für die englische Skalenversion bestätigten Vier-Faktoren-Modells. Hieraus kann auf eine weitgehende Maßäquivalenz der englischen und deutschen CES-D geschlossen werden. Beide Versionen erfüllen somit eine grundlegende Voraussetzung für den Einsatz in der kulturvergleichenden Forschung. Obwohl die Skala insgesamt eine für Forschungsinstrumente ausreichende Güte aufweist, zeigten einige Items in unserer Untersuchungspopulation unbefriedigende Itemkennwerte.

The applicability of the German version of the Center for Epidemiological Studies Depression Scale (CES-D) in gerontological research was investigated. The German CES-D shows satisfactory psychometric properties for the use as research instrument. It has a high internal consistency. The German CES-D is applicable as screening instrument for detection of subjects with „Depressive Disorders“ and „Major Depression“. It is, however, concluded that the so far common practice of using a unitary cutting point in different age-by-sex groups is not warranted for many gerontological research questions. Goodness indices of several potential cutting points are reported for the total sample and several sub-samples. These results are based on data of a representative sample of the at least 70 years old West-Berlin population and can, therefore, be used as a guideline for choosing relevant cutting points in investigations of old subjects. Confirmatory factor analysis was conducted testing the four-factor-structure repeatedly replicated in investigations of the English CES-D. The four-factor-solution fits the data in our sample satisfactorily well. This result indicates measurement equivalence of the English and German CES-D. It is concluded that both versions are applicable in cross-cultural comparisons of elderly persons. Although the current German CES-D exhibits a sufficient quality as research instrument, several items showed unsatisfactory item properties in our sample.

¹ Die vorliegende Arbeit ist im Rahmen der multidisziplinären Berliner Altersstudie (BASE) entstanden. BASE wird von der Arbeitsgruppe „Altern und gesellschaftliche Entwicklung“ (AGE) der Berlin-Brandenburgischen Akademie der Wissenschaften in Zusammenarbeit mit der Freien Universität Berlin, der Humboldt-Universität zu Berlin und dem Max-Planck-Institut für Bildungsforschung durchgeführt. Das Projekt wurde vom Bundesministerium für Forschung und Technologie (1989-1991, Förderkennzeichen: 13 TA 011 + 13 TA 011/A), vom Bundesministerium für Familie, Senioren, Frauen und Jugend (1992-1998, Förderkennzeichen: 314-1722-102/9 + 314-1722-102/9a) sowie von den beteiligten Institutionen gefördert. Dem Leitungsgremium der Berliner Altersstudie gehören Prof. Dr. P.B. Baltes (Sprecher), Prof. Dr. K.U. Mayer (stellvertr. Sprecher), Prof. Dr. H. Helmchen sowie Prof. Dr. E. Steinhagen-Thiessen an.

Riediger u.a., Die deutsche Version des CES-D

Die Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977) ist eine Depressivitäts-Skala, die vom Center for Epidemiological Studies des National Institute of Mental Health der USA speziell für den Einsatz in epidemiologischen Untersuchungen entwickelt wurde. Diese Skala zielt auf die Erfassung des aktuellen Zustands depressiver Symptomatik ab. Es wird die Häufigkeit des Auftretens von 20 Symptomen in der letzten Woche erfragt. Mit dem Ziel, Antwortstereotypen entgegenzuwirken und um positiven Affekt – bzw. dessen Abwesenheit – zu beurteilen, sind vier der Items in positiver Richtung formuliert (Items 4, 8, 12, 16; vgl. Abbildung 1). Antwortalternativen sind in vier Abstufungen von „kaum oder überhaupt nicht“ bis „meistens oder die ganze Zeit“ vorgegeben. Jedes Item wird entsprechend der gewählten Antwort mit einem Punktwert zwischen Null und Drei bewertet. Die Auswertung erfolgt nach Umpolung der positiv formulierten Items durch Addition der einzelnen Itempunkte. Der Gesamtskalenwert kann zwischen 0 und 60 liegen. Höhere Werte entsprechen einer größeren Anzahl bzw. Häufigkeit von in der vergangenen Woche aufgetretenen Symptomen. Von den Skalenentwicklern wird ein Cut-off-Wert von 16 zur Identifikation von Personen mit „hohem Depressionsrisiko“ (Radloff & Teri, 1986) vorgeschlagen. Die Festlegung dieses Schwellenwertes kann den Erfordernissen der jeweiligen Untersuchung entsprechend verschoben werden. Die meisten der Studien, die die CES-D anwenden, nutzen jedoch diesen kritischen Wert von 16, der inzwischen als „Standard-Cut-off“ gilt.

In nordamerikanischen Untersuchungen gehört die CES-D zu den am häufigsten eingesetzten Depressivitäts-Fragebögen (Fechner-Bates, Coyne &

Schwenk, 1994). Die Skala wurde von Hautzinger (1988) in die deutsche Sprache übersetzt. Diese Skalenversion wird unter dem Namen „Allgemeine Depressions Skala“ (ADS) vom Beltz Test Verlag vertrieben (Hautzinger & Bailer, 1992). Die Autoren empfehlen, den Standardabweichung über dem Mittelwert der Normstichprobe (N = 1205) liegenden Skalenwert 23 als Screening-Kriterium der ADS einzusetzen. Die Überprüfung der Frage, ob sich dieser Skalenwert tatsächlich als effizienter Cutting-Punkt zur Identifikation depressiver Störungen bewährt, steht noch aus. Die im Testmanual dargestellte Untersuchung einer Stichprobe von N = 86 Patienten mit akuter Depression ermöglichte lediglich die Ermittlung des Gütekennwertes Sensitivität (94%). Güteindizes zur Beurteilung der Effektivität der Screening-Entscheidungen (z.B. Spezifität, positiver und negativer Vorhersagewert) werden nicht berichtet. In Abbildung 1 sind zum Vergleich die deutsche Version der CES-D¹ und das englische Original (Radloff, 1977) einschließlich der zugehörigen Instruktionen dargestellt.

