

Capitolo 7

Relative Stress Scale e analisi di Rasch. Contributo all'«oggettivazione» della persona negli studi socio-sanitari

di Piergiorgio Lovaglio*; Giorgio Casale** e Anna***

7.1. Introduzione

Sono oltre 7 milioni le famiglie italiane che hanno un anziano in casa. Circa 2 milioni e mezzo di esse si prendono cura di un disabile. Oltre il 50% degli ultraottantenni presenta problemi di disabilità. Rispetto a cinquant'anni fa la popolazione anziana è pressoché raddoppiata. L'esercito di coloro che gratuitamente si prendono cura di anziani e disabili in genere è costituito dal 20,5% della popolazione con più di 14 anni. Questo esercito è formato da figli, parenti, amici o da volontari che impegnano mediamente 331 milioni di ore al mese e svolgono un lavoro pari a un milione e mezzo di euro l'anno. I 2/3 di queste ore sono a carico delle donne, soprattutto le attuali quarantenni e le loro madri. Le donne dedicano alle attività di aiuto 12 ore al mese e gli uomini 8. Nel 78,4% dei casi gli aiuti vengono erogati a un solo individuo. L'età dei *caregiver*, così viene chiamato chi aiuta, è generalmente compresa tra i 55 e i 64 anni. Il 30% di essi fornisce aiuto alla propria famiglia di origine o a quella del partner (Sabbadini, 2001). Il *caregiving*, cioè il prendersi cura come aiuto informale, è un processo non soltanto di

* Ricercatore, Dipartimento di Statistica, Facoltà di Statistica, Università degli Studi di Milano Bicocca.

** Direttore medico, Istituto Geriatrico «Piero Redaelli», Milano.

*** Statistico, IRCCS Fondazione Salvatore Maugeri, Pavia.

rilevanti dimensioni nel mondo degli anziani, ma è anche oggetto di sempre maggiore interesse da parte degli studiosi. La banca dati del Centro Maderna¹ di documentazione, formazione e ricerca sulla condizione anziana raccoglie, al gennaio 2003, circa 1000 contributi e pubblicazioni sul caregiving in età avanzata.

Il caregiving dell'anziano è una trasformazione della mutua relazione di aiuto, che esiste tra moglie e marito, tra genitori e figli o tra amici, a causa della comparsa di malattie o del venir meno dell'autosufficienza nel compiere gli atti della vita quotidiana. La relazione tra i componenti della famiglia viene in un certo senso «occupata», «riempita» dal processo di caregiving: il caregiving, che è fatto di «gesti che curano», da significante diventa significato della relazione stessa, o meglio «sostanza» tra chi dà e chi riceve. Per questi motivi il caregiving ha una sua storia, spesso inattesa, della durata di anni, che, ad esempio, per la demenza ha una mediana intorno ai 6,5 anni (Aneshensel, Pearlin, Mullan *et al.*, 1995), ma anche una sua gravosità, un carico (*burden*) con effetti sia sul ricevente sia sul donatore e con dimensioni che includono aspetti fisici, economici, sociali e psicologici.

Lo stress, che nell'inglese del XVII secolo significava difficoltà, avversità, afflizione, ha acquistato successivamente il significato di forza o tensione o sforzo, per indicare più recentemente lo stato di tensione o resistenza di una persona (o di un sistema biologico o di un oggetto) che si oppone a forze esterne che agiscono contro di lui. Esistono numerose interpretazioni dello stress del caregiver. L'approccio basato sulla transazione dello stress di Lazarus e Folkman (1984), che è la più diffusa, postula che sia la personalità del caregiver, in particolare il suo grado di autostima (*self-esteem*) e di senso di controllo (*mastery over life*), sia la situazione (*situational factors*), mediati dalla valutazione cognitiva e dalle tecniche di *coping* (strategie di superamento) messe in atto, influenzano la risposta emotiva del *carer* e gli *outcomes*. Lo stress del caregiver può diventare distress, ferita o malattia, oppure divenire, come noi pensiamo, nella pienezza del donarsi, valore e significato della relazione d'aiuto. Il complesso modello concettuale di Pearlin *et al.* (1990) identifica, in particolare, quattro domini del caregiving – il contesto, i fattori di stress, i mediatori e gli esiti (tabella 1) – e integra i sistemi fisiologici, psicologici e sociali che lo controllano.

Nel rapporto fra *life events* e fattori di stress è fondamentale tenere conto delle risorse che sono disponibili al caregiver. Le risorse disponibili possono infatti intervenire nel caregiving modificando le condizioni che sono problematiche, influenzando sia sullo stressor sia sulla risposta emotiva del caregiver. L'autostima e il supporto sociale possono, ad esempio, modificare la dinamica dello stress (*distress-deterring theories*).

¹ <http://www.centromaderna.it>.

Tabella 1 – Elementi del modello di stress del caregiver secondo Pearlin *et al.*, 1990. Sono state aggiunte alcune integrazioni

FATTORI DI STRESS
<p>a) Fattori primari</p> <ul style="list-style-type: none"> · Indicatori oggettivi (originano direttamente da chi riceve il <i>caregiving</i>): disturbo dello stato cognitivo, disturbi psicologici e del comportamento, disturbo dello stato funzionale (grado di autosufficienza) · Indicatori soggettivi (percezione che ha il <i>career</i>): sovraccarico e tensione, isolamento, deprivazione relazionale <p>b) Fattori secondari di stress (di ruolo)</p> <ul style="list-style-type: none"> · Conflitti familiari · Conflitti con il lavoro · Problemi economici <p>c) Fattori secondari (intrapsochici)</p> <ul style="list-style-type: none"> · Globali: autostima (<i>self-esteem</i>), senso di controllo (<i>mastery over life</i>) · Situazionali (Imprigionamento, Competenza, Crescita personale)
FATTORI CONTESTUALI
<ul style="list-style-type: none"> · Variabili sociodemografiche · Rete familiare e sociale
MEDIATORI
<ul style="list-style-type: none"> · <i>Coping</i> (fronteggiamento) definito da Lazarus & Folkman come «processo volto a fronteggiare le richieste esterne e interne percepite come pressanti o eccessive rispetto alle risorse personali. O si cerca di risolvere il problema (<i>problem-focused strategies</i>) oppure di controllare la reazione emotiva (<i>emotion-focused strategies</i>). Esempi di tecniche di <i>coping</i> strumentale focalizzato sul problema comprendono raccolta di informazioni, <i>problem solving</i>, <i>training</i> alle abilità di comunicazione e interpersonali, tecniche di gestione del tempo, mutamenti dello stile di vita, attivazione di varie fonti di aiuto, sforzi per modificare l'ambiente circostante o le situazioni cariche di stress. Tecniche di <i>coping</i> palliativo di regolazione dell'emozione comprendono la ricerca del significato degli stress vissuti, la negazione della realtà o della gravità di un evento, la manifestazione aperta delle proprie emozioni, <i>training</i> di rilassamento, respirazione, meditazione, dialogo interno. · Supporto sociale da parte della famiglia, amici, colleghi, etc.
OUTCOME
<ul style="list-style-type: none"> · Salute mentale e salute fisica · Depressione · Ansia

Gli studi sullo stress del caregiving, come si è accennato in precedenza, sono numerosi e il loro numero sta crescendo di anno in anno. La tabella 2, tratta dal database del Center for Gerontology and Health Care Research, Brown Medical School, ne elenca ben 35. L'elenco ha lo scopo anche di sottolineare la complessità dell'argomento e di suggerire, per così dire, le difficoltà finora incontrate nel definire un quadro interpretativo condiviso delle problematiche associate al processo del caregiving.

La Relative Stress Scale (RSS) (Greene, Smith, Gardiner, Timbury, 1982) è una scala ordinale di 15 item a 5 gradi che definisce la forma e l'intensità dello stress generale che si sviluppa in un parente o in un *caregiver* che assista una persona anziana demente. La scala include tre componenti che misurano il distress personale del caregiver (*personal distress*), lo sconvolgimento nella sua vita (*life upset*) e la presenza di emozioni negative (*negative feelings*). La RSS (Bergmann, Foster, Justice, Matthews, 1978; Bergmann, 1979), che è uno dei primi strumenti di valutazione dello stress associato al caregiving della demenza, è basata sugli studi di Gurel, Linn e Linn (1972) e di Wilkinson e Graham-White (1980). Ancora recentemente è stata impiegata come sistema di confronto per la validazione della Neuropsychiatric Inventory Caregiver Distress Scale (Kaufer, Cummings, Christine *et al.*, 1998).

Tabella 2 – Principali sistemi di valutazione del caregiving e del benessere del caregiver (da Center for Gerontology and Health Care Research, Brown Medical School, www.cher.brown.edu)

<p>AIDS Caregiver Scale (Ferrari <i>et al.</i> 1993). Appraisal of Caregiving Scale, ACS (Oberst <i>et al.</i> 1989; Carey <i>et al.</i> 1991). Bakas Caregiving Outcomes Scale, BCOS (Bakas & Champion 1999). Caregiver Activity Survey, CAS (Davis <i>et al.</i> 1997). Burden Interview, BI (Zarit SH 1980; Zarit JM 1982). Care Work Impact Appraisals (Orbell <i>et al.</i> 1993). Caregiver Burden and Generic Well-Being (Stull <i>et al.</i> 1994). Caregiver Burden Inventory, CBI (Novak & Guest 1989). Caregiver Burden Scale, CB (Elmstahl <i>et al.</i> 1996). Caregiver Load Scale, CLS (Oberst <i>et al.</i> 1989). Caregiver Perceived Burden Scale, CPB (Strawbridge & Wallhagen 1991). Caregiver Reaction Assessment, CRA (Given <i>et al.</i> 1992). Caregiver Self-Efficacy (Zeiss <i>et al.</i> 1999). Caregiver Social Impact Scale (Poulshock & Deimling 1984). Caregiver Strain Index, CSI (Robinson 1983). Caregivers' Stress Scales (Pearlin <i>et al.</i> 1990). Caregiving Appraisal Scale, CAS (Lawton <i>et al.</i> 1989). Caregiving Appraisal Scale, Modified (Hughes and Caliandro 1996).</p>

