Reference	Sample	Measurement Instrument			Psychometric Characteristics	
		Name of the Scale	Domains and Constructs	Length and Format of Instrument	Validity	Reliability
Kosberg *et al.*(1990)[6] United States	ADRD	The Cost of Care Index (CCI)	Consequences (or costs) of caretaking Five factors/components: (1) Personal and Social Restriction; (2) Physical and Emotional Health; (3) Value of Providing Care; (4) Patient as Provocateur; (5) Economic Costs	20 items, 4-point Likert scale (ranging from 1=Strongly agree to 4=Strongly disagree)	Content validity was established by first reviewing the literature related to the "costs" of providing care to dependent elderly persons and defining the dimensions needed to develop a bank of items. Twenty-seven items were initially derived from the input of professionals working with family CGs of frail and impaired elderly relatives. These 27 items were pilot tested to determine their ability to distinguish groups of carers caring for elders with different physical and mental impairment levels. The initial inspection of the structural validity of the CCI was not conducted in a sample of dementia CGs. Instead, authors recruited a sample of 137 CGs of clients seeking nursing home placement under the Florida Medicaid Program. A PCA with Varimax rotation yielded a final 20-item scale with 5-components/factors. The factorial structure of the CCI scale was not examined in the present study of dementia CGs. Concurrent validity was demonstrated by statistically significant Pearson's correlations between CCI scores and, for example, measures of caregiving functioning assessed by a) the Short Psychiatric Evaluation Schedule (SPES; r=0.27, p<0.01), b) self-reported mental health (r=0.36, p<0.001), and c) physical health (r=0.22, p <0.05). Significant correlations were also obtained between CCI scores and measures of "consequences of caregiving" assessed by a) the "ADL trouble due to patient" (0.24, p < 0.01) and b) "Tolerance for patient behavior" (-0.33, p < 0.001).	Cronbach's α, full scale=0.79
Goodman (1991)[7] United States	ADRD	Perceived Social Support for Caregiving (PSSC)	Perceived social support One factor: Availability or adequacy of social support (or help)	9 items, 5-point Likert scale (ranging from 1=Not at all to 5=Extremely)	Content validity was demonstrated by a team of researchers writing an initial 21-item pool based on a review of literature and empirically-determined reasons for joining self-help groups. Structural validity was established through PCA with Varimax rotation using an initial 21-item scale. Inspection of item loading reduced the scale to 12 items. PCA analyses were repeated yielding two separate scales (each with one factor/component): The Perceived Support for Caregiving (PSSC) and the Social Conflict (SC) scale. The PSSC explained 42.8% of the variance. Concurrent validity was established by significant positive Pearson correlations between PSSC total scores and Natural Network Indices (r=0.26 to 0.39; p<0.02 to 0.001). These results were collected from a subsample of respondents (N=70 to 79).	Cronbach's α, full scale=0.84
		Social Conflict (SC)	Social conflict One factor: Lack or inadequacy of social support (or help)	3 items, 5-point Likert scale (ranging from 1=Not at all to 5=Extremely)	Structural validity. A PCA with Varimax rotation yielded the SC factor/component explaining 18% of the variance. Concurrent validity was established by a significant positive Pearson correlation between the total scores on the SC and the ZBI item "Do you feel that your relative currently affects your relationship with other family members or friends in a negative way" (r=0.34, p=0.001).	Cronbach's α, full scale =0.72
Theut *et al.* (1991)[8] United States	ADRD	Anticipatory Grief Scale (AGS)	Anticipatory grief (bereavement of wives whose spouses had been diagnosed with dementia) Seven domains: (1) Anger; (2) Guilt; (3) Anxiety; (4) Irritability; (5) Sadness; (6) Feelings of loss; (7) Decreased function	27 items, 5-point Likert scale (1=Strongly disagree, 2=Disagree, 3=Somewhat agree, 4=Agree, 5=Strongly agree)	Content validity. Authors report selecting items from previous grief scales, reviewing the literature on the dimensions of anticipatory grief and developing additional items based on clinical experience with wives of patients with dementia. Structural validity. No formal examination of the underlying structure of the scale is presented. Concurrent validity was established by positive and significant (p<0.001) correlations between total scores in the AGS scale and the depression, anxiety, and hostility dimensions of the Hopkins Symptom Checklist (SCL-90-R). Note: Coefficients were not reported.	Cronbach's α, full scale =0.84