Der „Transfer“ der psychometrischen Güte eines Instruments von einem kulturellen Kontext in einen anderen kann nicht als selbstverständlich angesehen werden. Durch die Übersetzung eines Instruments können die ursprünglich gute Reliabilität und Validität beeinträchtigt werden. Skalenübersetzungen verlangen daher eine erneute Überprüfung der psychometrischen Eigenschaften im veränderten kulturellen Kontext (van de Vijver & Hambleton, 1996). Bis-

¹ Dargestellt ist die in der vorliegenden Untersuchung eingesetzte deutsche Version, die geringfügig von der ADS abweicht.

es lediglich einen einzigen Faktor („Depressivität“) gibt, der allen 20 Items zugrundeliegt. Informationen über die Prüfung eines derart vereinfachten Alternativmodells liegen für die von McCallum et al. (1995) zusammengefaßten Studien nicht vor.

Ergebnisse

Bei 23 Studienteilnehmern bzw. 4.5% der Gesamtstichprobe wurde die CES-D nicht bzw. nicht vollständig erhoben. Damit ergibt sich eine effektive Stichprobengröße von $N = 493$ mit einem durchschnittlichen Stichprobenalter von 84.67 Jahren.

In Tabelle 1 sind Kennwerte zur Beschreibung der Verteilungseigenschaften der CES-D-Werte in der Gesamtstichprobe sowie in verschiedenen Alters-Geschlechts-Gruppen zusammengefaßt. Frauen berichteten mehr depressive Symptome und erreichten daher tendenziell höhere CES-D-Werte als männliche Studienteilnehmer derselben Altersgruppe (Mann-Whitney-U-Test, $p < .01$). Dies stimmt mit aus der Literatur bekannten Ergebnissen überein (Gebhardt & Kliemitz, 1986). Gleichzeitig ist zu beobachten, daß mindestens 85jährige Personen durchschnittlich höhere Skalenwerte erhielten als 70 bis 84jährige Studienteil-

nehmer gleichen Geschlechts (Mann-Whitney-U-Test, $p < .01$).

Die Mehrzahl der Probanden wählte bei den meisten Items die Antwortmöglichkeit mit dem Punktwert Null („kaum oder überhaupt nicht“). Auf Grund dieser Antwortcharakteristika wurde für die Berechnung des Item-Kennwertes Schwierigkeit festgelegt, daß Items dann als „in Schlüsselrichtung beantwortet“ gelten, wenn sie – nach entsprechender Umpolung bei den vier positiv formulierten Items – mit einem, zwei oder drei Punkten bewertet werden. Die Mehrzahl der Items weist mittlere Schwierigkeitsindizes auf (.20 – .80, Fisseni, 1990). Hohe Schwierigkeiten ($< .20$, Fisseni, 1990) sind für die Items 9 („... dachte ich, mein Leben ist ein einziger Fehlschlag“), 15 („... waren die Leute unfreundlich zu mir“), 17 („... mußte ich manchmal weinen“) und 19 („... hatte ich das Gefühl, daß mich die Leute nicht leiden können“) zu beobachten. Bei diesen Items wählte ein vergleichsweise kleiner Prozentsatz der untersuchten Personen eine „in Schlüsselrichtung“ liegende Antwortalternative. Lediglich Item 8 („... dachte ich voller Hoffnung an die Zukunft“) ist durch eine niedrige Schwierigkeit ($> .80$, Fisseni, 1990) gekennzeichnet. Hier trat die mit einem Punktwert von Null zu bewertende Antwort

Tabelle 1: Kennwerte der CES-D-Verteilungen in der Gesamtstichprobe und verschiedenen Alters-Geschlechts-Gruppen

	Mittelwert	Median	Varianz	Schiefe
Gesamtstichprobe	14.02	12	88.44	1.39
70 bis 84 Jahre				
männlich	10.75	10	47.89	1.12
weiblich	13.94	12	86.63	1.72
85+ Jahre				
männlich	13.95	11.5	76.27	1.15
weiblich	17.67	14	124.39	1.02

Tabelle 2: Reliabilität in der Gesamt- und in verschiedenen Teilstichproben

	Spearman-Brown	Cronbach's α
Gesamtstichprobe (N = 493)	.88	.87
„Junge“ Alte (70-84 Jahre, N = 252)	.86	.85
„Alte“ Alte (85+ Jahre, N = 241)	.89	.87
Demenzdiagnose ¹ (N = 91)	.93	.91

¹ Hierbei handelt es sich um die DSM-III-R-Diagnosen Demenz mit Wahnphänomenen (N = 3), Demenz mit Depression (N = 5) sowie Senile Demenz, NNB (N = 83).

in lediglich 12.4% der Fälle auf. Mit Ausnahme der Items 8 und 15 weisen alle CES-D-Items mittlere (.30 – .50, Fisseni, 1990) bis hohe ($> .50$, Fisseni, 1990) Trennschärfe auf.

Tabelle 2 enthält die für die Gesamtstichprobe sowie verschiedene Teilstichproben ermittelten Reliabilitätskoeffizienten. Die CES-D weist eine stabile und zufriedenstellende Reliabilität in allen untersuchten Gruppen auf.

Abbildung 2 stellt vergleichend in der Gesamtstichprobe (N = 493) ermittelten ROC-Kurven für die breit definierte „Diagnosegruppe depressiver Störungen“ und das „engere Kriterium „Major Depression“ dar. In der „Diagnosegruppe depressiver Störungen“ wurden Diagnosen einer (nicht vollständig remittierten) depressiven Störung der DSM-III-R-Kategorie „Affektive Störungen“, einer „Anpassungsstörung mit depressiver Verstimmung“ sowie „Demenz mit Depression“ zusammengefaßt. Speziell handelt es sich um die Diagnosen „Major Depression, Einzelepisode“ (N = 23), „Major Depression, Rezidivierend“ (N = 2), „Dysthyme Störung“ (N = 11), „Anpassungsstörung mit depressiver Verstimmung“ (N = 5), „Demenz mit Depression“ (N = 5) sowie „Nicht näher

bezeichnete depressive Störung“ (N = 81).

Beide ROC-Graphen sind deutlich von der „Zufallsgeraden“ abgehoben, was für die Fähigkeit der CES-D spricht, zwischen Personen mit und ohne breit definierter „Depressiver Störung“ sowie mit und ohne „Major Depression“ zu unterscheiden. Eingezeichnet sind die Grenzen des für die Anwendung als Forschungsinstrument „akzeptabel“ erscheinenden Bereichs einer Sensitivität von mindestens 75% bei gleichzeitiger Spezifität von mindestens 65%. In diesen Bereich fallen die Cut-off-Werte 13 bis 16 in Bezug auf die Identifikation von Fällen der Diagnosegruppe „Depressive Störungen“ sowie die Schwellenwerte 15 bis 23 für die Identifikation von Fällen mit „Major Depression“ (vgl. Anhangstabelle 1).