Care-giving Burden Scale (Gerritsen & van der Ende 1994).
 Caregiving Hassles and Uplifts Scale (Kinney & Stephens 1989).
 Caregiving Indices (Bass & Bowman 1990).
 Caregiving Self-Efficacy Scale, CSS (Denney 1994).
 Cost of Care Index, CCI (Kosberg & Cairl 1986).
 Derogatis Stress Profile, DSP (Derogatis 1987).
 Family Assessment Device, FAD (Epstein, Baldwin & Bishop 1983).
 Family Caregiving Consequences Inventory, FCCI (Shyu et al. 1999).
 Family Caregiving Inventory (Archbold et al. 1992).
 Family Caregiving Scale (Schofield et al. 1997).
 Family Environment Scale, FES (Moos 1990/1974).
 Family Hardiness Index, FHI (McCubbin et al. 1987).
 Family Impact Interview (Covinsky 1994).
 Finding Meaning Through Caregiving, FMTC (Farran et al. 1999).
 Frustration Scale (Motenko 1989).
 Guilt Scale (Wells & Jorm 1987).
 Impact-on-Family Scale (Stein & Riessman 1980).
 Memory and Behavior Problems Checklist, MBPC (Zarit JM 1982, Zarit SH et al. 1985).
 Memory and Behavior Problems Checklist, Revised, RMBPC (Teri et al. 1992).
 Multidimensional Caregiver Strain Index, MCSI (Stull 1996).
 Neuropsychiatric Inventory Caregiver Distress Scale, NPI-D (Kaufert et al. 1998).
 Objective and Subjective Burden Scales (Montgomery et al. 1985).
 Parent Caregiver Strain Questionnaire, PCSQ (England & Roberts 1996).
 Patient-Caregiver Functional Unit Scale, PCFUS (Fredman & Daly).
 Perceived Caregiver Burden Scale, PCB (Stommel et al. 1990).
 Perceived Caregiver Burden Scale, Revised, PCB-13 (Gupta R 1999).
 Perceived Social Support for Caregiving, PSSC, and Social Conflict, SC, scales (Goodman 1991).
 Perceived Stress Scale, PSS (Cohen et al. 1983).
 Picot Caregiver Rewards Scale, PCRS (Picot et al. 1997).
 Relatives Stress Scale, RSS (Greene et al. 1982).
 Screen for Caregiver Burden, SCB (Vitaliano et al. 1991).
 Sense of Competence Questionnaire, SCQ (Vernooij-Dassen et al. 1996).
 Short Sense of Competence Questionnaire, SSCQ (Vernooij-Dassen et al. 1999).
 Subjective Burden Scale, SBS (Matsuda 1999).
 Ways of Coping Checklist (Folkman & Lazarus 1980).

7.2. La RSS come indicatore di caregiver stress in letteratura

Lavori originali: Greene, Smith, Gardiner, Timbury, 1982. John Greene, Department of Psychology Gartnavel Royal Hospital 1055 Great Western Road Glasgow G12 0XH, Great Britain. Lingua: versione originale inglese.

Tabella 3 – Sintesi tecnica dello strumento

TIPO DI STRUMENTO	Scala ordinale di 15 item a 5 gradi.
ORIGINE.	L'autore si è ispirato a lavori esistenti in letteratura.
SCOPO	Lo strumento originale ha diversi scopi sperimentali: valutare il tipo e l'intensità dello stress per un parente che si assuma il carico o la responsabilità di una persona anziana demente; completare la valutazione fatta dalla famiglia sui dementi con una valutazione dei meccanismi psicologici di adattamento della famiglia stessa.
POPOLAZIONE	Familiari di dementi che vivono a domicilio.
SOMMINISTRAZIONE	<i>Esaminatore:</i> La famiglia o chi ha in carico il soggetto. <i>Tempo di somministrazione:</i> Da 10 a 15 minuti. <i>Preparazione:</i> Non necessaria. <i>Valutazione:</i> Scala a 5 gradi da 0 a 4.
COMPONENTI	Tristezza, sconvolgimento, rifiuto.
LAVORI DI VALIDAZIONE	Hanno incluso: Fedeltà test-retest; Analisi fattoriale.
ALTRI LAVORI citati dall'autore	Studio longitudinale di un programma terapeutico realizzato in un «day hospital» su persone anziane dementi.
APPLICAZIONI	Clinica: studi di previsione del comportamento, programma terapeutico, presa in carico, atteggiamenti comportamentali. Ricerca: inchieste psicosociali.

In questo studio la traduzione della scala è stata affidata a un gruppo di tre esperti di madrelingua italiana con specifica esperienza in valutazione psicometrica e funzionale. Non è ancora stata applicata la procedura di contro-traduzione in Inglese.

7.2.1. Alcune premesse

La RSS, nata per applicazioni allo studio di processi assistenziali al paziente demente, si è rivelata molto ben applicabile a misure di «caregiver stress» nelle più varie situazioni inclusa, come in questo studio, l'assistenza geriatrica.

La validazione e, più in generale, la «rifinitura» dei punteggi osservati hanno seguito la tecnica nota come «analisi di Rasch». Nei confronti tra RSS e altre variabili si sono analizzate misure Rasch-trasformate.

Poiché l'analisi di Rasch non è ancora molto nota, il paragrafo che segue ne dà una sintesi generale. Il lettore che già conosca il metodo o che semplicemente decida di fidarsi degli autori può passare direttamente al paragrafo 7.4.

Tabella 4 – Gli item della rss e i livelli di risposta originali

item 1	Pensate mai di non riuscire a fronteggiare la situazione?				
item 2	Pensate mai di avere bisogno di aiuto?				
item 3	Vi capita di sentirvi depressi dalla situazione?				
item 4	La vostra salute ne ha in qualche modo sofferto?				
item 5	Vi preoccupate che succedano incidenti?				
item 6	Pensate mai che il problema non abbia via d'uscita?				
item 7	Avete difficoltà ad andare via per le vacanze?				
item 8	In che misura è stata modificata la vostra vita sociale?				
item 9	In che misura è stato turbato in vostro ménage familiare?				
item 10	Il vostro sonno viene interrotto?				
item 11	Si è abbassato il vostro tenore di vita?				
item 12	Vi sentite mai imbarazzati?				
item 13	Siete impossibilitati a ricevere visite?				
item 14	Vi capita mai di essere contrariati o arrabbiati?				
item 15	Vi capita a volte di sentirvi frustrati?				
codifica	1	2	3	4	5
descrizione	mai, assolutamente	no di rado un po'	a volte	frequentemente molto	sempre moltissimo

7.3. L'analisi di Rasch: principi generali e potenziale impatto sulla ricerca sociale

In ambito sanitario è di primaria attualità il concetto di valutazione dei risultati o *outcome*: outcome viene definito ogni indicatore sintetico che rifletta lo stato di «benessere» globale di ogni utente e che sia generato dalla fornitura di un bene o dall'erogazione di un servizio alla persona; in ambito sanitario l'outcome è interpretato quale costruito multidimensionale di indicatori osservati che sottendono tutti una medesima dimensione inerente qualche aspetto legato alla salute (Gertler, 1998).

In particolare si fa riferimento a concetti come autosufficienza motoria e cognitiva (Tesio, Franchignoni, Perucca, Porta, 1996), qualità della vita (Apolone, Mosconi, Ware, 2000), ritardo mentale (Warr, 1987) ecc.

Le considerazioni principali per la scelta della metodologia da affrontare per la stima degli outcome fanno emergere principalmente due ordini di problemi:

- gli indicatori utili per stimare un outcome sono tipicamente di natura categorial-ordinale (tipo 0/1, presenza/assenza, lieve/moderato/grave). I livelli ordinali sono raccolti in questionari o test;
- si presentano tipicamente situazioni di dati mancanti che generano distorsioni.

7.3.1. *La stima di outcome latenti in letteratura*

In letteratura il problema della stima di outcome latenti è stato affrontato per valutare abilità, percezioni, capacità o più in generale attitudini accomunate dalla definizione di *latent trait*; questi non sono direttamente osservabili, ma desumibili da un insieme di item (prove attitudinali, voci su questionari, prove fisiche o mentali, test psicologici ecc.), che misurano a meno di errori il *latent trait* con la metodologia denominata in generale *item response theory* (Lord, Novick, 1968).

Le metodologie utilizzate per la stima di variabili non osservabili o *latent traits* (LT) hanno attinto dai modelli di analisi fattoriale (FA) (Bartholomew, 1987), suscitando tuttavia una serie di critiche. Tra queste spiccano la rigidità delle ipotesi su cui si basa il modello e soprattutto la difficoltà di stimare i risultati latenti in maniera oggettiva (Vittadini, Lovaglio, 2001; Lovaglio, 2003; Vittadini, 1999).

7.3.2. *Caratteristiche della metodologia per la stima di outcome oggettivi*

Dalle modalità dicotomiche oppure ordinali attraverso cui gli individui rispondono agli item occorre poter ricavare misure continue invariante (cioè stabili in diversi contesti) compatibili con le caratteristiche di seguito descritte.

Unidimensionalità

L'insieme di item deve essere unidimensionale, ovvero deve sottendere un'unica dimensione latente. Qualora invece esso sottenda più di una dimensione latente finisce per fornire risultati contraddittori in quanto non risulta chiaro se e in che misura ogni item e gli item nel loro complesso si riferiscano all'una o all'altra dimensione latente oggetto di studio. Ciò non significa che situazioni empiriche possano e debbano essere necessariamente caratterizzate da un insieme di item perfettamente unidimensionali: si deve ottenere un'omogeneità sufficiente per lo scopo dello studio (Smith, Johnson, 2000).