Vitaliano *et al.* (1991)[9] United States	ADRD	The Screen for Caregiver Burden (SCB)	Burden or distress Two domains: (1) Objective burden or number of negative experiences; (2) Subjective burden or distress in response to experiences	25 items, Objective burden: 2-point scale (0=Did not Occur, 1=Occurrence) Subjective burden: 5-point scale (ranging from 0=No occurrence to 4=Occurrence with severe distress) (Each item received two ratings: one for objective burden and a second for subjective burden	Content validity was established by a review of extant literature on problems in AD caregiving and by asking a sample of spouse CGs of individuals with AD (N=68) what burden experiences were most salient to them. Items are scored based on two general "domains:" Objective and Subjective. That is, the scaling of the items was assesses both objective burden (OB) and subjective burden (SB). Structural validity. No formal examination of the underlying structure of the scale/subscales is presented. Concurrent validity: The SCB OB and SB subscales correlated significantly (p<0.05) with depression (0.54, 0.41, respectively), anxiety (0.43, 0.26, respectively), and suppressed anger (subjective=0.42). The SCB OB and SB subscales also had significant (p<0.05) negative correlations with morale (-0.51, -0.48, respectively). Depression, anxiety, suppressed anger, and morale were measured by the Beck Depression Inventory-Short (BDI-S), the Symptom Checklist--90 Anxiety Scale (SCL-90), the Suppressed Anger Subscale from the Anger Expression Scale, and the Satisfaction with Life Scale, respectively.	Cronbach's α by subscales: Objective burden (α=0.85) Subjective burden (α=0.89) Test-retest reliability (15-18 months apart) was estimated with Pearson's correlations between scale administrations: Objective burden (r=0.64, p<0.001) Subjective burden (r=0.70, p<0.001)
Given *et al.* (1992)[10]	Mixed	Caregiver Reaction	Reactions to caregiving burden	24 items; 5-point scale (ranging	Content validity was established by requiring that all the members of a 5-group team agree that each item from a 101-item pool belonged to a particular construct. After a pilot test with 99 CGs, items were	Cronbach's α by subscales: Impact on health (α=0.80)

United States		Assessment (CRA)	Five factors: (1) Impact on health (2) Impact on CG's daily schedule; (3) Impact on CG's finances; (4) Relationship to CG's sense of self-worth; (5) Friends/family support	from 1=Strongly disagree to 5=Strongly agree)	dropped due to lack of clarity, variability, or association with any other items reducing the item pool to a 40-item scale. Structural validity. An initial EFA with a sample of 377 participants (29.2% dementia CGs) led to the reduction of the 40-item scale to 35 items. A PCA with oblique rotation yielded a 5-factor/component solution. A scree plot confirmed the 5-factor structure. Further elimination of items due to low loadings resulted in a 24-item scale. A final re-run of the PCA model with the 24-item scale yielded a 5-factor solution accounting for 65.1% of the variance in items. Measurement (factorial) invariance tests. Using a second independent sample (also N=377; 26.8% dementia CGs) a multiple-group CFA was conducted to test the factorial invariance of the CRA instrument across diseases (Alzheimer's vs. cancer), spouse vs non-spouses, and across time. Results showed that the subscale structure of the CRA remains stable across diverse groups of CGs and across time. The latter suggests suitability of the CRA to measure change in CG reactions. Concurrent validity was established by calculating Pearson correlations between CRA subscales and a) the CES-D and b) the ADL Dependencies Scale using the combined independent samples (N=754). Finance (r=0.25, r=0.34), family support (r=0.20, r=0.39), and health (r=0.29, r=0.57) were significantly and positively correlated with ADL dependencies and depression respectively. CG esteem was significantly and negatively associated with depression (r=-0.23).	Impact on schedule (α=0.82) Impact on finances (α=0.81) Sense of self-esteem (α=0.90) Friends/family support (α=0.85)
Semple (1992)[11] United States	ADRD	Family Conflicts Scales (FCS)	Family conflict Three factors: (1) Definitions & strategies conflict; (2) Treatment of patient conflict; (3) Treatment of CG conflict	12 items, 4-point Likert scale (ranging from 1=No disagreement to 4=Quite a bit of disagreement)	Content validity was established by in-depth interviews with 20 CGs that resulted in the identification of three family conflict domains and the creation of 4 items per domain. Structural validity was established through CFA using the 12-item scale. After comparing competing model that conformed underlying theories, a 3-factor model yielded the best fit as measured by a GFI=0.98 and a chi-square/df ratio=2.9 (less than 3 is desirable). As evidence of concurrent validity the authors used Pearson correlations to show relationships between the three FCS factors/subscales and the Hopkins Symptom Checklist (HSC) measures of *anger* and *depression*. All correlations between the HSC-Anger measure and the FCS subscales were significant (p<0.001): Definitions & strategies conflict (r=0.23), Treatment of patient conflict (r=0.25), and Treatment of CG conflict (r=0.34). All correlations between the HSC-*depression* measure and the FCS subscales were also significant (p<0.001): Definitions & strategies conflict (r=0.23), Treatment of patient conflict (r=0.28), and Treatment of CG conflict (r=0.28).	Cronbach's α by subscales: Definitions & strategies conflict (α=0.80) Treatment of patient conflict (=0.86) Treatment of CG conflict (reported as "within the range between the two other subscales")
