

**UNIVERSITA' DEGLI STUDI DI ROMA
"LA SAPIENZA"
II° FACOLTA' DI MEDICINA E CHIRURGIA
CORSO DI LAUREA IN TECNICO
DELLA RIABILITAZIONE PSICHIATRICA
SEDE DI CIVITAVECCHIA**

Disciplina di Psicometria

Corso Integrato di Psichiatria, Psicologia clinica, Psicometria
II Anno - II Semestre

Dispense di Psicometria

Docente: Dott. Vincenzo Alfano

Anno Accademico 2002-2003

Indice

1. PRINCIPI ELEMENTARI DI MISURAZIONE

1. LA MISURAZIONE
2. CARATTERISTICHE GENERALI DEGLI STRUMENTI DI VALUTAZIONE
3. PRINCIPI ELEMENTARI DI MISURAZIONE
 1. LE MISURE
 1. *Livello Nominale*
 2. *Livello Ordinale*
 3. *Livello Intervallare*
 4. *Livello Razionale*
4. ELEMENTI DI INTERPRETAZIONE DELLE MISURAZIONI
5. BIBLIOGRAFIA

2. LA VALUTAZIONE STANDARDIZZATA

1. GENERALITA'
2. I PRINCIPI DELLA VALUTAZIONE
3. CENNI STORICI
4. LE BASI TEORICHE DELLA VALUTAZIONE
5. BIBLIOGRAFIA

3. FUNZIONI E PROPRIETA' DELLE RS

1. GENERALITA'
2. PROPRIETA' DELLE RS
 1. Validità
 2. Validità per contenuto
 3. Validità per criterio (o empirica)
 4. Validità concorrente
 5. Validità predittiva
 6. Validità per costrutto
 7. Validità fattoriale
 8. Validità nominale
3. SENSIBILITA'
4. AFFIDABILITA'
 1. Consistenza interna
 2. Test-retest reliability
 3. Inter-rater reliability
 4. Intra-rater reliability
 5. Parallel forms reliability
 6. Errore standard
5. BIBLIOGRAFIA

4. CARATTERISTICHE DELLE RS

1. AREA ESPLORATA
2. NUMERO DEGLI ITEM E MODALITA' DI ATTRIBUZIONE DEL PUNTEGGIO
3. TEMPO DELLA VALUTAZIONE
4. CHI EFFETTUA LA VALUTAZIONE
5. COMPLESSITA' DELLA VALUTAZIONE
6. BIBLIOGRAFIA

5. PRINCIPI DI VALUTAZIONE

1. L'INTERVISTA
2. IL PERIODO DI RIFERIMENTO
3. NORMA INDIVIDUALE E INTERINDIVIDUALE
4. IL PROCEDIMENTO DECISIONALE
5. LA SCELTA DEGLI STRUMENTI DI VALUTAZIONE - Principi Generali
6. LA SCELTA DEGLI STRUMENTI DI VALUTAZIONE - Elementi Pratici
7. BIBLIOGRAFIA

6. ALTRE MODALITA' DI MISURAZIONE

1. LA VALUTAZIONE COMPORTAMENTALE
2. SCALE DI VALUTAZIONE PERSONALIZZATE
3. DIARIO
4. MISURE NON INTRUSIVE
5. MISURAZIONI ELETTROFISIOLOGICHE

7. VALUTAZIONE DELLA SCHIZOFRENIA E DEGLI ALTRI DISTURBI PSICOTICI

1. GENERALITA'
2. LE SCALE PER LA SCHIZOFRENIA
3. LA VALUTAZIONE DI ALTRI ASPETTI DELLA SCHIZOFRENIA
4. BIBLIOGRAFIA

PRINCIPI ELEMENTARI DI MISURAZIONE

LA MISURAZIONE

Molti di noi hanno difficoltà ad accostarsi alle tecniche di valutazione standardizzata dei problemi psichiatrici poiché i concetti ad esse connessi risultano di difficile comprensione. Infatti, come in qualsiasi campo, i "tecnici" della misurazione creano un loro linguaggio, un gergo, che finisce per disorientare e mettere in difficoltà chi tecnico non è. Così, mentre noi parliamo semplicemente di numeri, i tecnici parlano di "interi" e definiscono "variabili" quello che noi chiamiamo sintomi, comportamenti, eccetera.