«Test free» e «sample free» calibration

Perché siano possibili confronti tra individui in tempi e situazioni diverse occorre trasformare le misure dicotomiche e ordinali in misure espresse in unità oggettive affinché le nuove scale siano «invarianti», ovvero indipendenti dal tipo di item considerati (*test free*) e indipendenti dal particolare campione di persone sottoposte alla prova (*sample free*) (Wright, Linacre, 1989). Altrimenti si può ben dire che «solitamente non si riesce a sapere quanto le caratteristiche di un insieme di item forniscano informazioni chiave soltanto sull'abilità di un particolare campione di persone e quanto i risultati ottenuti dipendano soltanto dalla difficoltà delle prove piuttosto che anche dall'abilità delle persone» (Smith, Johnson, 2000).

Misure lineari e quantitative

Le misure devono essere esprimibili su scale che consentano operazioni algebriche che invece i dati categoriali iniziali non consentono (Smith, Johnson, 2000). Esse devono mantenere in qualche modo le corrispondenze di ordine gerarchico e distanze fra i soggetti proporzionali ai punteggi. Ad esempio se i livelli no/lieve/grave/gravissimo sono categorizzati come 0/1/2/3 non vi è garanzia che sia vero che $3 - 2 = 2 - 1$, e con questo ogni operazione diventa criticabile.

Effetto di compressione sulle «code»

I questionari prevedono necessariamente punteggi minimi e massimi, ma il LT esplorato è potenzialmente infinito. Nelle matrici di dati ordinali è dunque tipico il problema dell'effetto compressione sulle code qualora un numero di individui (o di item) avessero conseguito punteggi iniziali vicino al minimo (o al massimo) e quindi tali da confondere gradi di abilità (o difficoltà) nettamente differenti tra loro. Le misure di outcome stimate devono essere «calibrate» lungo l'intero arco dei numeri reali per smascherare differenze di quantità lungo il LT esplorato anche sulle code della distribuzione.

«Missing data»: dati mancanti

Le misure di abilità dei soggetti e di difficoltà degli item non devono essere influenzate dal problema dei dati mancanti in una o più osservazioni. Occorre dunque un valido modello di stima dei punteggi mancanti stessi.

7.3.3. La metodologia di Georg Rasch per la stima di outcome oggettivi

Per poter essere utilizzati ai fini della stima degli outcome gli indicatori categoriali devono essere trasformati in *objective measures* quantitative

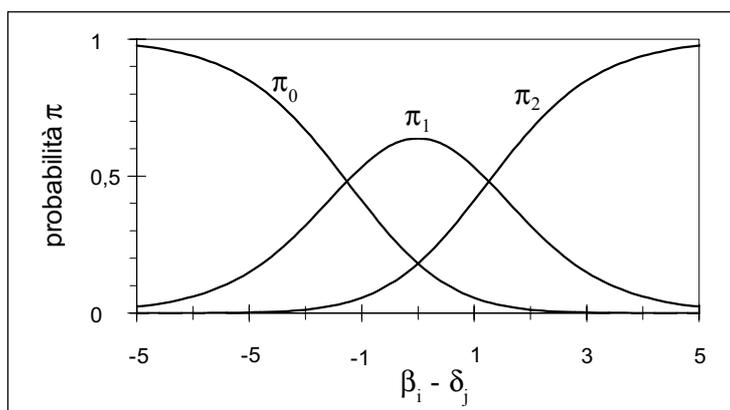
e continue, «calibrate» lungo l'intero arco dei numeri reali. Queste misure non possono che essere stimate secondo modelli statistici ed essendo stime esse forniscono misure «probabilistiche» ottenute con metodi inferenziali che riproducono le loro componenti sistematiche, e stimano al meglio i valori mancanti di uno o più individui su uno o più item, senza introdurre né informazioni, né modelli causali (comunque arbitrari) esterni alla pura logica di valutazione dei risultati (Bergmann, 1979).

La metodologia di Rasch (Rasch, 1960; Andrich, 1978; Wright, Mok, 2000; Wright, Masters, 1982), che si colloca all'interno dei modelli di *item response theory* (Pearlin *et al.*, 1990), perviene a stimare misure di outcome (abilità dei soggetti lungo un LT) soddisfacendo tutte le proprietà ideali di una misura quantitativa e oggettiva, così come viene sintetizzato nella definizione di Wright e Linacre (Bergmann, Foster, Justice, Matthews, 1978).

Secondo questa definizione, attraverso un modello item-response, dato un insieme di risposte di un campione di individui a item che sottendono un unico LT , si costruiscono misure quantitative, continue, probabilistiche, indipendenti dal campione di individui considerato e, per gli individui, dal particolare insieme di item utilizzati.

In particolare il modello Rasch suppone che la probabilità per un soggetto di rispondere con una certa categoria (0/1/2 ecc.) a un determinato item dipende soltanto dalla differenza tra due parametri: l'abilità degli individui (funzione diretta) e le difficoltà di rispondere con modalità via via più difficili agli item (funzione inversa), attraverso un modello non lineare, ma di tipo logistico. Ovvero la misura non è proporzionale alla probabilità di risposta ma a una sua particolare trasformazione, detta LOGIT (Tesio, 2003) (per una descrizione formale del modello si veda anche l'inserito a fine capitolo).

Figura 1 – Esempio di categorie di risposta: 0 = non passa la prova; 1 = passa con difficoltà; 2 = passa la prova



La figura 1 dà una rappresentazione del modello di probabilità di risposta specificato dall'analisi di Rasch: la risposta positiva nell'item i , prevede tre livelli. In ordinata sono riportate le curve che esprimono la probabilità di rispondere 0, 1, 2, e in ascissa la differenza fra l'abilità del soggetto e la difficoltà dell'item. Soggetti via via più abili avranno probabilità sempre più elevate di punteggio superiore, mentre il contrario vale per soggetti meno abili. A soggetti per i quali la differenza fra abilità (β) e difficoltà (δ) sia pari a 0 corrisponde una probabilità di superare l'item del 50% (si dice che l'item è «on target»); le curve mostrano la probabilità per il soggetto di ottenere 0 (π_0), la probabilità di ottenere 1 (π_1) e di ottenere 2 (π_2) nell'item j ; i parametri che descrivono la difficoltà della *soglia* tra risposta 1 e 0: d_{j0} (tra la risposta 2 e 1: d_{j1}) è l'intersezione delle curve π_0 e π_1 (π_1 e π_2).

7.3.4. Analisi dell'adattamento del modello

Una volta costruite le misure di outcome occorre verificare se e in che misura le stime ottenute (sulle unità e sugli item) distorcano le modalità iniziali di risposta, espressione diretta delle risposte dell'individuo di fronte agli item. Occorre quindi verificare l'esistenza di errori sistematici che impediscano una buona approssimazione dei dati di partenza a quanto previsto dal modello. Si possono individuare due tipi di errori:

- 1) *misfitting*, risposte non prevedibili in quanto troppo in contrasto con il grado di abilità dell'individuo e con il grado di difficoltà dell'item, ovvero risposte troppo positive a item troppo difficili o risposte insoddisfacenti a item troppo facili, data l'abilità del soggetto;
- 2) *overfitting*, risposte troppo sistematicamente legate all'abilità dell'individuo e alla difficoltà dell'item senza alcun effetto della componente casuale. In sostanza risposte troppo prevedibili e perciò «sospette» di essere determinate da fattori esterni sconosciuti, dato che ci si aspetta una componente casuale dell'esperimento.

Il modello di Rasch permette quindi di rispondere alle seguenti domande:

- a) Quale abilità o propensione ha ogni individuo rispetto al latent trait considerato?
- b) Qual è la difficoltà di ogni item in generale (*item location*) e di ogni «soglia» che scandisce le differenti categorie degli item?
- c) Con quale errore sono stimati i parametri di difficoltà, di difficoltà dell'item e delle soglie?
- d) Quali soggetti (item) presentano risultati non attesi data la difficoltà dell'item (l'abilità del soggetto)?

Le applicazioni dell'analisi di Rasch in letteratura sono molteplici e riguardano, per esempio, la stima dei seguenti outcome: l'interesse degli studenti verso l'osservazione scientifica (Wright, Masters, 1982); la capacità di lettura da parte di un insieme di scolari delle classi elementari (MacManara, 1996); l'abilità manuale di pazienti con artrite reumatoide postoperati (Penta, Thonnard, Tesio, 1998); la misura della disattenzione e dell'iperattività degli studenti (Smith, Johnson, 2000); la *customer satisfaction* percepita dai pazienti in seguito al ricovero in ospedale (CRISP, 2001); la misura delle prestazioni accademiche nei college americani (Bassiri, Schulz, 2002; 2003).

7.4. Analisi di Rasch sulla RSS: trasformazione lineare dei punteggi grezzi; unidimensionalità della scala, *reliability* degli item, qualità delle rilevazioni

La Relative Stress Scale (RSS) è stata applicata allo studio di monitoraggio del buono socio-sanitario (BSS) di cui si parla estesamente in altre parti di questo volume. Il tema del presente paragrafo consiste nella valutazione delle proprietà metriche e nell'analisi delle informazioni che si possono trarre da questo indicatore. Si rimanda ai capitoli specifici per la descrizione del disegno epidemiologico.

Sono stati analizzati tutti i pazienti beneficiari del BSS sia alla prima sia alla seconda intervista. Da questo campione sono stati eliminati i *missing* totali cioè tutti i casi che hanno un valore mancante per tutti e 15 gli item della RSS; i casi validi risultano 1236 (653 al Tempo 1 e 583 al Tempo 2). La tabella 5 mostra il tipo di caregiver dell'anziano nei due tempi di rilevazione.

Sui 15 item della RSS è stata effettuata un'analisi di Rasch al fine di validare la scala e rendere i punteggi di stress sui caregiver delle misure attendibili, attraverso il software RUMM (Sheridan, 1998).

L'analisi Rasch ha fornito indicazione sul collasso a tre categorie di risposta per ogni item, sulle 5 iniziali (con modalità: «mai», «talvolta», «sempre» 0, 1, 2). Dei 1236 dati iniziali se ne sono analizzati soltanto 1212 casi poiché i dati riferiti a 24 soggetti hanno evidenziato punteggi estremi (tutte le risposte ai 15 item uguali a 0) e quindi, secondo la «filosofia Rasch», non informative: il punteggio 0 rappresenta meno di 1, ma si ignora «quanto» di meno (non si può distinguere il grado di stress fra coloro che hanno una «abilità» che meriterebbe punteggi inferiori al minimo).