Teri *et al.* (1992)[12] United States	Mixed	The Revised Memory and Behavior Problem Checklist (RMPBC)	CG reactions to patient behavior problems Three factors/components: (1) Memory-related problems; (2) Depression problems; (3) Disruptive behaviors (The scale uses two scoring methods per item: frequency of patient behavior problems and CG distress or reaction to the patient behavior problems.)	24 items, Each item is rated on two scales. 1) Frequency of patient behavior: 5-point Likert scale (0=Never occurred, 1=Not in the past week, 2=1 to 2 times in the past week, 3=3 to 6 times in the past week, 4=Daily or more often) 2) Reaction of "upset" by CG: 5-point Likert scale (0=Not at all, 1=A little, 2=Moderately, 3=Very much, 4=Extremely)	Content validity was shown by raters sorting items into hypothesized content areas, rating items, and agreeing on items. This method reduced the original pool of 64 items to 47 items. A PCA approach with Varimax rotation was used to study the underlying dimensions of the scale using "frequency scorings". The analysis yielded a 24-item, 3-component/factor scale explaining 53.4% of the variance. Concurrent validity was examined calculating Pearson correlations between RMPBC subscales and *well-known* (benchmark) scales measuring similar constructs. Correlations were estimated separately by scoring method Behavior Frequency and Caregiver Reaction. For the Behavior Frequency scoring, significant positive Pearson correlations were obtained between the RMPBC Depression subscale and the HAM-D Scale (r=0.44, p<0.01) as well as between the RMPBC Memory-Related Problems subscale and the MMSE. For the Caregiver Reaction scoring, validity was demonstrated by significant positive Pearson correlations between all RMPBC subscales and the CES-D scale and the Caregiver Stress Scale (CSS) (all p-values < 0.01). Depression measured by the CES-D correlated with the RMPBC Memory-Related Problems (r=0.29), Depression (r=0.31), and Disruption (r=0.26) subscales. Burden measured by the CSS correlated with the RMPBC Memory-Related Problems (r=0.32), Depression (r=0.42), and Disruption (r=0.41) subscales. Discriminant validity was established for RMPBC Behavior Frequency by non-significant correlations between the RMPBC Depression subscale and the Mini-Mental State Exam (r=-0.04, p>0.05) as well as non-significant correlations between the RMPBC Memory-Related problems subscale and the HAM-D Scale (r=0.001, p>0.05).	Behavior *Frequency* Scoring: Cronbach's α, full scale =0.84 Cronbach's α by subscales: Depression (α=0.80) Memory-Related problems (α=0.79) Disruption (α=0.67) Caregiver *Reaction* Scoring: Cronbach's α, full scale =0.90 Cronbach's α by subscales: Depression (α=0.89) Memory-Related problems (α=0.88) Disruption (α=0.84)
Macera *et al.* (1993)[13] United States	ADRD	Caregiver Burden Scale (CBS)	Perceived burden Three domains: (1) Activity for which patient required help; (2) Activity for which CG provided help; (3) Stress by providing help	15 items, 2-point scale (0=No, 1=Yes)	Content validity and the creation of items was not addressed in the article. Authors reviewed the literature on perceived burden and state the importance of measuring burden associated with specific caregiving tasks. Results of the authors-developed CBS scale are presented as a pilot study. Structural validity. No examination of the underlying structure of the scale is presented. Concurrent validity for the CBS scale was established by a significant positive Pearson correlation with the CES-D (r=0.38, p<0.001).	Cronbach's α, full scale =0.87
Gerritsen *et al.* (1994)[14]	ADRD	Care-Giving Burden Scale	Subjective burden Two factors:	13 items, 5-point Likert scale	Content validity was appraised by researchers and colleagues screening items for caregiving burden from previous scales, in particular, the Vernooij-Dassen's Sense of Competence Questionnaire. The	Cronbach's α, full scale =0.84. Cronbach's α by subscales:

The Netherlands		(C-GBS)	(1) Personal consequences (subjective impact of care-giving on the lives of the carers) (2) Relationship (evaluation/opinions of the care relationship)	(1=Disagree very much, 2=Disagree, 3=Agree on the one hand, disagree on the other, 4=Agree, 5=Agree very much) Note: Items were recoded to binary, 2-point scale (1,2=0; 3,4,5=1)	screening process reduced the original 27-item Sense of Competence scale, as well as additional author-developed items, to a final pool of 20 items. Structural validity was established through a PCA with Varimax rotation. The analysis yielded a two-factor/component solution that explained 34.4% of the variance. (A replication of the PCA at a second time point (after 3 months) produced similar results explaining 37.6% of the variance.) Based on these results and an inspection of item loadings, authors further reduced the 20-item scale to a 13-item scale. The concurrent validity was established by statistically significant (p<0.001) positive Pearson correlations between the C-GBS scores and CG depression measured by the Zung Self-Rating Depression Scale (r=0.53). C-GBS scores were significantly associated with both, patient deviant behavior and memory/orientation subscales from the RMBPC (r=0.53 and 0.31, respectively).	Personal consequences (α=0.74) Relationship (α=0.77) Note: Reliability estimates from an independent sample of CGs (N=42) were similar (full scale α=0.84; Subscales: Relationship α=0.77, Personal Consequences α=0.75)
Gilleard *et al.* (1994)[15] United Kingdom	ADRD	Dementia Quiz (DQ)	Dementia knowledge Three domains: (1) Biomedical knowledge; (2) Services knowledge; (3) Coping knowledge	25 items, 5-point, multiple-choice scale (including a fifth "don't know" option)	Content validity. Thirty-six items were gathered from unpublished questionnaires, the original Alzheimer's Disease Knowledge Test (ADK), and experience working with health care staff and families of those with dementia. Several "experts" experienced in aging and mental health guided the rewording and reduction of the item pool to 34 items. To provide evidence of face validity, a panel of 10 experts pilot tested the 34-item scale, and all items were scored 'correct' by at least eight out of 10. Structural validity. No formal analysis to study the underlying structure (dimensionality) of the 34-item scale is conducted. Authors reported further reducing the scale to 25 items due to low item-subscale (domain) correlations (r's < 0.25). Concurrent validity was established by correlating Dementia Quiz (DQ) scores with the Alzheimer's Disease Knowledge Test (ADK). The results indicated highly significant associations between the ADK score and the three DQ subscale scores: Biomedical Knowledge subscale (r= 0.59, p < .001); Services Knowledge subscale (r= 0.37, p < .001); and Coping Knowledge subscale (r= 0.52, p < .001).	Cronbach's α, full scale =0.88 Spearman-Brown (SB) split-half reliability estimate for subscales: Biomedical Knowledge (SB=0.78) Services Knowledge (SB=0.71) Coping Knowledge (SB=0.71)
Hinrichsen & Niederehe (1994)[16] United States	ADRD	The Dementia Management Strategies Scale (DMSS)	Management strategies Three factors: (1) Managing criticism; (2) Encouragement; (3) Active management	28 items, 5-point Likert scale (1=Never, 2=Seldom, 3=Sometimes, 4=Often, 5=Most of the time)	Content validity was established in a prior study by Niederehe & Scott (1987).[17] A 34-item pool was developed based on literature reviews, clinical work with dementia patients and family members, and pilot interviews. Structural validity was established through EFA using PAF extraction and Varimax rotation that yielded a 3-factor solution. The original 34-item pool was reduced to 28 items based on factor loadings.	Cronbach's α by subscales: Criticism (α=0.85) Encouragement (α=0.80) Active management (α=0.77)
Carruth (1996)[18] United States	Mixed	Caregiver Reciprocity Scale (CRS)	CG reciprocity Four factors: (1) Warmth and regard; (2) Intrinsic rewards for giving; (3) Love and affection; (4) Balance within family caregiving	26 items, 5-point Likert scale (ranging from 1=Strongly disagree to 5=Strongly agree)	Content validity was established by an initial 50-item pool developed from a literature review and interviews with family CGs. Two panels of experts rated items relevance and CVIs were computed. Items with low CVIs were eliminated reducing the pool to 32 items. A pilot test with 30 CGs provided data for further item reduction by "item-to-item", "item-to-subscale", and "item-to-total" correlations further reducing the scale to 30 items. Structural validity. Before attempting to establish validity, an inter-item analysis dropped four poorly-correlated items reduce the 30-item scale to 26 items. The sample (N=303) was randomly split into two subsamples to perform EFA (N=130) and CFA (N=173). An EFA performed by factor analysis with Varimax rotation yielded a 22-item, 4-factor solution that accounted for 62.9% of the variance. The CFA with the cross validation sample established the acceptability of the 4-factor model with adequate fit indexes (e.g., GFI=0.88; AGFI=0.85; RMR=0.05; TLI=0.95). AVE was used to assessed the convergent validity of the 4 factors extracted by CFA. AVE values ranged from 0.47 to 0.64. (Three of the four AVEs were slightly below the recommended threshold of 0.50.)	Cronbach's α by subscales: Warmth and regard (α=0.89) Intrinsic rewards for giving (α=0.82) Love and affection (α=0.86) Balance within family caregiving (α=0.78) Test-retest reliability was estimated using Pearson's correlations with a convenience sample of N=35 who retested 2 weeks after the initial test. Test-retest reliability by subscales: Warmth and regard (r=0.70); Intrinsic rewards for giving (r=0.69); Love and affection (r=0.88); Balance within family caregiving (r=0.58)