Die für beide Diagnosegruppen ermittelten ROC-Kurven in den Teilstichproben der 70 bis 84jährigen Männer (N = 127) und Frauen (N = 125) sowie der mindestens 85jährigen Männer (N = 124)

¹ Die Festlegung eines „akzeptablen“ Sensitivitäts-Spezifitäts-Bereiches kann in dieser Arbeit nur willkürlich erfolgen. Im Anwendungsfall sind die Anforderungen der konkreten Fragestellung zu berücksichtigen.

menhänge zwischen den Items – und somit die Faktorenstrukturen – erhalten bleiben (Watkins, 1989). Vorliegende Untersuchungen zur Faktorenstruktur der englischen CES-D (Golding & Aneshensel, 1989; Hertzog et al., 1990; Davidson et al., 1994; McCallum et al., 1995) replizierten im wesentlichen die vier ursprünglich von Radloff (1977) aufgefundenen Faktoren „Depressiver Affekt“ (Items 3, 6, 9, 10, 14, 17, 18), „Fehlendes Wohlbefinden“ (Items 4, 8, 12, 16), „Psychomotorische Verlangsamung und Somatische Probleme“ (Items 1, 2, 5, 7, 11, 13, 20) sowie „Interpersonelle Probleme“ (Items 15, 19). Im Gegensatz dazu erbrachten explorative Faktorenanalysen der Skalenstruktur der deutschen CES-D wiederholt lediglich einen eindeutigen Faktor, auf dem alle bzw. der Großteil der Items hoch laden (Hautzinger, 1988; Weyerer et al., 1992). Es ergibt sich die Frage, inwieweit das Antwortverhalten deutscher Probanden von dem englischsprachiger Personen verschieden ist, bzw. ob die CES-D in ihrer deutschen Version andere Merkmale erfaßt als das englische Original.

Methodik

Studienteilnehmer und Untersuchungsinstrumente

Der vorliegende Artikel basiert auf Daten der Ersterhebung der Berliner Altersstudie, einer interdisziplinär angelegten Untersuchung alter und sehr alter Menschen aus Berlin (West). Es wurde eine nach Fünf-Jahres-Altersgruppen und Geschlecht geschichtete Zufallsstichprobe von 516 Personen im Alter von 70 bis 103 Jahren (mittleres Alter: 84.92 Jahre) untersucht (Mayer & Baltes, 1996).

Drei Forschungspsychiatern führten mit jedem Studienteilnehmer das halbstrukturierte klinische Interview *Geriatric Mental State – Version A. History and Aetiology Schedule* (GMS-A/HAS) durch (Copeland et al., 1988; Gurland et al., 1976; Henderson et al., 1983; McWilliam et al., 1988). Anhand der Informationen dieser standardisierten GMS-A/HAS-Untersuchung wurden Diagnosen nach DSM-III-R gestellt. Pro Studienteilnehmer wurde ebenfalls die *Hamilton Depression Scale* (HAMD, Hamilton, 1960, 1967) erhoben. Hierbei handelt es sich um eine Fremdbeurteilungs-Skala zur Einschätzung des Schweregrads depressiver Störungen. Dieses Fremdrating nahm der untersuchende Psychiater vor.

Die CES-D wurde in der Berliner Altersstudie in Form eines Interviews eingesetzt. Instruktion und Items wurden von forschungstechnischen Assistentinnen vorgelesen. Den Studienteilnehmern lag eine Lesetafel mit den Antwortmöglichkeiten vor.

Analysemethodik

Die Reliabilitätsprüfung erfolgte unter Anwendung der Testhalbierungs-Methode und der Analyse der internen Konsistenz.

Zur Überprüfung der Leistungsfähigkeit der deutschen CES-D als Screening-Instrument zur Identifikation von Fällen der Diagnosegruppe „Depressive Störungen“¹ sowie der spezifischen Diagnose einer „Major Depression“ wurden

¹ In diese Störungskategorie wurden neben depressiven Störungen der DSM-III-R-Klasse „Affektive Störungen“ ebenfalls „Anpassungsstörungen mit depressiver Verstimmung“ und „Demenzen mit Depression“ aufgenommen.

ROC (Receiver-Operating-Characteristics)-Kurven erstellt. Die Receiver Operating Characteristics Analysis ist ein Verfahren zur Beschreibung der Güte diagnostischer Tests unter Einbeziehung aller möglichen Schwellenwerte. In der ROC-Kurve werden die Anteile richtig positiver und falsch positiver Screening-Entscheidungen (respektive Sensitivität und 1 minus Spezifität) für jeden möglichen Cut-off-Wert – bezogen auf die CES-D 1 bis 60 – gegeneinander abgetragen. Je größer der Abstand einer ROC-Kurve von der „Random ROC“ – Geraden, desto geringer ist die Überlappung der Werteverteilungen von „Fällen“ und „Nicht-Fällen“ und desto größer ist die Diskriminationsfähigkeit des Screenings (Murphy et al., 1987; Mossman & Somoza, 1989; Hsiao et al., 1989).

Die Stichprobe (N = 516) der Berliner Altersstudie ist, wie Selektivitätsanalysen ergaben (Lindenberger et al., 1996), für die Gesamtbevölkerung alter Menschen in West-Berlin repräsentativ und somit als Referenzstichprobe geeignet. Daher wurden für die Gesamtstichprobe und die nach Geschlecht unterteilten Substichproben der „alten“ (mindestens 85jährigen) und „jungen“ (70 bis 84jährigen) Studienteilnehmer Sensitivität, Spezifität, positiver und negativer Vorhersagewert, Gesamteffizienz sowie Kappa für einen breiten Bereich potentieller Cut-off-Punkte (10 bis 30) ermittelt. Diese Daten können in zukünftigen Untersuchungen alter Menschen bei der Auswahl des in der jeweiligen Anwendungssituation geeigneten Screening-Kriteriums zugrunde gelegt werden (Anhangstabellen 1 bis 5).