Tabella 5 – Caregiver nei due tempi: numerosità e tipologia

	CODICE	TEMPO1	TEMPO2	Totale
Soggetto	0	10	11	21
Coniuge	1	31	43	74
Figlio	2	69	61	130
Figlia	3	278	240	518
Nipote	4	37	36	73
Altri Parenti	5	70	63	133
Altri	6	9	10	19
Operatore Professionale	7	4	5	9
Personale retribuito	8	48	46	94
Personale religioso	9	81	68	149
Dato mancante	-1	16	0	16
totale		653	583	1236

Nella tabella 6 si forniscono le statistiche degli item (ne vengono forniti sia il codice, sia la descrizione). In particolare si notino la difficoltà degli item (Location), l'errore standard (SE), la componente residua (Residual) tra il punteggio atteso dal modello e quello osservato, il numero di casi validi (Data) e l'adattamento di ognuno (Prob), indice di quanto il punteggio atteso e osservato siano significativamente diversi, ottenuti dalla statistica di adattamento chi quadrato (Chi SQ) separatamente per i due tempi di rilevazione.

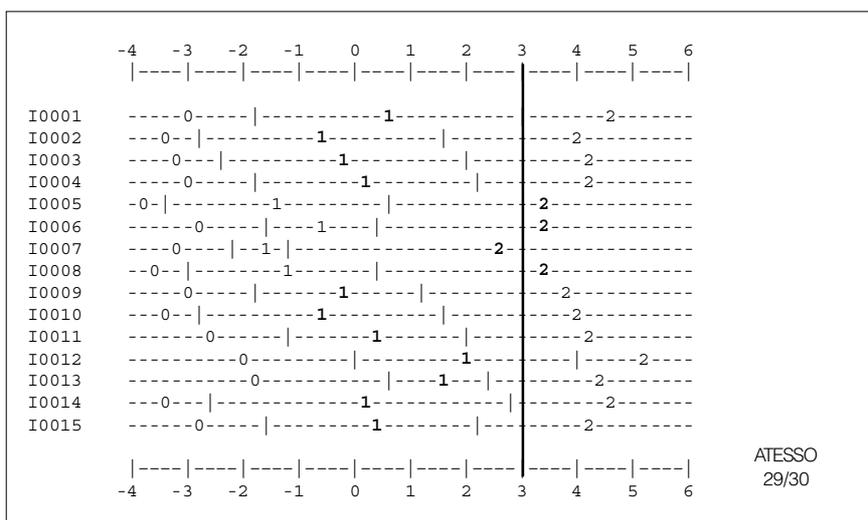
Si noti che le «Location» rappresentano trasformazioni lineari dei punteggi grezzi. La differenza di un'unità (qui definita LOGIT) ha lo stesso significato a qualsiasi livello della misura (mentre lo stesso non vale per i punteggi grezzi: $3 - 2 \neq 2 - 1$); convenzionalmente 0 è la location della difficoltà media degli item e di un soggetto che abbia il 50% di probabilità di superare l'item medio. Valori negativi indicano pertanto non un'assurda «anti-abilità» o «anti-difficoltà», ma soltanto livelli inferiori a quelli medi. Per entrambi i tempi si nota che «il più facile» è l'item 7 (*Problemi per vacanze*) nel senso che è massimamente probabile osservare caregiver in accordo con tale affermazione (stress frequente) anche per livelli di stress complessivo non particolarmente elevati, mentre il più difficile è l'item 12 (*La situazione vi crea imbarazzo*), intendendosi che l'accordo su tale item (stress frequente) è probabile solo con caregiver particolarmente stressati in generale. Ovvero quasi tutti affermano di avere problemi per le vacanze, mentre ben pochi arrivano a manifestare imbarazzo per la situazione. Nelle figure 2 e 3 vengono mostrate le mappe delle *soglie* che descrivono la difficoltà di

Tabella 6 – *Item difficulty* nei due tempi

Tempo 1							
item	Descrizione	Location	SE	Residual	Data	Chi Sq	Prob
1	fronteggiare situazione	0,678	0,089	0,640	629	5,529	0,237
2	bisogno di aiuto	-0,522	0,086	2,310	629	34,684	0,000
3	sentirsi depressi	-0,074	0,087	-3,569	629	18,873	0,001
4	risentire in salute	0,360	0,085	-2,702	629	14,073	0,007
5	preoccupazione infortuni	-1,280	0,082	1,812	629	32,939	0,000
6	senza via di uscita	-0,497	0,071	1,447	629	19,410	0,001
7	problemi per vacanze	-1,581	0,065	0,650	629	1,601	0,809
8	modifiche vita sociale	-1,243	0,076	-3,567	629	25,625	0,000
9	turbamento ménage familiare	-0,256	0,077	-3,770	629	18,416	0,001
10	interruzione sonno	-0,483	0,085	1,128	629	11,606	0,021
11	tenore di vita ridotto	0,466	0,082	-1,739	629	5,062	0,281
12	vi crea imbarazzo	2,093	0,103	1,146	629	1,362	0,851
13	impossibile ricevere visite	1,627	0,105	1,520	629	17,797	0,001
14	arrabbiati contrariati	0,235	0,093	2,304	629	19,649	0,001
15	sentirsi frustrati	0,477	0,084	-4,019	629	23,648	0,000
Tempo 2							
item	Descrizione	Location	SE	Residual	Data	Chi Sq	Prob
1	fronteggiare situazione	0,388	0,098	-1,326	536	3,718	0,445
2	bisogno di aiuto	-0,797	0,097	-1,485	536	0,908	0,923
3	sentirsi depressi	-0,344	0,093	-3,359	536	8,881	0,064
4	risentire in salute	0,375	0,096	-2,870	536	9,900	0,042
5	preoccupazione infortuni	-1,207	0,096	0,464	536	19,645	0,001
6	senza via di uscita	-0,594	0,077	0,398	536	11,673	0,020
7	problemi per vacanze	-1,746	0,074	1,405	536	12,501	0,014
8	modifiche vita sociale	-1,235	0,090	-2,822	536	16,923	0,002
9	turbamento ménage familiare	-0,396	0,087	-3,887	536	13,935	0,008
10	interruzione sonno	-0,526	0,092	2,423	536	31,966	0,000
11	tenore di vita ridotto	0,887	0,095	-0,623	536	4,724	0,317
12	vi crea imbarazzo	2,660	0,113	2,237	536	22,878	0,000
13	impossibile ricevere visite	2,135	0,116	-0,246	536	6,436	0,169
14	arrabbiati contrariati	0,008	0,100	2,192	536	22,547	0,000
15	sentirsi frustrati	0,392	0,092	-2,338	536	8,983	0,062

risposta alle diverse categorie all'interno di ogni item, in un range tra -4 e 6 LOGIT. Si noti che le categorie intermedie («di rado», «a volte», «frequentemente») sono state collassate in un'unica categoria centrale, come da indicazioni emerse dall'analisi Rasch. Da una struttura 1/2/3/4/5 si è dunque passati a una struttura 0/1/2. I tratti verticali tra le categorie indicano il livello di abilità in cui è ugualmente probabile prendere uno dei punteggi adiacenti: al Tempo 1 un soggetto il cui stress si collochi su un livello intermedio, per esempio 3 LOGIT (riga nera verticale in figura 2) ha un punteggio globale atteso di 29/30; nei singoli item il punteggio è pari a 1 sull'item 12 e pari a 2 su tutti gli altri.

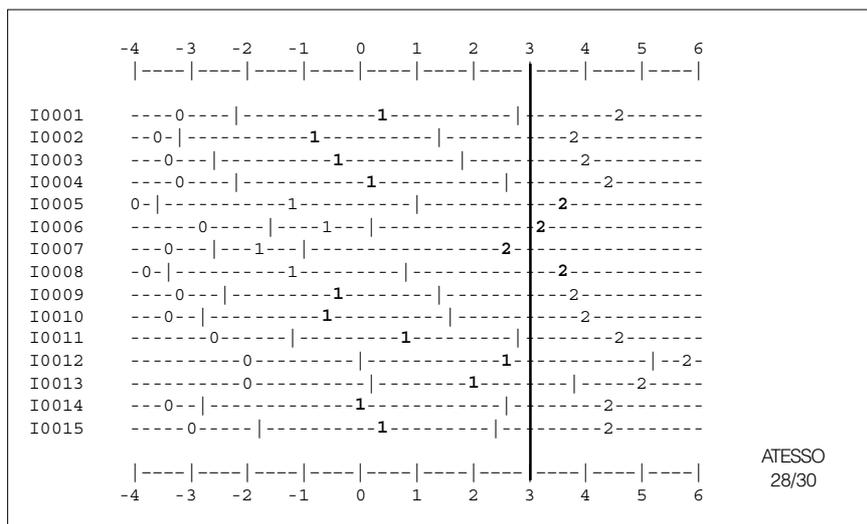
Figura 2 – Mappa delle soglie al Tempo 1



Si noti come la scala RSS sia intrinsecamente stabile: al Tempo 2 (figura 3) alla stessa abilità complessiva (3 LOGIT) corrisponde lo stesso punteggio anche nei singoli item, tranne che nell'item 13.

La difficoltà delle risposte da ogni item è da intendersi come la quantità di stress, in accordo con una certa affermazione su un item; per entrambi i tempi di rilevazione la categoria più difficile (nel senso di più rara) risulta l'ultima («sempre») relativa all'item 12 («La situazione vi crea imbarazzo»), ovvero la risposta coerente con un imbarazzo molto frequente richiede uno stress complessivo maggiore rispetto alla risposta «sempre» sull'item 6: «La situazione non ha via d'uscita», mentre la meno difficile è la prima categoria dell'item 5 («Preoccupazione per possibili infortuni»).