Keady & Nolan (1996)[19] United Kingdom	ADRD	Behavioral and instrumental stressors in Dementia (BISID)	CG stress Three domains: (1) Behavior of patient (2) Activities of daily living (ADL) (3) Continence	22 items, (Each item is rated using the scale below and also according to "Way of coping" and "Perceived stress level") Ratings for Behavior and Continence domains: 5-point Likert scale (from 0=Never, to 4=Very frequently (> once a day) Ratings for the ADL domain: 4-point Likert scale (from 0=No help needed to	Content validity. Items were drawn from a comprehensive review of the literature on CGs needs and stressors, existing measures, and experiences of local dementia professionals. A pilot study of the 22 items with 38 dementia CGs confirmed the scale's content acceptability to CGs. Structural validity. No formal examination of the underlying factor structure of the scale using factor analysis is presented.	Cronbach's α estimates from the BISID subscales were obtained from *two independent samples*. The first sample comprised 205 caretakers and the second independent sample included 264 caretakers. Cronbach's α by subscales (N=205): Behavioral (α=0.89). ADL (α=0.90) Continence (α=0.92) Cronbach's α by subscales in the *second independent sample* (N=264) were very close and also within acceptable ranges:

				3=Totally unable to complete the activity) Ratings for "Perceived stress level" 4-point Likert scale (from 0=Not stressful to 3=Very stressful)		Behavioral (α=0.92) ADL (α=0.92) Continence (α=0.94)
Vernooij-Dassen *et al.* (1996)[20] The Netherlands	ADRD	Sense of Competence Questionnaire (SCQ)	Feelings of competence Three factors: (1) Satisfaction with the demented patient; (2) Satisfaction with one's CG performance; (3) Consequences of caregiving for one's personal life	27 items, 4-point Likert scale (1=Disagree Very Much, 2=Disagree, 3=Agree, 4=Agree Very Much)	Content validity was determined through classification of items by a 39-person panel of experts. Structural validity was established through EFA. Authors reported conducting an EFA that yielded the same 3-factor structure that the panel of experts had previously predicted. No further details of the EFA extraction procedures were provided. Note: The 7-item abbreviated version of the SCQ scale (S-SCQ) developed later by Vernooij-Dassen *et al.* (1999)[21] also produced the same 3-factor structure through an EFA. Using the same sample of CGs, authors found significant Pearson's correlation between the S-SCQ and the original SCQ (r=0.88).	Cronbach's α, full scale =0.79 Cronbach's α by subscales: Satisfaction with the demented patient (α=0.55); Satisfaction with one's CG performance (α =0.63); Consequences of caregiving for one's personal life (α = 0.50) (Cronbach's α for the abbreviated 7-item S-SCQ scale=0.76.)
Davis *et al.* (1997)[22] United States	ADRD	Caregiver Activity Survey (CAS)	Time spent in caregiving activities (One "total score" measure) Scores were the hours and minutes engaged in the activity during a 24 hour period (day and night) during weekdays.	6 items, The six items included: (1) communicating; (2) using transportation (3) dressing; (4) eating (5) looking after one's appearance; (6) supervising	Content validity. Special efforts were made to find terms that could be used with a variety of populations in different cultures. Several versions of the scale were pilot-tested. in different cultural settings. Specialists reviewed the scale drafts to develop cultural and linguistic equivalents in several languages. Specific definitions were provided, along with examples of what was meant for the various categories of assistance. As a result, an initial pool of 13-items was developed and tested. The results of the first analysis led to the reduction of the original 13-item scale to a 6-item CAS scale. The concurrent validity of CAS was established by significant (p-values < 0.05) Pearson correlations with the Alzheimer's Disease Assessment Scale Cognitive Subscale (ADAS-Cog) (r = 0.61), MMSE (r = −0.57) and Physical Self Maintenance Scale (PSMS) (r=0.43).	Test-retest reliability was established by retesting N=42 CGs within a 2-week interval (i.e., week 1 and week 3), and calculating the ICC. The ICC=0.85, p<.001.