Zur Überprüfung der Frage, ob die deutsche CES-D dieselben Inhalte wie das englische Original erfaßt, wurde untersucht, inwieweit das Vier-Faktoren-Modell, das ursprünglich von Radloff

(1977) für die englische Version der Skala aufgefunden und nachfolgend wiederholt repliziert wurde, dem Antwortverhalten der untersuchten Population alter und sehr alter deutscher Personen entspricht. Ausgangspunkt dieser Betrachtung ist der Artikel von McCallum et al. (1995), in dem vergleichend die Ergebnisse konfirmatorischer Faktorenanalysen in Untersuchungen älterer kanadischer, australischer und japanischer Personen sowie einer nordamerikanischen Bevölkerungsstichprobe dargestellt sind. Getestet wurde jeweils ein Modell mit vier interkorrelierten Faktoren. Die Berechnungen der vorliegenden Arbeit können als Erweiterung der Zusammenfassung von McCallum et al. (1995) um eine Stichprobe älterer deutscher Probanden aufgefaßt werden.

Strukturgleichungsmodellierung wurde in der vorliegenden Untersuchung als rein konfirmatorisches Datenanalyseinstrument gehandhabt. Ziel der vorliegenden Untersuchung war zu überprüfen, welche Ergebnisse die Anwendung derselben Analysemethodik (konfirmatorische Faktorenanalyse) bei der Prüfung identischer Modellannahmen (Vier-Faktoren-Modell) für verschiedene sprachliche Versionen der CES-D erbringt. Ließen sich die Ergebnisse konfirmatorischer Faktorenanalysen des englischen Originals an Daten einer deutschen Stichprobe replizieren, würde dies auf die Meßäquivalenz der englischen und deutschen CES-D hindeuten.

Neben der Prüfung des Vier-Faktoren-Modells wurde ermittelt, inwieweit eine Ein-Faktoren-Struktur, welche bisher wiederholt in explorativen Faktorenanalysen der deutschen CES-D ermittelt wurde, dem Antwortverhalten der hier untersuchten Personen entspricht. Dieses Ein-Faktoren-Modell unterstellt, daß

Abbildung 1: CES-D - Deutsche Version und Original

Jetzt möchte ich Ihnen einige Sätze vorlesen, bei denen es um Ihr Befinden in der letzten Woche geht. Sie sollen mir bitte jedesmal sagen, ob dieser Satz auf Ihr Befinden in der letzten Woche zutrifft. Dafür haben Sie eine Skala von 0 bis 3 zur Verfügung. Die „0“ bedeutet, daß der Satz kaum oder überhaupt nicht zutrifft, auf weniger als einen Tag in der letzten Woche. Die „1“ bedeutet, daß der Satz manchmal zutrifft, auf 1 bis 2 Tage der Woche. Die „2“ bedeutet, daß der Satz öfter zutrifft, auf 3 bis 4 Tage der Woche. Und „3“ bedeutet, daß der Satz meistens oder die ganze Zeit zutrifft, auf 5 bis 7 Tage der Woche.

Below is a list of the ways you might have felt or behaved. Please tell me how often you have felt this way during the past week. „0“ - Rarely or none of the time (less than 1 day); „1“ - Some or little of the time (1-2 days); „2“ - Occasionally or a moderate amount of time (3-4 days); „3“ - Most or all of the time (5-7 days).

Während der letzten Woche ...
During the last week:

- | | |
|--|---------|
| 1. haben mich Dinge beunruhigt, die mir sonst nichts ausmachen.
<i>I was bothered by things that usually don't bother me.</i> | 0 1 2 3 |
| 2. hatte ich kaum Appetit.
<i>I did not feel like eating; my appetite was poor.</i> | 0 1 2 3 |
| 3. konnte ich meine trübsinnige Laune nicht loswerden.
<i>I felt that I could not shake off the blues even with help from my family or friends.</i> | 0 1 2 3 |
| 4. kam ich mir genauso gut vor wie andere.
<i>I felt that I was just as good as other people.</i> | 0 1 2 3 |
| 5. hatte ich Mühe, mich zu konzentrieren.
<i>I had trouble keeping my mind on what I was doing.</i> | 0 1 2 3 |
| 6. war ich deprimiert/niedergeschlagen.
<i>I felt depressed.</i> | 0 1 2 3 |
| 7. war alles anstrengend für mich.
<i>I felt that everything I did was an effort.</i> | 0 1 2 3 |
| 8. dachte ich voller Hoffnung an die Zukunft.
<i>I felt hopeful about the future.</i> | 0 1 2 3 |
| 9. dachte ich, mein Leben ist ein einziger Fehlschlag.
<i>I thought my life had been a failure.</i> | 0 1 2 3 |
| 10. hatte ich oft Angst.
<i>I felt fearful.</i> | 0 1 2 3 |
| 11. habe ich schlecht geschlafen.
<i>My sleep was restless.</i> | 0 1 2 3 |
| 12. war ich fröhlich gestimmt.
<i>I was happy.</i> | 0 1 2 3 |
| 13. habe ich weniger als sonst geredet.
<i>I talked less than usual.</i> | 0 1 2 3 |
| 14. fühlte ich mich einsam.
<i>I felt lonely.</i> | 0 1 2 3 |
| 15. waren die Leute unfreundlich zu mir.
<i>People were unfriendly.</i> | 0 1 2 3 |
| 16. habe ich das Leben genossen.
<i>I enjoyed life.</i> | 0 1 2 3 |
| 17. mußte ich manchmal weinen.
<i>I had crying spells.</i> | 0 1 2 3 |
| 18. war ich traurig.
<i>I felt sad.</i> | 0 1 2 3 |
| 19. hatte ich das Gefühl, daß mich die Leute nicht leiden können.
<i>I felt that people dislike me.</i> | 0 1 2 3 |
| 20. bin ich überhaupt nicht in Schwung gekommen.
<i>I could not get „going“.</i> | 0 1 2 3 |

her liegt lediglich eine Untersuchung vor, die sich mit der Güte der deutschen CES-D in älteren Populationen befaßt (Weyerer et al., 1992). Die effektive Stichprobengröße ist jedoch mit N = 62 zu klein, um endgültige Aussagen über die Anwendbarkeit der deutschen CES-D in der gerontologischen Forschung zu treffen. In der Normstichprobe der „Allgemeinen Depressions Skala“ sind lediglich N = 87 Personen im Alter von mehr als 60 Jahren vertreten (Hautzinger & Bailer, 1992). Fragen des Einsatzes der Skala in hochbetagten Bevölkerungsgruppen werden nicht diskutiert.