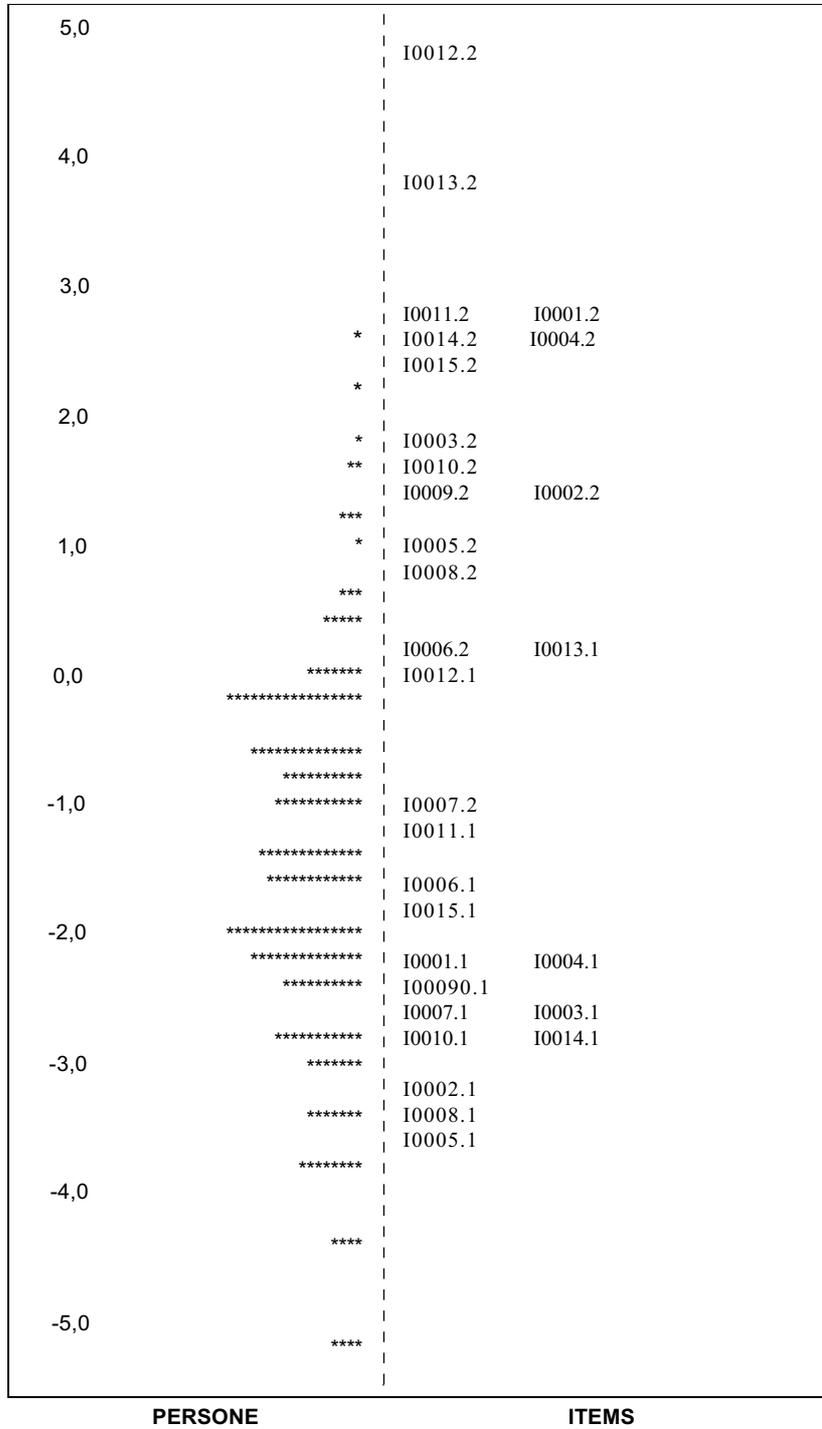
Figura 3 – Mappa delle soglie al Tempo 2



Nelle figure 4 e 5 vengono fornite le mappe (per Tempo 1 e Tempo 2) che congiuntamente mostrano la distribuzione (* = 3 persone) dei punteggi di stress dei caregiver e dell'accordo alle risposte degli item valutata come difficoltà di superare le soglie tra categorie attigue in ogni item (parte sinistra).

Come evidenziato nella figura 4 (ma le stesse conclusioni riguardano anche la figura 5 sulla parte sinistra), dal basso verso l'alto vengono riportati i punteggi di stress dei caregiver via via maggiori, mentre nella parte destra, dal basso verso l'alto sono mostrate le difficoltà delle risposte di ogni item, via via più difficili.

Figura 5 – Mappa abilità e soglie, Tempo 2. Altre definizioni come nella figura 4



7.4.1. Analisi di affidabilità della scala: misure grezze

L'analisi di affidabilità può essere usata per costruire nuove scale di misura, per migliorare le scale esistenti e per valutare l'affidabilità delle scale già in uso.

Questa analisi si rende necessaria poiché una serie di item tipicamente carpisce solo una parte del LT e lascia una componente di errore da valutare. In questo contesto la definizione di *affidabilità* è il rapporto tra la variabilità del concetto latente rispetto alla variabilità complessiva dell'insieme degli item e viene valutata tipicamente con l' α di Cronbach: quanto più α è vicina a 1 tanto più gli item nel complesso sono affidabili. Sui punteggi grezzi cumulati degli item al Tempo 1 si ricava $\alpha = 0,8536$ e al Tempo 2 $\alpha = 0,8821$, valori molto soddisfacenti.

Nella tabella 7 si mostra per ogni item la «correlazione punto biseriale» indicativa della coerenza delle risposte di ciascun soggetto. La «P-Biseriale» è calcolata come coefficiente di correlazione tra le risposte del soggetto a ciascun item e il punteggio totale del soggetto sul complesso degli item. Tutti gli indici mostrano un buon grado di coerenza degli item nel loro complesso.

Tabella 7 – Analisi di affidabilità sui punteggi grezzi

<i>P-Biseriale</i>		
<i>items</i>	Tempo 1	Tempo 2
1	0,49	0,58
2	0,35	0,59
3	0,60	0,65
4	0,58	0,63
5	0,36	0,48
6	0,48	0,58
7	0,55	0,57
8	0,64	0,65
9	0,64	0,67
10	0,38	0,41
11	0,54	0,53
12	0,34	0,34
13	0,38	0,43
14	0,31	0,39
15	0,64	0,61

7.4.2. Analisi di affidabilità della scala: misure Rasch

Una misura analoga all' α di Cronbach sui punteggi Rasch è il person separation index (PSI) (Wright, Masters, 1982) che ha fornito analoga affidabilità per ciascun tempo (PSI = 0,847 al Tempo 1 e PSI = 0,898 al Tempo 2).

7.4.3. Stabilità nel tempo delle misure

Punteggi grezzi degli item tra i due tempi

Sono state valutate le stabilità nel tempo degli item attraverso tre test statistici² basati sul test di Wilcoxon (W), sul Kappa di Cohen (K) e sul coefficiente di correlazione intraclass (ICC) (tabella 8).

Il K misura l'accordo tra due valutatori, o tra due istanti temporali quando entrambi gli strumenti stanno stimando lo stesso oggetto. Esso si basa sulla differenza (Tempo 1 – Tempo 2) tra la proporzione di casi osservata nella quale gli stimatori sono in accordo; vale 1 se vi è perfetto accordo e 0 se l'accordo può essere considerato interamente casuale.

L'ICC, che si basa sulla misurazione degli item in due occasioni diverse sullo stesso campione, rappresenta la porzione di variabilità tra i tempi rispetto alla variabilità totale; se la misura è stabile ci si aspetta che i punteggi degli item nelle due occasioni abbiano un'alta correlazione positiva (max ICC = 1).

Tabella 8 – Test di stabilità intertemporale sui punteggi grezzi

items	Wilcoxon		Kappa		Intraclass	
	valore	p	valore	p	valore	p
1	-2,601	0,009	0,450	<.001	0,673	<.001
2	-0,238	0,812	0,411	<.001	0,670	<.001
3	-1,160	0,246	0,470	<.001	0,694	<.001
4	-1,058	0,290	0,520	<.001	0,725	<.001
5	-2,752	0,006	0,528	<.001	0,716	<.001
6	-0,822	0,411	0,518	<.001	0,759	<.001
7	-0,657	0,511	0,580	<.001	0,818	<.001
8	-0,536	0,592	0,537	<.001	0,753	<.001
9	-1,935	0,053	0,505	<.001	0,741	<.001
10	-0,444	0,657	0,559	<.001	0,762	<.001
11	-1,084	0,278	0,588	<.001	0,733	<.001
12	-0,462	0,644	0,464	<.001	0,644	<.001
13	-0,905	0,366	0,464	<.001	0,678	<.001
14	-0,639	0,523	0,570	<.001	0,731	<.001
15	-0,473	0,636	0,520	<.001	0,719	<.001

² Per tutti i test e gli indici citati in questo paragrafo, si veda Landenna, Marasini, Ferrari, 1998.

Il test di Wilcoxon (w) è una procedura non parametrica per verificare l'ipotesi che le distribuzioni degli item nei due istanti temporali siano le stesse, senza fare assunzioni sulla forma, prendendo in considerazione le differenze dei ranghi tra coppie di osservazioni.

I valori di significatività di K e ICC in entrambi i test (p) mostrano che per ogni item si rifiuta decisamente l'ipotesi di correlazione intraclassa e K nullo, mentre per W in soli due item vi sono valori di significatività che indicano che si possa accettare l'ipotesi di stessa distribuzione dei punteggi tra i due tempi.

Difficoltà degli item nei due tempi

Attraverso test statistici è emersa la stabilità dei punteggi di difficoltà degli item nei due istanti temporali; la correlazione tra i ranghi (Rho di Spearman) è risultata pari a 0,90 ($p < 0,0001$), il test di Wilcoxon ($p = 0,191$) e il test dei segni ($p = 0,118$) sull'uguaglianza della distribuzione sono risultati altamente significativi. Nella tabella 9 si mostrano i ranghi dei 15 item nei due tempi, in cui il valore 1 indica l'item più «facile» e 15 il più «difficile».