"Misurare" significa quantificare, assegnare cioè un valore numerico alle "variabili". Viene definito come "variabile" tutto ciò che è osservabile direttamente o indirettamente e che può assumere valori diversi; per noi psichiatri le variabili sono il comportamento, i pensieri, i sentimenti, le emozioni, l'autostima, eccetera. Il numero che assegniamo a queste variabili quando le misuriamo è un'attribuzione quantificativa che traduce in numeri ciò che l'uomo ha fatto da sempre, misurare cioè la "qualità" della realtà che lo circonda utilizzando termini che contengono in sé una stima quantitativa.

Quando diciamo, infatti, di aver provato un dolore "grandissimo", "grande", "moderato", "lieve" o "nessun dolore", oppure che un pensiero è presente nella nostra mente "sempre", "molto spesso", "spesso", "qualche volta", "raramente" o "mai", noi facciamo implicitamente una misurazione, collocando la nostra esperienza in qualche punto lungo il *continuum* delimitato dai due gradi estremi di giudizio (grandissimo/assente, sempre/mai). Un'operazione analoga la facciamo quando confrontiamo la nostra variabile ad uno standard di riferimento, quando diciamo, per esempio, che l'intelligenza di un individuo è "di poco", "molto", "di gran lunga" maggiore o minore rispetto alla norma. Gli aggettivi e gli avverbi che noi usiamo hanno implicita in sé questa funzione di misurazione (assoluta o comparativa). Il problema è che non possiamo sottrarre "molto" da "moltissimo", "spesso" da "quasi sempre", "moderato" da "grave" e così via, ed utilizzare quindi le operazioni matematiche come potremmo fare se avessimo stabilito che, per esempio, l'assenza di un sintomo corrisponde a zero, il grado "lievissimo" ad 1 e che ogni grado successivo (lieve, moderato, grave, molto grave, gravissimo) aumenta di un punto (2, 3, 4, 5, 6); in tal caso, il paziente che fosse passato da "gravissimo" a "grave", sarebbe migliorato di 2 punti. Questo è, in effetti, ciò che fanno gli strumenti di valutazione, riproporre cioè i giudizi in termini numerici avendo, gli Autori che li hanno creati, associato un valore numerico a quegli aggettivi o a quegli avverbi.

Generalmente la scelta della scala numerica è arbitraria e risponde a ciò che l'Autore della RS ritiene adeguato al proprio scopo. In realtà esistono delle regole alle quali sarebbe opportuno attenersi; ad esempio l'ampiezza della scala non dovrebbe essere né troppo ridotta, per non perdere dei livelli significativi di definizione, né troppo estesa per non accrescere eccessivamente la variabilità dell'assegnazione dei punteggi da parte dei valutatori. Le regole matematiche vorrebbero anche (e soprattutto) che gli intervalli tra i punteggi fossero uguali fra loro, cioè che la differenza tra "lieve" e "moderato", tra "moderato" e "grave" e tra "grave" e "gravissimo" fosse uguale così come uguale è la differenza fra 10 e 20 centimetri e quella fra 20 e 30 e fra 30 e 40. In ambito psichiatrico (e psicologico in generale) questo è praticamente impossibile ed è perciò necessario far ricorso a convenzioni e adattamenti che impongono una lettura dei risultati ottenuti con questi strumenti che lasci ampio spazio al buon senso e che tenga conto della realtà clinica. Con questa premessa, tenendo conto che la misurazione numerica è principalmente finalizzata all'elaborazione statistica dei dati e che le analisi statistiche comunemente impiegate sono abbastanza "robuste", le RS sono da considerarsi ampiamente accettabili.