Gegenstand dieses Artikels ist daher die Frage, ob die deutsche Version der CES-D für den Einsatz in epidemiologischen Untersuchungen alter und sehr alter Menschen geeignet ist. Es wird die Reliabilität des Instruments überprüft sowie untersucht, ob sich die Skala für den Einsatz als Screening-Instrument in der ersten Phase eines zweiphasigen Fallidentifikationsprozesses eignet, und inwieweit der bisher praktizierte Einsatz eines einheitlichen Screening-Kriteriums für verschiedene Subpopulationen der alten und sehr alten Bevölkerung gerechtfertigt ist.

Die vorliegende Untersuchung geht von der Grundannahme aus, daß spezifische Anforderungen an die Eigenschaften einer Screening-Entscheidung erst durch konkrete Fragestellungen bestimmt werden. Anhand einer Stichprobe, die für die West-Berliner Gesamtbevölkerung im Alter ab 70 Jahren repräsentativ ist (Lindenberger et al., 1996), wurde für mehrere Schwellenwerte geprüft, inwieweit sie zur Identifikation von Fällen der breit definierten Diagnosegruppe „Depressive Störungen“ und der speziellen Diagnose „Major Depression“ geeignet sind. Da jeder der N = 516 Studienteilnehmer an einer standar-

disierten psychiatrischen Untersuchung teilnahm, konnten folgende Gütekennwerte ermittelt werden:

- Sensitivität (Prozentsatz der „Fälle“ mit positivem Screening)
- Spezifität (Prozentsatz der „Nicht-Fälle“ mit negativem Screening)
- Positiver Vorhersagewert (Prozentsatz der Personen mit positivem Screening, für die sich die Diagnose bestätigt)
- Negativer Vorhersagewert (Prozentsatz der Personen mit negativem Screening, auf die die Diagnose nicht zutrifft)
- Gesamteffizienz (Prozentsatz der insgesamt durch den Test richtig identifizierten „Fälle“ und „Nicht-Fälle“)
- Kappa (zufallskorrigierte Gesamteffizienz)

Diese Güteindizes können in zukünftigen Untersuchungen als Orientierungshilfe bei der Auswahl des für die jeweilige Fragestellung optimalen Screening-Kriteriums dienen. Vorteil einer solchen Vorgehensweise ist, daß potentielle Cutting-Punkte anhand der Übereinstimmung von Screening-Entscheidung und Diagnose beurteilt werden können. Damit wird die Auswahl des für eine spezifische Fragestellung effizientesten Screening-Kriteriums möglich.

Ein weiterer Schwerpunkt dieses Artikels ist die Frage, ob die deutsche Version der CES-D die Voraussetzungen für interkulturelle Vergleiche mit dem englischen Original erfüllt. Zur Überprüfung der Maßäquivalenz von Original und Übersetzung wird auf den Vergleich von Faktorenstrukturen zurückgegriffen. Es wird von der Überlegung ausgegangen, daß, wenn verschiedensprachige Skalenversionen in unterschiedlichen Kulturen dieselben Merkmale erfassen, die Zusam-

Tabelle 4: Güteindizes zur Beurteilung der Gesamtanpassungsgüte des Vier-Faktoren-Modells und des Ein-Faktoren-Modells

Güteindex	Vier-Faktoren-Modell	Ein-Faktoren-Modell
GFI	0.98	0.97
AGFI	0.97	0.96
RMR	0.065	0.079
RMSEA	0.059	0.075

kennen. In den nordamerikanischen, kanadischen und australischen Untersuchungen der englischen CES-D wurde ebenfalls auf eine zufriedenstellende Anpassung des Vier-Faktoren-Modells geschlossen (Hertzog et al., 1990; McCallum et al., 1995). Das Ein-Faktoren-Modell wurde in diesen Untersuchungen nicht überprüft.

Ein Vergleich der quadrierten Multiplen Korrelationskoeffizienten – als Maß für die durch die Faktoren aufgeklärte Varianz der Items – macht deutlich, daß das Vier-Faktoren-Modell in seinen Teilstrukturen eine bessere Anpassungsgüte als das Ein-Faktoren-Modell aufweist. Dies wird an deutlich höheren quadrierten Multiplen Korrelationskoeffizienten der Items, die im Vier-Faktoren-Modell den hypothetischen Faktoren „Fehlendes Wohlbefinden“ und „Interpersonelle Probleme“ zugeordnet sind, deutlich. Auffallende Abweichungen von den Annahmen beider Faktorenmodelle sind bei Item 8 („... dachte ich voller Hoffnung an die Zukunft“) zu beobachten. Es läßt weder im Vier-Faktoren-Modell auf den hypothetisch zugrundegelegten Faktor „Fehlendes Wohlbefinden“, noch weist es Beziehungen zu dem Faktor „Depressivität“ auf, wenn angenommen wird, daß dieses der einzige Faktor ist, der allen Items zugrundeliegt (Ein-Faktoren-Modell).

Diskussion

Die deutsche Version der CES-D weist in der untersuchten Stichprobe alter und sehr alter Menschen eine zufriedenstellende interne Konsistenz auf. Bisher vorliegende Ergebnisse zur Reliabilität der englischen, spanischen und deutschen CES-D werden somit bestätigt (Radloff, 1977; Himmelfarb & Murrell, 1983; Orme et al., 1986; Mahard, 1988; Hautzinger, 1988; Hautzinger & Bailer, 1992; Weyerer et al., 1992). Reliabilitätsuntersuchungen in den Teilstichproben „junger“ (70 bis 84 Jahre) und „alter“ Studienteilnehmer (mindestens 85 Jahre) zeigen eine deutliche Stabilität der internen Konsistenz der Skala in verschiedenen Altersgruppen an. Die gute interne Konsistenz innerhalb der Substichprobe dementiell erkrankter Personen verdeutlicht, daß kognitive Einbußen sofern sie nicht so gravierend sind, daß sie eine sinnvolle Skalenanwendung unmöglich machen, keinen negativen Einfluß auf die Reliabilität der deutschen CES-D haben.