Picot *et al.* (1997)[23] United States	ADRD	Picot Caregiver Rewards Scale (Picot-CRS)	Perceived CG rewards Two domains/subscales: (1) External rewards: communication from patient, health care professionals, or other entity regarding quality of care of the caregiving (2) Internal rewards: feelings of achievement and growth	24 items, 5-point Likert scale (0=Not at all, 1=A little, 2=Somewhat, 3=Quite a lot, 4=A great deal)	Content validity was established by interviews with eight family CGs to identify themes about positive feelings and changes (i.e., rewards) resulting from caregiving. Twenty-seven items were generated from caregiving literature and considering caregiving's external and internal rewards. A pilot test with 20 CGs led to a reduction from 27 to 24 items. The underlying factorial structure of the scale was not examined. Concurrent validity was demonstrated by a significant positive Pearson correlation between PCRS scores and "caregiving demands" (r=0.22, p<0.05) measured by Texas Research Institute of Mental Sciences Behavioral Problem Checklist (TRIMS BPC) as well as by a significant positive association between PCRS scores and palliative coping (r=0.26, p<0.05) measured by the Jalowiec Coping Scale. A hypothesized negative association between rewards and costs as measured by the Costs of Care Index (CCI) was not found (r=0.07, p>0.05).	Cronbach's α, full scale =0.83
Schoefield *et al.* (1997)[24] Australia	Mixed	Comprehensive instrument to assess the experience of caregiving: A battery of scales Scale 1: Social Support	Social Support Three factors/components: (1) Family support; (2) Friends support; (3) Esteemed by family and friends	7 items, 5-point Likert scale (ranging from 1=Strongly disagree to 5=Strongly agree)	Content validity was demonstrated by reviewing literature and instruments and conducting interviews with CGs to generate key domains and a preliminary bank of items. A pilot test with 98 CGs that included a comparison group of 78 non-CGs was also conducted that further refined the initial item pool. Items were organized into five main domains for analysis. The structural validity of the seven-item scale administered to CGs was determined through a PCA with Varimax rotation yielding a three-factor structure accounting for 66% of the variance. (All the scales in the battery were analyzed using the same sample, N=976).	Cronbach's α by subscales: Family support (α=0.64) Friend's support (α=0.57) Esteemed by family and friends (α=0.56)
		Scale 2: Family environment	Family environment Two factors/components: (1) Closeness; (2) Conflict	6 items, 3-point Likert scale (1=Less, to 3=More)	The structural validity for the 6-item scale administered to CGs was determined through a PCA with Varimax rotation yielding a 2-factor/component model explaining 63% of the variance.	Cronbach's α by subscales: Closeness (α=0.68) Conflict (α=0.70)
		Scale 3: Caring role	Caring role Three factors/components: (1) Satisfaction/Love; (2) Resentment; (3) Anger	16 items, 5-point Likert scale (1=Strongly disagree to 5=Strongly agree)	The structural validity for the 16-item scale administered to CGs was assessed through a PCA with Varimax rotation that produced a 3-factor/component structure explaining 44.2% of the variance.	Cronbach's α by subscales: Satisfaction (α=0.71) Resentment (α=0.69) Anger (α=0.71)
		Scale 4: Help Needed by Recipient	Help needs by care recipient Two factors/components: (1) ADLs; (2) IADLs	11 items, 3-point Likert scale (from 1=No help, 2=Some help, 3=A lot of help)	The structural validity of the 11-item scale administered to CGs was evaluated through a PCA with Varimax rotation that resulted in a 2-factor/component solution accounting for 57.1% of the variance.	Cronbach's α by subscales: ADL (α=0.82) IADL(α=0.68)
		Scale 5: Behavior Problem	Behavior problems Three factors/components: (1) Aggressive; (2) Depressive; (3)	18 items, 4-point scale (0=Never, 1=Rarely, 2=Sometimes, 3=Often)	Finally, the structural validity for the 18-item scale administered to CGs was determined through a PCA, also with Varimax rotation that produced a 3-factor/component solution accounting for 41% of the variance.	Cronbach's α by subscales: Aggressive (α=0.84) Depressive (α =0.60) Forgetfulness/Confusion (α =0.73)

			Forgetfulness/confusion			
Kaufer *et al.* (1998)[25] United States	ADRD	The Neuropsychiatric Inventory Caregiver Distress (NPI-D) Scale	Subjective CG distress One domain: Psychological distress Note: The NPI-D was developed as an adjunct scale of Neuropsychiatric Symptoms (NPI) scale. The NPI-D assesses the impact of neuropsychiatric symptoms in Alzheimer's disease (AD) patients on CG distress.	10 items, 6-point Likert scale (from 0=Not at all distressing to 5=Very distressing) Note: The 10 items represent symptoms obtained from one of the three subscales from the original NPI scale. Items assess AD CG distress for each of these symptoms.	Content validity. A preliminary version of the NPI-D included items from the three subscales contained in the original NPI scale (physical, social, and psychological distress). An initial field testing of the NPI-D scale revealed that AD CGs viewed the impact of neuropsychiatric disturbances primarily in terms of psychological or emotional distress. As a result, a revised version of the NPI-D scale excluded items from the physical and social subscales. No examination of the underlying factorial structure of the NPI-D scale is presented. Concurrent validity was established by correlating scores in the NPI-D scale in a subsample (N=69) with scores in an abridged version of the Relatives' Stress Scale (RSS) using Pearson's correlation (r=0.60, p<0.001). (The abridged RSS included 2 of the 3 subscales: personal distress and negative feelings.) The correlation between total NPI and NPI-D scores was 0.83 (p < 0.001).	Test-retest reliability was established by retesting 23.5% (N=20) of the CGs within an average of 4.5 days and correlating the scores using Pearson's correlation (r=0.92, p<.001). Interrater reliability was also calculated with the ICC between two raters of the NPI-D in 16 CGs (ICC=0.96, P < 0.001).