CARATTERISTICHE GENERALI DEGLI STRUMENTI DI VALUTAZIONE

Le RS esplorano, come vedremo più avanti, i diversi aspetti che caratterizzano un'entità psicopatologica e ne danno una valutazione quantitativa (in termini di gravità o di frequenza). Gli strumenti che richiedono una risposta dicotoma (SI/NO, presente/assente) sono strumenti di rilevazione più che di misurazione ed è improprio perciò chiamarli RS. Le RS, che esplorano prima la presenza/assenza delle variabili e poi la gravità di quelle presenti, sono tanto strumenti di rilevazione che di misura. Come strumenti di rilevazione, esse rappresentano un punto di riferimento sia personale, perché evitano che alcune variabili non vengano indagate dall'esaminatore per disattenzione o dimenticanza, sia interpersonale, perché tutti coloro che le usano devono valutare tutte le variabili che l'Autore ha ritenuto necessarie per esplorare compiutamente quella specifica entità.

Ogni variabile (o definizione, o voce, o "item") esplora un singolo comportamento. Di solito le RS sono composte da un certo numero di item; solo alcune, come, ad esempio, la *Global Assessment of Functioning scale* - GAF (Endicott et al., 1976) che rappresenta l'asse V della classificazione diagnostica multiassiale del DSM-IV (1994), o la *Visual Analogue Scale* - VAS di Aitken (1969), sono composte da un solo item.

Spesso gli item vengono associati tra di loro a formare dei cluster¹ o dei fattori; si parla di *cluster* quando l'Autore indica a priori quali sono gli item che, a suo giudizio, concorrono alla definizione di una certa condizione psicopatologica più complessa, di *fattori*, quando è l'analisi fattoriale, condotta su di una casistica rappresentativa, che indica quali item sono più strettamente correlati tra loro nella definizione della condizione psicopatologica.

PRINCIPI ELEMENTARI DI MISURAZIONE

LE MISURE

TAB. 1.I - TIPI DI MISURE	
Livello	Proprietà
Nominale	Esclusività
Ordinale	Esclusività + Ordine di grandezza
Intervallare	Esclusività + Ordine di grandezza + Equivalenza
Razionale	Esclusività + Ordine di grandezza + Equivalenza + Zero assoluto

Esistono 4 livelli diversi di misura: nominale, ordinale, intervallare e razionale (Tab. 1.I).

Livello Nominale

Quello nominale (o categorico) è un livello puramente qualitativo, è una modalità di misurazione che si caratterizza per l'esclusività, nel senso che il numero assegnato ad un attributo della variabile è distinto da tutti gli altri e rappresenta sempre e soltanto quell'attributo. Variabili nominali sono, ad esempio, il sesso, lo stato civile, l'occupazione, eccetera. Queste variabili non possono essere impiegate per operazioni matematiche, mentre possono essere usate come criterio di raggruppamento dei soggetti di un campione.

Livello Ordinale

Il livello ordinale possiede, oltre all'esclusività, la proprietà di esprimere i valori in ordine di grandezza (ranking) e di confrontare le posizioni relative all'interno dei numeri assegnati ad una variabile. Così è, ad esempio, il livello di scolarità, all'interno del quale possiamo stabilire una gerarchia che, partendo dal valore più basso (1 = Analfabeta), raggiunga il valore massimo (8 = Specializzazione post-laurea o più di una laurea) attraverso gradi crescenti di scolarizzazione (p. es., 2 = Alfabetata, 3 = Terza elementare, 4 = Licenza elementare, 5 = Licenza media inferiore, 6 = Maturità, 6 = Diploma universitario, 7 = Laurea). È chiaro che il punteggio "3" indica un livello di scolarizzazione inferiore rispetto al punteggio "6" e questo ci consente di confrontare livelli diversi di una stessa variabile, ma non di stabilire l'entità della differenza tra i due livelli mediante un'operazione matematica: non si può, in altri termini, sottrarre la "Terza elementare" dalla "Maturità".