Die Untersuchung der Eignung der deutschen Version der CES-D als Screening-Instrument zeigt, daß die Skala prinzipiell in der Lage ist, zwischen Fällen und Nicht-Fällen sowohl der breiten Störungsgruppe „Depressive Störungen“ als auch der speziellen Erkrankung einer

„Major Depression“ zu differenzieren. Die deutsche CES-D ist somit prinzipiell für den Einsatz als Screening-Instrument eines zweiphasigen Fallidentifikationsprozesses epidemiologischer Untersuchungen alter und sehr alter Menschen geeignet. Damit stimmt die vorliegende Untersuchung mit Studien überein, die eine Eignung der englischen CES-D als Screening-Instrument feststellten (Myers & Weissman, 1980; Boyd et al., 1982; Zimmermann & Coryell, 1994).

Die Auswahl eines geeigneten Cutting-Punktes als Screening-Kriterium ist stets von den konkreten Erfordernissen der jeweiligen Anwendungssituation abhängig. Aus diesem Grund ist es nicht Ziel dieser Arbeit, einen „optimalen“ Cut-off-Wert zu identifizieren. Es wurden vielmehr Gütekriterien für einen breiteren Bereich potentiell anwendungsrelevanter Cutting-Werte dokumentiert (Anhangstabellen 1 bis 5). Da die Ermittlung der Güteindizes auf den Daten einer repräsentativen Stichprobe der mindestens 70jährigen Bevölkerung West-Berlins beruht, kann die angelegte Dokumentation als Orientierungshilfe bei der Auswahl der bzw. des für die jeweilige Fragestellung geeigneten Cutting-Werte(s) dienen.

Werden Skalenwerte ermittelt, die als potentieller Cut-off-Punkt bestimmte Güteanforderungen erfüllen, so handelt es sich in den Teilstichproben der „jungen“ und „alten“ Männer und Frauen jeweils um unterschiedliche Werte, die sich jedoch für einige der Gruppen teilweise überlappen. Zur Verdeutlichung soll folgendes Beispiel genügen: Der Skalenwert 18 zeigt in der Gruppe der mindestens 85jährigen Frauen eine zufriedenstellende Leistungsfähigkeit zur Identifikation von Fällen mit allgemein definierter „Depressiver Störung“ (Sensitivität: 87.8, Spezifität: 86.8, Kappa: 72.6; vgl. An-

hangstabelle 5). Derselbe Skalenwert 18 weist jedoch in der Gruppe der 70 bis 85jährigen Männer deutlich abweichende Kennwerte auf (Sensitivität: 40.9, Spezifität: 91.4, Kappa: 34.8; vgl. Anhangstabelle 2). Hier ist zwar mit einem geringen Anteil „falsch positiver“ Screening-Entscheidungen zu rechnen, der zu erwartende Prozentsatz „falsch negativer“ Zuordnungen ist jedoch unakzeptabel hoch. Tendenzuell erscheinen in den männlichen Teilstichproben niedrigere Cutting-Werte sinnvoll als in der weiblichen Gruppe der gleichen Altersklasse. Ähnliches ist für die „jungen“ im Vergleich zu den „alten“ Studienteilnehmern festzustellen. Aus diesem Ergebnis ist zu schlußfolgern, daß die Anwendung eines einheitlichen Cutting-Punktes in verschiedenen Alters-Geschlechts-Gruppen der alten Bevölkerung für eine Vielzahl von Fragestellungen ungeeignet erscheint. Der bisher von der Mehrzahl der Anwender praktizierte einheitliche Einsatz des „Standard“-Cut-off-Wertes 16 kann bestenfalls als globaler Kompromißwert akzeptiert werden. Ein einheitlicher Einsatz des von Hautzinger und Bailer (1992) vorgeschlagenen Cutting-Punktes 23 kann nicht empfohlen werden. Er weist in allen untersuchten Substichproben eine zu geringe Sensitivität zur Identifikation von Fällen mit „Depressiven Störungen“ auf. Für die spezielle Diagnose „Major Depression“ zeigt dieser Skalenwert lediglich in den weiblichen Teilstichproben eine zufriedenstellende Leistungsfähigkeit (vgl. Anhangstabellen 2 bis 5).

Eine Itemanalyse erbrachte befriedigende Güteindizes für die Mehrheit der CES-D Items. Vereinzelt Schwächen weisen die Items 8, 9, 15, 17 und 19 auf. Mehr als 80% der untersuchten Personen wählten in Reaktion auf die Items 9 („... dachte ich, mein Leben ist ein ein-

ziger Fehlschlag“) und 17 („... mußte ich manchmal weinen“) die Antwortmöglichkeit „kaum oder gar nicht“. Bei den Items 15 („... waren die Leute unfreundlich zu mir“) und 19 („... hatte ich das Gefühl, daß mich die Leute nicht leiden können“) handelt es sich sogar um mehr als 90% der Studienteilnehmer, die hier Itemwertwerte von Null aufweisen. Diese Items sind somit nicht geeignet, zwischen verschiedenen Ausprägungen des Merkmals „Depressivität“ zu unterscheiden. Die Items 8 („... dachte ich voller Hoffnung an die Zukunft“) und 15 („... waren die Leute unfreundlich zu mir“) weisen nur geringe Trennschärfen auf. Kritik an den beiden „interpersonellen“ Items 15 und 19 („... waren die Leute unfreundlich zu mir“ und „... hatte ich das Gefühl, daß mich die Leute nicht leiden können“) wurde bereits für die englische Skalenversion publiziert. Liang et al. (1989) führen beispielsweise an, daß diese Items eher das Fehlen sozialer Unterstützung und nicht spezifische Depressionssymptome erfassen. Sie argumentieren, daß Fehlen sozialer Unterstützung zwar als Ursache von, nicht jedoch als Indikator für Depressivität angesehen werden sollte. Der Gesamtscore der CES-D kann trotz dieser Kritik an einzelnen Items durchaus sinnvoll interpretiert werden.

Der CES-D-Skalenwert korreliert mäßig mit der Fremdbeurteilung des Schweregrades depressiver Symptomatik unter Anwendung der Hamilton Depression Scale (HAMD, Hamilton, 1960, 1967). Dieses Ergebnis stimmt mit den von Radloff (1977) beschriebenen mittleren Korrelationen zwischen der CES-D und verschiedenen Fremdbeurteilungs-Ratings überein. Fremdbeurteilungen der Symptomschwere werden vom klinischen Urteil des Untersuchers beeinflusst. Im Gegensatz dazu bilden Selbst-

beurteilungs-Skalen subjektive Beschwerden ungefiltert ab. Sie erfassen somit teilweise unspezifische Beschwerden und nicht-depressive Gefühle des Belastetseins. Der Skalenwert einer Selbstbeurteilungsskala kann des Weiteren infolge beeinträchtigter Fähigkeit zur Selbstbeurteilung zum Beispiel aufgrund depressiver Kognitionen oder Anosognosie beeinflusst werden (Helmchen & Linden, 1993; Helmchen et al., 1996). Die beobachteten mittleren Korrelationsstärken können als Ausdruck der methodischen Unterschiede zwischen Selbst- und Fremd-Ratings interpretiert werden.