Tabella 9 – Ranghi delle difficoltà degli item nei due tempi

items	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Tempo 1	13	4	8	10	2	5	1	3	7	6	11	15	14	9	12
Tempo 2	11	4	8	10	3	5	1	2	7	6	13	15	14	9	12

La correlazione tra i ranghi (Rho di Spearman) ottenuta sul punteggio cumulato dei valori grezzi degli item risulta altamente significativa e pari a 0,71 (non illustrata).

Misure Rasch nei due tempi

L'analisi seguente mira a valutare se le misure lineari Rasch dei livelli di stress nelle due occasioni temporali variano o meno, il che dà un'indicazione della stabilità della misura nel tempo.

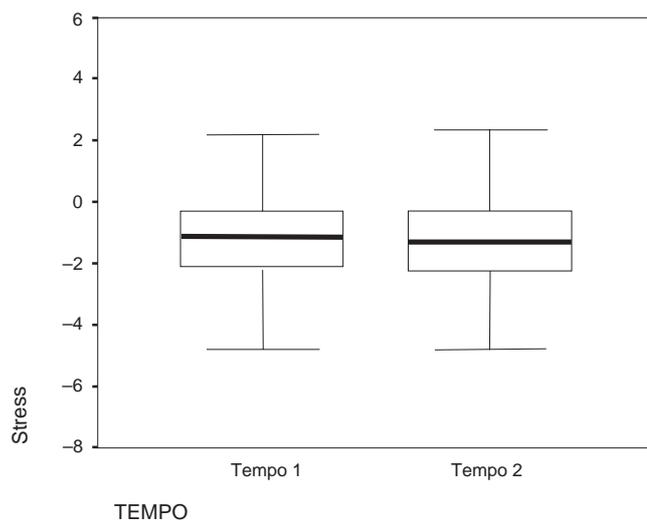
Nella tabella 10 le statistiche relative alle misure di stress ottenute con l'analisi di Rasch nei due tempi fanno emergere che al Tempo 2 si notano livelli di stress minore (media, e mediana inferiori al Tempo 2) e una variabilità maggiore (Deviazione standard, differenza interquartile), rispetto al Tempo 1.

Tabella 10 – Statistiche delle distribuzioni dei punteggi (misura in LOGIT) nei due istanti temporali

	Statistiche	
	Tempo 1	Tempo 2
Media	-1,4261	-1,6050
Mediana	-1,3760	-1,5410
Moda	-1,8960	-1,8180
Deviazione std.	1,4270	1,7481
Varianza	2,0365	3,0560
Range	8,8410	9,7460
Minimo	-6,0470	-6,3130
Massimo	2,7940	3,4330
Percentile 25	-2,1660	-2,4550
Percentile 75	-0,3290	-0,4000
Diff. interquartile	1,8370	2,0550

Rispetto alla stabilità delle misure di stress nei due istanti temporali, nella figura 6 viene fornita la distribuzione delle misure di stress nelle due occasioni.

Figura 6 – Boxplot («whiskers» al 10° e 90° percentile; box 25°-75° percentile)



Successivamente viene fornita una serie di test statistici su campioni appaiati volti alle seguenti valutazioni:

- 1) stabilità della distribuzione dei punteggi (test di Friedman, di Wilcoxon e del segno sull'uguaglianza delle distribuzioni, di Kendall sulla concordanza dei ranghi delle distribuzioni);
- 2) correlazione delle misure nei due istanti (Rho di Pearson);
- 3) differenza tra le medie dei caregiver nei due istanti (ANOVA con campioni appaiati).

Dalla tabella 11 si nota un elevato livello della correlazione ottenuta con le misure LOGIT dello stress (Rho di Pearson), ma gli altri test mostrano che la distribuzione nei due istanti temporali è significativamente differente e che i punteggi medi di stress sono significativamente differenti tra i due tempi (Anova con misure ripetute).

Tabella 11 – Test sulla stabilità intertemporale su dati appaiati

<i>Statistica</i>	<i>Valore</i>	<i>p</i>
Rho di Pearson	0,715	0.000
Test di Friedman	29,610	0.000
Test di Kendall	29,609	0.000
Test del Segno	-5,399	0.000
Test di Wilcoxon	-3,640	0.000
ANOVA con misure ripetute	8,396	0.004

Le differenze tra le medie dei punteggi grezzi sono reperibili nel capitolo 4.

Si possono trarre le seguenti conclusioni:

- a) la scala è metricamente valida, come viene evidenziato dalle statistiche di affidabilità classica, sia su punteggi Rasch sia su quelli grezzi;
- b) la struttura di difficoltà degli item non mostra significative differenze tra i due tempi;
- c) le distribuzioni delle misure Rasch di stress (e dei punteggi medi) evidenziano un'evoluzione temporale, mostrando che i soggetti al Tempo 2 sono in media meno stressati rispetto al Tempo 1, come verrà mostrato nel paragrafo 7.6.2.

7.5. Una serie di ulteriori questioni

Nel presente paragrafo si analizzeranno due ordini di questioni:

- 1) come variano i punteggi di stress al variare del tipo di caregiver;
- 2) il livello di stress dei caregiver che trasferiscono gli anziani in RSA al Tempo 2 è significativamente diverso da quello di chi invece non trasferisce l'anziano?

Rispetto al primo punto si nota che per entrambi i tempi si può rifiutare decisamente ($p < 0,00001$) l'ipotesi di livelli medi di stress uguali per tutti i tipi di caregiver. Nella tabella 12 si riportano i punteggi medi di stress per ogni tipo di caregiver e tempo. Si noti che lo stress aumenta per le categorie «nipote», «altri» e «personale retribuito», mentre diminuisce per le altre categorie.

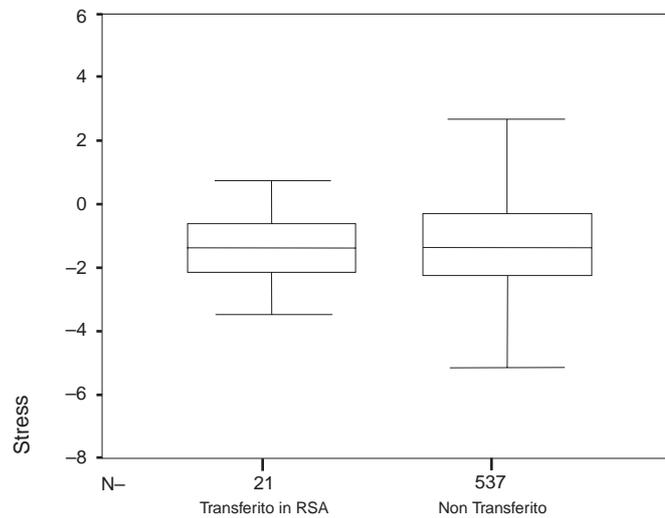
Tabella 12 – Statistiche dei punteggi di stress per *caregiver*

Caregiver	codice	Tempo 1			Tempo 2		
		Media	N	Dev. std.	Media	N	Dev. std.
Soggetto	0	-1,69	10	1,74	-2,25	10	1,55
Coniuge	1	-0,34	30	1,44	-1,22	41	2,11
Figlio	2	-1,16	67	1,39	-1,44	58	1,60
Figlia	3	-1,19	273	1,29	-1,23	229	1,60
Nipote	4	-1,60	37	1,57	-1,29	36	1,53
Altri parenti	5	-1,28	67	1,08	-1,67	61	1,41
Altri	6	-2,80	8	1,21	-1,74	9	1,67
Operatore Professionale	7	-2,80	3	0,60	-3,99	4	1,73
Personale retribuito	8	-2,03	45	1,50	-1,62	45	1,47
Personale religioso	9	-2,46	76	1,24	-3,15	65	1,87

Sembra dunque che una maggiore contiguità affettiva parentale si associ a una riduzione di stress nel periodo di osservazione

Rispetto alla questione se lo stress del caregiver cambi tra coloro che hanno trasferito l'anziano in RSA o meno, emerge che i punteggi medi di stress non variano significativamente tra i due sottogruppi ($F = 0,41$; $p = 0,519$), come mostrano i «boxplot» nella figura 7.

Figura 7 – Variazione di misura di stress tra caregiver che hanno e che non hanno ricoverato l'anziano in RSA



7.6. Legame tra stress e caratteristiche cliniche dell'anziano

La formulazione di un modello statistico di regressione consente di stimare quali possano essere i fattori da cui può dipendere lo stress del caregiver (variabili indipendenti o regressori).

Per «spiegare» la variazione dei punteggi di stress ottenuti dal modello Rasch vengono proposti come regressori i punteggi di autosufficienza degli item FIM³ (FIM1-FIM18), i punteggi di co-morbidità CIRS del paziente (Sheridan, 1998) (CIRS1-CIRS14), l'età e il minutaggio medio settimanale di assistenza (minuti totali), distinguendo poi tra minutaggio solo di supervisione (*minuti supervisione*) e minuti assistenziali «fisici» e quindi non limitati a supervisione (*minuti no supervisione*).

La definizione degli indici FIM e CIRS è riportata rispettivamente nei capitoli 3 e 4.

In via iniziale le correlazioni dei punteggi di stress al Tempo 1 con le variabili di minutaggio e con le componenti FIM (motoria, cognitiva, totale e legata ai trasferimenti) sono significative, ma alquanto basse.

³ Su tali argomenti si vedano i capitoli 3 e 4.

Tabella 13 – Matrice di correlazione: sottolineate le correlazioni non significative a livello 0,01

TEMPO 1	Stress	Min. totali	Min. no superv	Min. Superv	Fim Totale	Fim Motoria	Fim Cognitiva	Fim Trasferim.
Stress	1,000							
Min. totali	0,228	1,000						
Min. no superv	0,225	0,243	1,000					
Min. Superv	0,184	0,975	<u>0,022</u>	1,000				
Fim Totale	-0,189	-0,245	-0,305	-0,183	1,000			
Fim Motoria	-0,174	-0,248	-0,318	-0,183	0,953	1,000		
Fim Cognitiva*	-0,168	-0,175	-0,200	-0,134	0,825	0,616	1,000	
Fim Trasferimenti	-0,134	-0,249	-0,279	-0,193	0,853	0,927	0,491	1,000

* Somma nei tre items di trasferimento wc, vasca/doccia, letto/sedia.