Livello Intervallare

L'uso delle quattro operazioni matematiche è invece appropriato con il livello intervallare poiché questo tipo di misura possiede, oltre alle proprietà dell'esclusività e dell'ordine, quella dell'equivalenza, per cui le distanze tra gli attributi di una variabile sono uguali (p. es., in una scala a 5 punti, la distanza - o la differenza - fra 1 e 3 è uguale a quella fra 2 e 4). Le scale intervallari, tuttavia, non hanno uno zero assoluto in quanto, o non si verifica mai la totale assenza della variabile, o lo strumento non è in grado di valutarla. Questo comporterebbe l'impossibilità di usare le operazioni matematiche con questo tipo di misure poiché, in queste condizioni, non possiamo affermare, ad esempio, che una temperatura di 30° alla scala Celsius (nella quale gli zero gradi non corrispondono allo zero assoluto) è calda il doppio rispetto ad una di 15° o che, se misuriamo la tolleranza di un farmaco su una scala da uno a quattro (1 = cattiva, 2 = discreta, 3 = buona, 4 = ottima tolleranza), un farmaco con una tolleranza pari a "4" sia tollerato esattamente il doppio rispetto ad uno con tolleranza pari a "2". L'introduzione di un convenzionale punto "zero", che indica arbitrariamente l'assenza della variabile, consente l'applicazione delle quattro operazioni matematiche alle misure intervallari. Con questo artificio le misure intervallari possono essere applicate alle RS ed agli strumenti di valutazione che esplorano l'intensità di segni e sintomi e possono essere elaborate mediante le analisi statistiche di tipo parametrico (quali l'analisi di varianza o il t-test di Student) le quali prevedono che i dati provengano da campioni ricavati da "universi" distribuiti secondo la distribuzione normale (vedi oltre).

Livello Razionale

Quello razionale è il livello più elevato e possiede, oltre a quelle dei livelli precedenti, la caratteristica di avere lo zero assoluto. In altri termini, le variabili misurabili con questo tipo di misura possono essere del tutto assenti e gli strumenti di cui disponiamo sono capaci di rilevarne l'assenza. Sono misure razionali quelle di variabili quali l'*età*, l'*altezza*, il *tempo trascorso dall'inizio dei disturbi*, eccetera; eccezionali sono le variabili razionali in psichiatria e nelle scienze umane in genere.

La quantità di informazione fornita dagli strumenti di valutazione cresce passando dalle scale nominali, di più basso livello, a quelle razionali, di livello più elevato. Variabili esprimibili con scale intervallari o razionali (scale, cioè, quantitative) possono essere espresse anche con scale inferiori (cioè qualitative), mentre il contrario non è possibile.

ELEMENTI DI INTERPRETAZIONE DELLE MISURAZIONI

Usare strumenti di valutazione per misurare i problemi dei nostri pazienti ha senso solo se siamo capaci di dare un significato ai numeri assegnati alle variabili che vogliamo misurare ed ai punteggi che otteniamo con quelle misurazioni. Senza avere la pretesa di entrare nel merito della complessa problematica statistica, vogliamo ricordare brevemente, in questo capitolo, soltanto alcune nozioni relative ai punteggi ed ai confronti fra i punteggi, in modo da poter interpretare il risultato del nostro lavoro.

Nel singolo soggetto, una *variabile* (generalmente indicata con la lettera x) (Tab.1.II) assume, in un determinato momento, un valore specifico (x_i) fra tutti quelli ammissibili (x_1, x_2, \dots, x_n); misurando quindi la stessa variabile in tempi diversi si potranno ottenere anche valori diversi.