Zeiss *et al.* (1999)[26] United States	Mixed	Caregiver Self-Care Self-Efficacy	Self-care self-efficacy One domain: CG behaviors that reduce stress and enhance well being	10 items, Rating of confidence in performing item activity (ranging from 0%=No confidence to 100%=Completely confident).	Content validity was established through literature reviews and authors' own experiences working with CGs resulting in the development of items for two separate measures: Self-care self-efficacy and Problem-solving self-efficacy. The measures were pretested with ten CGs to improve the clarity of wording and to decide the best method for administration. As the result of the pretesting, the measures were administered by interview rather than a paper-and-pencil format. Both measures were field tested in the same sample of 217 CGs. No examination of the underlying structure of the scales is presented. Concurrent validity for the Self-Care Self-Efficacy scale was established by a significant positive Pearson correlation between Self-Care and the "network size" subscale of the Arizona Social Support Interview (r=0.30, p<.001).	Cronbach's α, full scale =0.76. Test-retest reliability with a subsample (N=39) retested after 11 weeks was a high Pearson coefficient (r=0.675, p<0.001).
		Problem-Solving Self-Efficacy	Problem-solving self-efficacy One domain: CG behaviors utilizing problem-solving skills shown to be positively related to psychological adjustment	4 items, Rating of confidence in performing item activity (ranging from 0%=No confidence to 100%=Completely confident).	Concurrent validity for the Problem-Solving Self-Efficacy scale was established by a significant positive Pearson's correlation between Problem Solving and the Logical Analysis subscale of the Daily Living Questionnaire (r=0.19, p<0.05).	Cronbach's α, full scale =0.83. Test-retest reliability with a subsample (18%) retested after 11 weeks was a high Pearson coefficient (r=0 .683, p<0.001).
Farran *et al.* (1999)[27] United States	ADRD	Finding Meaning Through Caregiving Scale (FMTCS)	Positive aspects of caregiving Three factors: (1) Loss/Powerlessness (LP); (2) Provisional meaning (PM); (3) Ultimate meaning (UM)	43 items, 5-point Likert scale (ranging from 1=Strongly disagree to 5=Strongly agree)	Content validity was demonstrated through a preliminary qualitative study of family CGs of demented patients. Their answers to open-ended questions became six major themes from which the authors developed an initial 135-item pool for the FMTCS measure. A pilot study (N=46) established preliminary psychometric properties for a shortened 43-item FMTCS. The shortened FMTCS measure resulted from the examination of item-to-scale, item-to-item, and item-to-total correlations. This pilot produced "acceptable" reliability estimates both by three factors/subscales (0.88 to 0.95) and total scale (0.91). The pilot test-retest reliability (one-month interval), estimated with Spearman correlation, ranged from 0.85-0.89 for the three subscales and 0.80 for the full FMTCS. Given that the three original subscales had a "strong" theoretical base, the authors used CFA to establish the factorial validity of the FMTCS using an independent sample of N=215 caretakers (only N=208 had available data on the FMTCS). The CFA model confirmed the 3-factor structure identified in the previous pilot study and provided an adequate overall fit (e.g., GFI=0.998 and a coefficient of determination=0.763). Concurrent validity was established by Pearson's correlations between FMTCS scores and existing measures hypothesized to be related. Scores on the LP subscale were significantly (all p-values < 0.01) correlated with scores on a) Patient Problem behaviors (r=0.44), b) Marital tension (r=0.38), c) Global role strain (r=0.70), and d) depression (r=0.61), as measured by the CES-D. Scores on the PM subscale were significantly correlated with a) Marital satisfaction (r=0.24), b) Caregiver Satisfaction (r=0.64), and c) Personal gain(r=0.57). Scores on the UM subscale were significantly correlated with a) Religious participation (r=0.53), b) Personal religion beliefs (r=0.61), and c) Religious support satisfaction (r=0.24). Total FMTSC scores revealed similar relationships. FMTSC total scores were positively associated with measures of a) Marital satisfaction (r=0.46), b) Caregiver satisfaction (r=0.58), c) Personal gain (r=0.39), c) Religious participation (r=0.37), d) Religious beliefs (r=0.54), and e) Religious support satisfaction (r=0.21). Total FMTSC scores, however, were *negatively associated* with Patient Problem behaviors (r=-0.35), Marital tension (r=-0.49), Role strain (-0.64), and Depression (-0.60).	Cronbach's α, full scale =0.91 Cronbach's α by subscales: Loss/Powerlessness (LP) (α=0.89) Provisional Meaning (PM) (α =0.88) Ultimate Meaning (UM) (α = 0.91)