Si noti che, come ci si aspettava, vi è correlazione negativa tra minutaggi FIM e livello di stress: al crescere dell'autosufficienza diminuiscono lo stress e i minuti di assistenza.

7.6.1. Un modello previsionale del livello di stress

Al fine di prevedere i punteggi di stress attraverso componenti legate alla dipendenza motoria e cognitiva, alla presenza di eventuali comorbilità, e a componenti legate ai tempi assistenziali a carico del caregiver (minutaggi assistenziali, suddivisi in minuti di supervisione e minuti di non supervisione) e l'età dell'anziano viene specificato un modello lineare generalizzato in cui tra le variabili indipendenti figurano come variabili «dummy» i contrasti (presenza/assenza) dei punteggi CIRS (1 = co-morbilità lieve, 2 = co-morbilità grave) e come covariate i punteggi nei 18 item della scala FIM, l'età e i minutaggi medi settimanali legati al tempo di supervisione (minuti supervisione) e a quelli di non sola supervisione (minuti no supervisione).

Il modello mostra un basso potere esplicativo ($R^2 = 0,113$), risultando comunque significativo ($p < 0,05$) in alcune delle sue componenti, come emerge nella tabella 14; in essa dal basso verso l'alto appaiono le variabili più significative (Sig) insieme alle stime dei parametri (Beta) e al grado di variabilità delle stime o errore standard (ES). L'analisi si riferisce al Tempo 1.

Tabella 14 – Coefficienti (Beta), significatività (Sig) ed errore standard (ES) dei parametri di regressione

Variabili	Beta	ES	Sig.
Minuti No superv.	0,115	0,028	0,000
FIM18	-0,118	0,037	0,001
Minuti Supervisione	0,020	0,006	0,002
FIM3	-0,152	0,060	0,011
FIM1	0,086	0,036	0,018
<i>Intercetta</i>	<i>-1,871</i>	<i>0,215</i>	<i>0,000</i>

7.6.2. Analisi del cambiamento nel tempo

L'analisi successiva ha lo scopo di indagare quali fattori determinino la riduzione dello stress nei due istanti temporali, come evidenziato nei paragrafi precedenti.

Viene specificato a tal proposito un modello di crescita individuale a effetti random (Rice, Leyland, 1996). I modelli di tale natura, a differenza dei modelli lineari classici, tengono conto della correlazione individuale tra gli istanti temporali (correlazione intersubjects) e suppongono che la relazione tra le variazioni della variabile dipendente e le variazioni delle variabili indipendenti non siano stabili, ma varino tra i soggetti in base a caratteristiche individuali. Nella fattispecie qualora esistesse una differenza tra i punteggi di stress nei due istanti temporali il modello riuscirebbe a prevedere quali fattori individuali determinano tale cambiamento nel tempo.

In un solo modello, dunque, si riescono a valutare sia le covariate che spiegano il valore dello stress al Tempo 0, sia il tasso di cambiamento tra i due tempi.

Tra le variabili individuali che possono spiegare la variazione tra i due tempi vengono specificate come covariate le differenze (Tempo 2 – Tempo 1) dei punteggi della FIM Motoria (Dfimmoto) e della FIM Cognitiva (Dfimcogn) e la FIM legata ai trasferimenti (Dfimtras), il minutaggio medio settimanale di assistenza relativo alla supervisione (Dsuperv.) e alla non supervisione (Dnosuperv.).

L'output, ottenuto con il software statistico SAS, mostra che il tasso di cambiamento medio (in tal caso una diminuzione di 0,15 LOGIT) dello stress nei due periodi è significativo ($p = 0,0039$) e risulta altrettanto significativa la fonte di variabilità da imputarsi ai diversi tassi di cambiamento dello stress tra soggetti diversi ($p < 0,0001$, non illustrata).

Le variabili che influenzano il valore al Tempo 0 sono coerenti con l'analisi di regressione del paragrafo precedente (tabella 14).

Rispetto alle variabili che influenzano il tasso di cambiamento sono stati stimati due modelli: nel primo vengono inserite come covariate le variazioni delle diverse componenti FIM, nel secondo modello le variazioni dei punteggi di tutti gli item FIM (DFIM1-DFIM18) e dei minutaggi assistenziali nei due periodi. La quota di variabilità residua (QS)⁴ dovuta alle covariate è pari al 56%.

Tabella 15 – Variabili che spiegano la diminuzione dello stress nei due istanti: dal basso verso l'alto variabili via via meno significative

<i>Variabili</i>	<i>Estimate</i>	<i>St. Error</i>	<i>t Value</i>	<i>Pr > t </i>
D FIM motoria	-0,03	0,01	-2,02	0,04
D Supervisione	0,01	0,01	0,98	0,33
D FIM cognitiva	-0,02	0,02	-0,97	0,33
D FIM trasferimenti	0,01	0,05	0,31	0,76
D non supervisione	0,00	0,04	0,09	0,93
<i>Intercetta</i>	<i>1,43</i>	<i>0,06</i>	<i>-22,16</i>	<i>0,00</i>

La tabella 15 mostra che l'unica variabile significativa è relativa alle variazioni della FIM motoria: all'aumentare di un punto nell'autosufficienza lo stress diminuisce di 0,03 LOGIT.

Per avere un'idea più concreta di quanto ciò significhi si consideri che un calo di punti FIM si associa a un aumento di stress con il seguente ordine di grandezza: se un soggetto avesse il 50% di probabilità di rispondere «talvolta» sull'item I0003 («Sentirsi depressi», tabella 4) un calo di 5 punti FIM lo porterebbe a una probabilità di circa il 54% di dare la stessa risposta.

Disaggregando i punteggi delle componenti FIM nelle variazioni dei singoli item il modello 2 seleziona l'item cognitivo DFIM17 (soluzione di problemi) e il tempo di non supervisione, come mostrato nella tabella 16.

⁴ Tali modelli, data la loro complessità dovuta alla specificazione di due livelli, il tempo (i) e gli individui (ii), non godono di misure di adattamento tipo R², ma di statistiche che valutano soltanto la quota di varianza residua (in tal caso coincidente con 1-R² del modello presentato nella tabella 12) spiegata dall'inclusione di covariate a livello di individuo, rispetto a un modello senza covariate (QS).

Tabella 16 – Variabili che spiegano la diminuzione di stress nei due istanti: dal basso verso l'alto variabili via via meno significative

<i>Variabili</i>	<i>Estimate</i>	<i>St. Error</i>	<i>t Value</i>	<i>Pr > t </i>
DFIM17	-0,25	0,09	-2,74	0,0064
Dnosuperv.	-0,10	0,05	-2,21	0,0277
DFIM18	-0,16	0,08	-1,91	0,0572
DFIM9	-0,17	0,09	-1,90	0,0585
DFIM3	-0,18	0,11	-1,62	0,1059
DFIM5	-0,21	0,13	-1,55	0,1215
DFIM4	0,15	0,12	1,32	0,1883
DFIM10	0,12	0,10	1,23	0,2205
DFIM12	0,09	0,08	1,14	0,2569
DFIM11	-0,10	0,10	-1,07	0,2871
DFIM1	-0,06	0,06	-1,04	0,2980
DFIM16	0,10	0,10	1,00	0,3183
DFIM2	-0,08	0,10	-0,80	0,4218
Dsuperv.	-0,01	0,01	-0,76	0,4503
DFIM7	-0,06	0,07	-0,75	0,4534
DFIM6	0,06	0,10	0,61	0,5397
DFIM14	0,04	0,08	0,52	0,6008
DFIM15	0,05	0,12	0,43	0,6695
DFIM13	0,04	0,10	0,42	0,6745
DFIM8	-0,01	0,06	-0,16	0,8755

7.7. Conclusioni

1. La scala RSS si rivela uno strumento metricamente valido nel senso che gli item proposti per la stima dello stress del caregiver descrivono omogeneamente un'unica dimensione di disagio/stress del caregiver dell'anziano.
2. L'analisi di Rasch suggerisce di semplificare ulteriormente lo strumento esistente. In particolare si è passati da item su 5 livelli di risposta a 3 livelli, collassando le categorie centrali.
3. La non autosufficienza, misurata su scala FIM e sui minuti assistenziali, è una determinante di rilievo del caregiver stress (circa il 13% di varianza spiegata, significativamente).
4. Anche altre variabili intervengono significativamente: il tema meriterebbe sicuramente un'ulteriore indagine specifica.

5. Il maggiore determinante dello stress per entrambi i tempi di rilevazione sembra essere il tempo di assistenza fisica alle attività della vita quotidiana e l'autosufficienza sull'item «Soluzione di problemi».
6. Il passare del tempo si associa a una diminuzione dei punteggi medi di stress, nonostante la perdita di autosufficienza.
7. Fra i diversi tipi di caregiver si notano differenze significative. In particolare lo stress aumenta nei caregiver professionali e diminuisce nei familiari più stretti.
8. Sia nell'analisi del caregiver stress, sia nella determinazione delle componenti che spiegano la diminuzione dello stress nei due periodi, gli indicatori di co-morbidity non sembrano giocare un ruolo rilevante.
9. La scala FIM e la scala RSS sembrano due strumenti complementari, da utilizzare insieme in studi di questo tipo. Infatti essi sono significativamente correlati, ma non in misura tale da renderli reciprocamente ridondanti. L'analisi delle loro co-variazioni fra categorie di caregiver e nel tempo consente utili valutazioni di qualità delle rilevazioni e consente riflessioni sul diverso atteggiamento di diverse figure assistenziali.