Mit Hilfe einer konfirmatorischen Faktorenanalyse konnte in der vorliegenden Arbeit bestätigt werden, daß das Vier-Faktoren-Modell als Ganzes mit einer zufriedenstellenden Anpassungsgüte dem Antwortverhalten der untersuchten alten und sehr alten deutschen Personen entspricht. Dies stimmt mit Ergebnissen konfirmatorischer Faktorenanalysen der englischen Skalenversion in einer nordamerikanischen Bevölkerungsstichprobe sowie in Stichproben älterer kanadischer und australischer Personen überein (Hertzog et al., 1990; McCallum et al., 1995). Auch die Betrachtung von Teilstrukturen des Modells deutet auf eine insgesamt zufriedenstellende Anpassung des Vier-Faktoren-Modells hin. Die einzelnen Items weisen jeweils mittlere bis hohe Ladungen auf den vier hypothetischen Faktoren auf. Es sind deutliche Ähnlichkeiten zu Faktorladungen der nordamerikanischen, kanadischen und australischen Untersuchungen beobachtbar.

Ein von den Modellannahmen abweichendes Verhalten zeigt lediglich Item 8 („Während der letzten Woche dachte ich voller Hoffnung an die Zukunft.“). Dieses Item steht in der untersuchten deut-

schen Stichprobe in keiner Beziehung zu dem modelltheoretisch zugrunde gelegten Faktor „Fehlendes Wohlbefinden“. Ein solches Ergebnis wurde für die englische Skalenversion nicht ermittelt. Das abweichende Verhalten dieses Items der deutschen CES-D weist auf einen möglichen Übersetzungsbias hin.

Insgesamt lassen sich sehr deutliche Parallelen in der Struktur des Antwortverhaltens der untersuchten deutschsprachigen Personen und der nordamerikanischen, kanadischen und australischen Stichproben erkennen. Es kann geschlußfolgert werden, daß die deutsche Skalenversion – mit Ausnahme von Item 8 – eine weitgehende Meßäquivalenz zum englischen Original aufweist. Dies spricht für die prinzipielle Möglichkeit kulturübergreifender Vergleiche der englischen und deutschen CES-D.

Ein Modell, das lediglich einen Faktor annimmt, der allen 20 Items der Skala zugrunde liegt, weist eine zum Vier-Faktoren-Modell vergleichbare Gesamtanpassungsgüte auf. Allerdings ist dieses Ein-Faktoren-Modell dem komplexeren Modell in der Anpassungsgüte seiner Teilstrukturen unterlegen. Es erklärt deutlich weniger Varianz der Items, die im Vier-Faktoren-Modell den Faktoren „Fehlendes Wohlbefinden“ und „Interpersonelle Probleme“ zugeordnet sind. Die bislang in explorativen Faktorenanalysen der deutschen CES-D ermittelte Ein-Faktoren-Struktur stellt demzufolge im Vergleich zum Vier-Faktoren-Modell keine „bessere“ Repräsentation der Struktur der deutschen Skalenversion dar.

LITERATURVERZEICHNIS

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J. & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 53-63.

Berkman, L.F., Berkman, C.S., Kasl, S., Freeman, D.H., Leo, L., Ostfeld, A.M., Corononi-Huntley, J. & Brody, J.A. (1986). Depressive Symptoms in Relation to Physical Health and Functioning in the Elderly. *American Journal of Epidemiology*, 124, 372-388.

Berry, J.M., Storandt, M. & Coyne, A. (1984). Age and Sex Differences in Somatic Complaints Associated with Depression. *Journal of Gerontology*, 39, 465-467.

Boyd, J.H., Weissman, M.M., Thompson, W.D. & Myers, J.K. (1982). Screening for depression in a community sample: Understanding the discrepancies between depression symptom and diagnostic scales. *Archives of General Psychiatry*, 39, 1195-1200.

Copeland, J.R.M., Dewey, M.E., Henderson, A.S., Kay, D.W.K., Neal, C.D., Harrison, M.A.M., McWilliam, C., Forshaw, D. & Shiwach, R. (1988). The Geriatric Mental State (GMS) used in the community: replication studies of the computerized diagnosis AGE-CAT. *Psychological Medicine*, 18, 219-233.

Davidson, H., Feldman, P.H. & Crawford, S. (1994). Measuring depressive symptoms in the frail elderly. *Journals of Gerontology*, 49, 159-164.

Fechner-Bates, S., Coyne, J.C. & Schwenk, T.L. (1994). The relationship of self-reported distress to depressive disorders and other psychopathology. *Journal of Consulting and Clinical Psychology*, 62, 550-559.

Fisseni, H. (1990). *Lehrbuch der psychologischen Diagnostik*. Göttingen: Verlag für Psychologie.

Gebhardt, R. & Kliemitz, H. (1986). Depressive Störungen, Geschlecht und Zivilstand. *Zeitschrift für Klinische Psychologie: Forschung und Praxis*, 15, 3-20.

Golding, J.M. & Aneshensel, C.S. (1989). Factor structure of the Center for Epidemiologic Studies Depression Scale among Mexican Americans and Non-Hispanic Whites. *Psychological Assessment*, 1, 163-168.

Gurland, B.J., Copeland, J., Sharpe, L. & Kelleher, M. (1976). The Geriatric Mental State Interview (GMS). *International Journal of Aging and Human Development*, 7, 303-311.

Hamilton, M. (1960). A Rating Scale for Depression. *Journal of Neurology, Neurosurgery and Psychiatry*, 23, 56-62.