Matsuda	ADRD	Subjective	Subjective burden	14 items,	The content validity of the SBS scale is not formally addressed by the author. However, a prior	Cronbach's α, full scale =0.87

(1999)[28] Japan		Burden Scale (SBS)	Three domains: (1) Wellbeing of CG (emotional, physical, social, and financial); (2) Wellbeing of CG's family; (3) Interpersonal stress among relatives	5-point Likert scale (0=No, 1=Yes, a little bit, 2=Yes, to some degree, 3=Yes, to much degree, 4=Yes, very much)	publication by the same author[29] described the development of items for the tool based on literature reviews on stress and coping theories as well as clinical experiences. Development of items also addressed differences in family context unique to Japan. For example, there is a higher proportion of three-generation households and daughter-in-law CGs in Japan with CG stressors unique to family members and relationships. No examination of the underlying factorial structure or dimensionality of the 14-item scale is presented. Concurrent validity was assessed by calculating a Pearson's correlation coefficient between the SBS total scores and a mental health criterion measured by the General Health Questionnaire (GHQ) (r=0.41, p<0.001). Group discriminant validity was established by comparing SBS scores for CGs with high scores in the GHQ (17 or higher-MU group) vs CGs with low GHQ scores (16 or under-MH group) using a t-test. The MU group showed significantly higher SBC scores that the MH group (t=5.45, p < 0.001).	Split-half reliability of the full scale was estimated using the Spearman-Brown formula (r= 0.80). Rest-retest reliability (6-month interval) was calculated in a subsample (N=50) using a Pearson's correlation coefficient (r=0.72).

Note: AD = Alzheimer's disease; ADRD = Alzheimer's disease and related dementias; ADL = Activities of Daily Living; AGFI = adjusted goodness-of-fit index; AVE = average variance extracted. A recommended threshold for convergent validity is an AVE > 0.50; CG = Caregiver; CATPCA = categorical principal component analysis; CES-D = Center for Epidemiological Studies Depression Scale; CFA = confirmatory factor analysis; CFI = comparative fit index; CR = composite reliability. A recommended threshold for convergent validity is a CR > 0.70; CVI = content validity index;[191] EFA = exploratory factor analysis; GFI = goodness of fit index; Hamilton Depression Rating Scale = HAM-D; Hospital and Anxiety Depression Scale = HADS; IADL = instrumental activities of daily living; ICC = Intra-class correlation coefficient; IFI = incremental fit index; IRT = item response theory; LSNS= Lubben Social Network Scale; ML = maximum likelihood; MLE = maximum likelihood estimation; MMSE = Mini-Mental State Examination; NPI = Neuropsychiatric Inventory; NFI = Normed Fit Index; NNFI = non-normed fit index; PAF = principal axis factoring; PCA = principal components analysis; POMS= Profile of Mood States; RMPBC = Revised Memory and Behavior Problems Checklist; RMSEA = root mean square error of approximation; SF-36 = Short form 36 Health Survey; SRMR = standardized root-mean-square residual; TLI = Tucker-Lewis Index; ZBI = Zarit Burden Interview; PSI = person separation index.[192] PSI values above 0.70 indicate good to excellent reliability in differentiating persons along the measured trait. Proposed rule of thumb thresholds for ICCs are: between 0.50 and 0.75 (moderate); ≥ 0.75 (good), and ≥ 0.90 (excellent).[193] Generally accepted threshold for "good" Cronbach's α test of reliability is considered to be ≥ 0.70. Responsiveness (longitudinal validity) refers to the ability of an instrument to detect clinically important changes over time.[194] Measures such as minimal important change (MIC), smallest detectable change (SDC), effect size (ES), and area under the receiver operating curve (ROC) can be used to describe responsiveness.