Inserito. Il modello di Georg Rasch

Si definisce «abilità» verso il *latent trait* (LT) ogni «propensione» β_i di un individuo i rispetto a un LT. Si definisce «difficoltà» δ_j la difficoltà (o l'accordo dei soggetti) su un particolare item j ($j = 1, \dots, m$). Nel caso di item con più di due categorie la risposta x su un determinato item j ($x = 0, 1, \dots, k_j$) sottintende implicitamente risposta positiva a ogni modalità precedente $0, 1, 2, \dots, x-1$, la difficoltà di rispondere alla categoria x nell'item j (δ_{jx}) viene considerata come il superamento delle difficoltà dei passi precedenti ($\delta_{jw} = 0, \dots, x$); a sua volta il parametro che esprime la difficoltà del passo w dell'item j (δ_{jw}) si può esprimere come somma di una componente definita «location» dell'item (δ_j) più una componente che è l'effetto differenziale del passo w rispetto alla location (τ_w): $\delta_{jw} = \delta_j + \tau_w$; il classico Dichotomous Rasch Model (Rasch, 1960) viene esteso a item con k_j possibili risposte nella versione Rating Scale (Andrich, 1978) in cui l'effetto differenziale è identico per ogni item ($\delta_{jw} = \delta_j + \tau_w$).

Nella versione Partial Credit le difficoltà di ogni passo sono differenti tra gli item ($\delta_{jw} = \delta_j + \tau_{jw}$) e la probabilità per un soggetto i di rispondere con modalità x nell'item j dipende solo dalla differenza tra i due parametri β_i e δ_j :

$$\pi_{ijx} = \frac{\exp \sum_{w=0}^x (\beta_i - \delta_{jw})}{\sum_{q=0}^{k_j} \exp \sum_{w=0}^q (\beta_i - \delta_{jw})} \quad x = 0, \dots, k_j \quad [1]$$

$$\text{con } \sum_{x=0}^{k_j} \pi_{ijx} = 1 \text{ e } \delta_{j0} = 0$$

Dalla [1] si può mostrare che la probabilità che l'*i*-esimo soggetto risponda *x* piuttosto che *x-1* sull'item *j* (odds) è funzione del parametro di abilità della persona β_i e del parametro relativo alla difficoltà secondo una funzione di tipo logistico:

$$\frac{\pi_{ijx}}{\pi_{ij(x-1)} + \pi_{ijx}} = \frac{\exp(\beta_i - \delta_{jx})}{1 + \exp(\beta_i - \delta_{jx})} \quad x = 1, \dots, k_j \quad [2]$$

Dalla [2] si deduce la relazione tra i punteggi LOGIT e i parametri di abilità del soggetto e difficoltà dell'item in termini lineari

$$\ln \left(\frac{\pi_{ijx}}{\pi_{ij(x-1)}} \right) = \beta_i - \delta_{jx} \quad [3]$$

La [3] mostra che l'abilità e la difficoltà degli item vengono stimate nella stessa unità di misura (LOGIT) e che il soggetto *i* ha una maggiore probabilità di rispondere *x* piuttosto che *x-1* sull'item *j* soltanto se la sua abilità è superiore alla difficoltà della categoria *x* per l'item *j*: se per tale soggetto questa differenza fosse nulla, esso avrebbe la stessa probabilità di rispondere *x* e *x-1* sull'item *j*.

La stima dei parametri del modello Rasch nella versione Partial Credit è ottenuta attraverso la massimizzazione della funzione di verosimiglianza basata sul vettore di risposte degli individui su ogni item, condizionatamente al punteggio complessivo ottenuto sulla totalità degli *m* item (Wright, Masters, 1982).

Bibliografia

- Andrich, D. (1978), «A rating formulation for ordered response categories», *Psychometrika*, n. 43, pp. 581-594.
- Aneshensel, C.S., Pearlin, L.I., Mullan, J.T. *et al.* (1995), *Profiles in caregiving: the unexpected career*, Academic Press, San Diego.
- Apolone, G., Mosconi, P., Ware, J. Jr (2000), *Questionario sullo stato di salute SF-36. Manuale d'uso e guida all'interpretazione dei risultati*, II ed., Guerini e Associati, Milano.
- Bartholomew, D.J. (1987), *Latent variable models and factor analysis*, Oxford University Press, New York.
- Bassiri, D., Schulz, E.M. (2002), *Using Adjusted GPA and Adjusted Course Difficulty Measures to Evacuate Differential Grading Practices in College*. Paper presented at the Eleventh International Objective Measurement.
- Bassiri, D., Schulz, E.M. (2003), *Alternative to grade point averages as measures of academic achievement in college*, ACT Research Report 2001/4, Iowa City (IA), ACT, Inc. Conditionally accepted for publication in *Journal of Applied Measurement*.
- Bergmann, K.B. (1979), «How to keep the family supportive», *Geriat Med*, n. 23, pp. 53-57.
- Bergmann, K.B., Foster, E.M., Justice, A.W., Matthews, V. (1978), «Management of the elderly demented patient in the community», *Brit J Psychiat*, n. 132, pp. 441-449.
- CRISP (2001), *Osservatorio regionale sulla qualità dei servizi sanitari*, report intermedio.
- Gertler, P.J. (1998), «A latent variable model of quality determination», *Journal of business & Economics Statistics*, n. 6, pp. 96-104.
- Greene, J.G., Smith, R., Gardiner, M. (1982), «Evaluating reality orientation with psychogeriatric day hospital patients», in R. Taylor, A. Gilmore (eds), *Recent trends in social gerontology*, Gower Publishing Co.
- Greene, J.G., Smith, R., Gardiner, M., Timbury, G.C. (1982), «Measuring behavioural disturbance of elderly demented patients in the community and its effects on relatives: a factor analytic study», *Age and Ageing*, n. 11, pp. 121-126.
- Greene, J.G., Timbury, G.C. (1979), «A geriatric psychiatry day hospital service: a five-year review», *Age and Ageing*, n. 8, pp. 49-53.
- Gurel, L., Linn, M.W., Linn, B.S. (1972), «Physical and mental impairment of function evaluation in the aged: The PAMIE scale», *J Geront*, n. 1, pp. 83-90.
- Kaufers, D.I., Cummings, J.L., Christine, D. *et al.* (1998), «Assessing the impact of neuropsychiatric symptoms in Alzheimer's disease:

- Neuropsychiatric Inventory Caregiver Distress Scale», *J Am Geriatr Soc*, n. 46, pp. 210-215.
- Landenna, G., Marasini, D., Ferrari, P. (1998), *La verifica di Ipotesi Statistiche*, il Mulino, Bologna.
- Lawton, M.P., Brody, E.M. (1969), «Assessment of older people: self-maintaining and instrumental activities of daily living», *Gerontologist*, n. 9, 3, pp. 179-186.
- Lazarus, R.S., Folkman, S. (1984), *Stress, appraisal, and coping*, Springer, New York.
- Lord, F.M., Novick, M.R. (1968), *Statistical theories of mental test scores*, Addison-Wesley, Reading (MA-USA).
- Lovaglio, P.G. (2003), «The estimate of customer satisfaction in a reduced rank regression framework», *Total Quality Management* (in press).
- MacManara, T. (1996), *Measuring second language performance. Applied linguistics and language study*, Longman Publications, London.
- Meer, B., Baker, J. (1966), «The Stockton geriatric rating scale», *J Gerontol*, n. 21, pp. 372-403.
- Pattie, A.H., Gillear, C. (1979), *Manual of the Clifton assessment procedures for the elderly*, Hodder and Stoughton, Sevenoaks (Kent-UK).
- Pearlin, L.I. *et al.* (1990), «Caregiving and the stress process: an overview of concepts and their measures», *Gerontologist*, n. 30, pp. 583-594.
- Penta, M., Thonnard, J.L., Tesio, L. (1998), «Abilhand: a Rasch-built measure of manual ability», *Arch Phys Med Rehabil*, n. 79, pp. 1038-1042.
- Rasch, G. (1960), *Probabilistic models for some intelligence and attainment tests*, University of Chicago Press, Chicago (www.rasch.org)
- Rice, N., Leyland, A. (1996), «Multilevel models: applications to health data», *Journal of health services research and policy*, n. 1, pp. 154-164.
- Sabbadini, L.L. (2001), *Sostenere chi cura*, Roma 25-26 ottobre (http://www.terzaeta.com/news/ottobre2001/15_10_2001.html).
- Sheridan, B. (1998), *RUMM item Analysis Package: Rasch Unidimensional Measurement Model*, *Rasch Measurement Transactions*, n. 11:4, p. 588 (<http://www.rummlab.com.au>).
- Smith, E.V. Jr, Johnson, B.D. (2000), «Attention deficit hyperactivity disorder: scaling and standard setting using Rasch measurement», *Journal of applied measurement*, n. 1, pp. 3-24.
- Tesio, L. (2003), «Measuring person's behaviours and perceptions: rasch analysis as a tool for rehabilitation research», *J Rehabil Med* (n. 35, 3, pp. 105-115).

- Tesio, L., Franchignoni, F.P., Perucca, L., Porta, G.L. (1996), «The influence of age on length of stay, functional independence and discharge destination of rehabilitation inpatients in Italy», *Disability Rehabilitation*, n. 18, pp. 502-508.
- Vittadini, G. (1999), «Analysis of qualitative variables in Structural Models with unique solutions», in M. Vichi, O. Opitz (eds), *Classifications and data analysis-Theory and Applications*, Springer and Verlag, New York, pp. 203-210.
- Vittadini, G., Lovaglio, P.G. (2001), *The estimate of latent variables in a structural model an alternative approach to PLS*, Second International Symposium on PLS and related methods, Cisia-Ceresta, Montreuil, France, pp. 423-434.
- Warr, P. (1987), *Work, unemployment and mental health*, Oxford University Press, Oxford.
- Wilkinson, I.A., Graham-White, J. (1980), «Psychogeriatric dependency rating scales (PGDRS). A method of assessment for use by nurses», *Brit J Psychiat*, n. 137, pp. 558-565.
- Wright, B.D., Linacre, J.M. (1989), «Observations are always ordinal; measurements, however, must be interval», *Arch Phys Med Rehabil*, n. 70, pp. 857-860.
- Wright, B.D., Masters, G.N. (1982), *Rating Scale Analysis*, Rasch Measurement, MESA, Chicago.
- Wright, B.D., Mok, M. (2000), «Rasch models overview», *Journal of applied measurement*, n. 1, pp. 83-106.