- Hamilton, M. (1967). Development of a Rating Scale for Primary Depressive Illness. *British Journal of Consulting and Clinical Psychology*, 6, 278-296.
- Hautzinger, M. (1988). Die CES-D Skala: Ein Depressionsmeßinstrument für Untersuchungen in der Allgemeinbevölkerung. *Diagnostica*, 34, 167-173.
- Hautzinger, M. & Bailer, M. (1992). *Allgemeine Depressions Skala*. Weinheim: Beltz Test GmbH.
- Helmchen, H., Baltes, M.M., Geiselman, B., Kanowki, S., Linden, M., Reischies, F.M., Wagner, M. & Wilms, H.U. (1996). Psychische Erkrankungen im Alter. In K.U. Mayer & P.B. Baltes (Hrsg.), *Die Berliner Altersstudie* (S. 185-219). Berlin: Akademie Verlag.
- Helmchen, H. & Linden, M. (1993). The Differentiation Between Depression and Dementia in the Very Old. *Ageing and Society*, 13, 589-617.
- Henderson, A.S., Duncan-Jones, P. & Finlay-Jones, R.A. (1983). The reliability of the Geriatric Mental State Examination. *Acta psychiatrica scandinavica*, 67, 281-289.
- Hertzog, C., Van Alstine, J., Usala, P.D. & Hultsch, D.F. (1990). Measurement properties of the Center for Epidemiological Studies Depression Scale (CES-D) in older populations. *Psychological Assessment*, 2, 64-72.
- Himmelfarb, S. & Murrell, S.A. (1983). Reliability and validity of five mental health scales in older persons. *Journal of Gerontology*, 38, 333-339.
- Hsiao, J.K., Bartko, J.J. & Potter, W.Z. (1989). Diagnosing Diagnoses. *Archives of General Psychiatry*, 46, 664-667.
- Kessel, J.B. & Zimmerman, M. (1993). Reporting Errors in Studies of the Diagnostic Performance of Self-Administered Questionnaires: Extent of the Problem, Recommendations for Standardized Presentation of Results, and Implications for the Peer Review Process. *Psychological Assessment*, 5, 395-399.
- Liang, J., Tran, T.V., Krause, N. & Markides, K.S. (1989). Generational differences in the structure of the CES-D scale in Mexican Americans. *Journals of Gerontology*, 44, 110-120.
- Linden, M., Borchelt, M., Barnow, S. & Geiselman, B. (1995). The impact of somatic morbidity on the Hamilton Depression
- Rating Scale in the very old. *Acta psychiatrica scandinavica*, 92, 150-154.
- Linden, M. & Geiselman, B. (1996). Sudiagnostische psychiatrische Morbidität. Beschwerdeprofil und Konsequenzen am Beispiel depressiver Störungen. In H. Sass (Hrsg.), *Psychopathologische Methoden und psychiatrische Forschung* (S. 106-116). Jena: Gustav Fischer Verlag.
- Lindenberger, U., Gilberg, R., Pötter, U., Little, T.D. & Baltes, P.B. (1996). Stichprobenselektivität und Generalisierbarkeit der Ergebnisse der Berliner Altersstudie. In K.U. Mayer & P.B. Baltes (Hrsg.), *Die Berliner Altersstudie* (S. 85-108). Berlin: Akademie Verlag.
- Mahard, R.E. (1988). The CES-D as a measure of depressive mood in the elderly Puerto Rican population. *Journals of Gerontology*, 43, 24-25.
- Mayer, K.U. & Baltes, P.B. (Hrsg.) (1996). *Die Berliner Altersstudie*. Berlin: Akademie Verlag.
- McCallum, J., Mackinnon, A., Simons, L. & Simons, J. (1995). Measurement properties of the Center for Epidemiological Studies Depression Scale: An Australian community study of aged persons. *Journals of Gerontology Series B Psychological Sciences and Social Sciences*, 50B, 182-189.
- McWilliam, C., Copeland, J.R.M., Dewey, M.E. & Wood, N. (1988). The Geriatric Mental State Examination As a Case-finding Instrument in the Community. *British Journal of Psychiatry*, 152, 205-208.
- Mossman, D. & Somoza, E. (1989). Maximizing Diagnostic Information From the Dexamethason Suppression Test. *Archives of General Psychiatry*, 46, 653-660.
- Murphy, J.M., Berwick, D.M., Weinstein, M.C., Borus, J.F., Budman, S.H. & Klerman, G.L. (1987). Performance of Screening and Diagnostic Tests. *Archives of General Psychiatry*, 44, 550-555.
- Myers, J.K. & Weissman, M.M. (1980). Use of a self-report symptom scale to detect depression in a community sample. *American Journal of Psychiatry*, 137, 1081-1084.
- Newmann, J.P. (1989). Aging and Depression. *Psychology and Aging*, 4, 150-165.
- Orme, J.G., Reis, J. & Herz, E.J. (1986). Factorial and discriminant validity of the Center for Epidemiological Studies Depression (CES-D) Scale. *Journal of Clinical Psychology*, 42, 28-33.

- Radloff, L.S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401.
- Radloff, L.S. & Teri, L. (1986). Use of the Center for Epidemiological Studies Depression Scale with older adults. *Clinical Gerontologist*, 5, 119-136.
- Roberts, R.E., Rhoades, H.M. & Vernon, S.W. (1990). Using the CES-D scale to screen for depression and anxiety: Effects of language and ethnic status. *Psychiatry Research*, 31, 69-83.
- van de Vijver, F. & Hambleton, R.K. (1996). Translating Tests: Some Practical Guidelines. *European Psychologist*, 1, 89-99.
- Watkins, D. (1989). The Role of Confirmatory Factor Analysis in Cross-Cultural Research. *International Journal of Psychology*, 24, 685-701.
- Weyerer, S., Geiger-Kabisch, C., Denzinger, R. & Pfeifer-Kurda, M. (1992). Die deutsche Version der CES-D Skala. Ein geeignetes Instrument zur Erfassung von Depressionen bei älteren Menschen? *Diagnostica*, 38, 354-365.
- Zemore, Z. & Eames, N. (1979). Psychic and Somatic Symptoms of Depression Among Young Adults, Institutionalized Aged and Noninstitutionalized Aged. *Journal of Gerontology*, 34, 716-722.
- Zimmerman, M. & Coryell, W. (1994). Screening for major depressive disorder in the community: A comparison of measures. *Psychological Assessment*, 6, 71-74.
- Zung, W.W.K. (1965). A self-rating depression scale. *Archives of General Psychiatry*, 12, 63-70.

Korrespondenzadresse:

Michaela Riediger
Max-Planck-Institut
für Bildungsforschung
Lentzeallee 94
14195 Berlin