Le variabili che vengono impiegate più comunemente sono di due tipi, *discrete* e *continue*:

* - le *variabili continue* possono teoricamente assumere infiniti valori nell'intervallo fra due valori dati per quanto vicini essi siano. Un esempio di variabile continua è il peso perché se un individuo pesa 72 kg ed un altro 73, si possono inserire tra loro gli individui il cui peso è di kg 72.1, 72.2, ...72.9 ed ancora, fra gli individui il cui peso è rispettivamente 72.1 kg e 72.2 kg si possono inserire quelli di kg 72.11, 72.12, ...72.19, in tal modo, aumentando il numero di cifre decimali, si può aumentare indefinitamente il numero di valori inseribili fra 72 e 73 kg; Le *variabili discrete* possono assumere solo un numero finito di valori fra i due valori dati per quanto lontani essi siano. La variabile "numero dei figli" è una variabile discreta, perché, ad esempio, tra chi ha 1 figlio e chi ne ha 4 non si possono inserire infiniti valori della variabile, ma solo due: 2 e 3.

DEFINIZIONE DELLE MISURAZIONI		
Variabile	• variabile continua	• può assumere infiniti valori fra due valori dati
	• variabile discreta	• può assumere solo un numero finito di valori fra due valori dati
Campione		Insieme di valutazioni di una o più variabili in un soggetto in tempi diversi e/o in un gruppo di soggetti
Distribuzione		Modalità di distribuzione dei punteggi di una variabile in un campione
	• tendenza centrale	• modalità di aggregazione dei punteggi intorno ai valori medi
	- <i>media</i>	- valore medio dei punteggi di una variabile in un campione
	- <i>moda</i>	- punteggio che si rileva con maggiore frequenza
	- <i>mediana</i>	- punteggio che occupa la posizione centrale quando i valori della variabile sono ordinati in ordine crescente
	• variabilità	• entità della deviazione dei punteggi dalla tendenza centrale
	- <i>range</i>	- differenza tra il punteggio più basso e quello più alto
	- <i>varianza</i>	- indice che misura la tendenza dei punteggi a disporsi intorno alla media della distribuzione

□

Tutti gli strumenti che noi usiamo (e non solo le RS) hanno una capacità di risoluzione più o meno limitata in funzione dell'unità di misura adottata, anche se, per convenzione, si accetta che questo non limiti, all'atto pratico, la possibilità di osservare ogni misura possibile.

Poiché una misura ha un significato soltanto in rapporto ad altre, è necessario disporre di più di una misurazione della stessa variabile (p. es., valutazioni ripetute nel tempo di uno stesso soggetto o valutazione della stessa variabile in un gruppo di soggetti); l'insieme dei punteggi così ottenuti costituisce un *campione* (sample). Ogni campione è descritto dalla distribuzione che illustra la frequenza con cui compare ciascun valore di una data variabile (vedi Fig. 1.1).

Per sintetizzare questi dati e renderli più comprensibili ed utilizzabili anche a fini di confronto si utilizzano le misure di *tendenza centrale* e quelle di *dispersione* che sono in grado di "riassumere" e rappresentare l'insieme di tutte le misure effettuate.

La *tendenza centrale* è descritta da media, moda e mediana:

- la *media* (M) corrisponde alla somma (\cdot) di tutti i valori della variabile ($x_1, x_2, x_3... x_n$) divisa per il numero dei valori stessi (n), cioè delle osservazioni;
- la *moda* è il punteggio che si osserva con maggiore frequenza;
- la *mediana* è il punteggio che occupa la posizione centrale quando si ordinano i punteggi dal più basso al più alto.

Quando media, moda e mediana hanno lo stesso valore si dice che la distribuzione è di tipo "normale". Gli indici di tendenza centrale ci dicono dove sono concentrati i valori della variabile; gli indici di dispersione ci dicono quanto i valori si disperdono intorno alle misure di tendenza centrale. Il concetto di dispersione dei valori di una variabile è inscindibilmente legato alla variabilità: se non ci fosse variabilità non ci sarebbe dispersione di valori. Sono misure di dispersione: il range, la varianza e la deviazione standard.

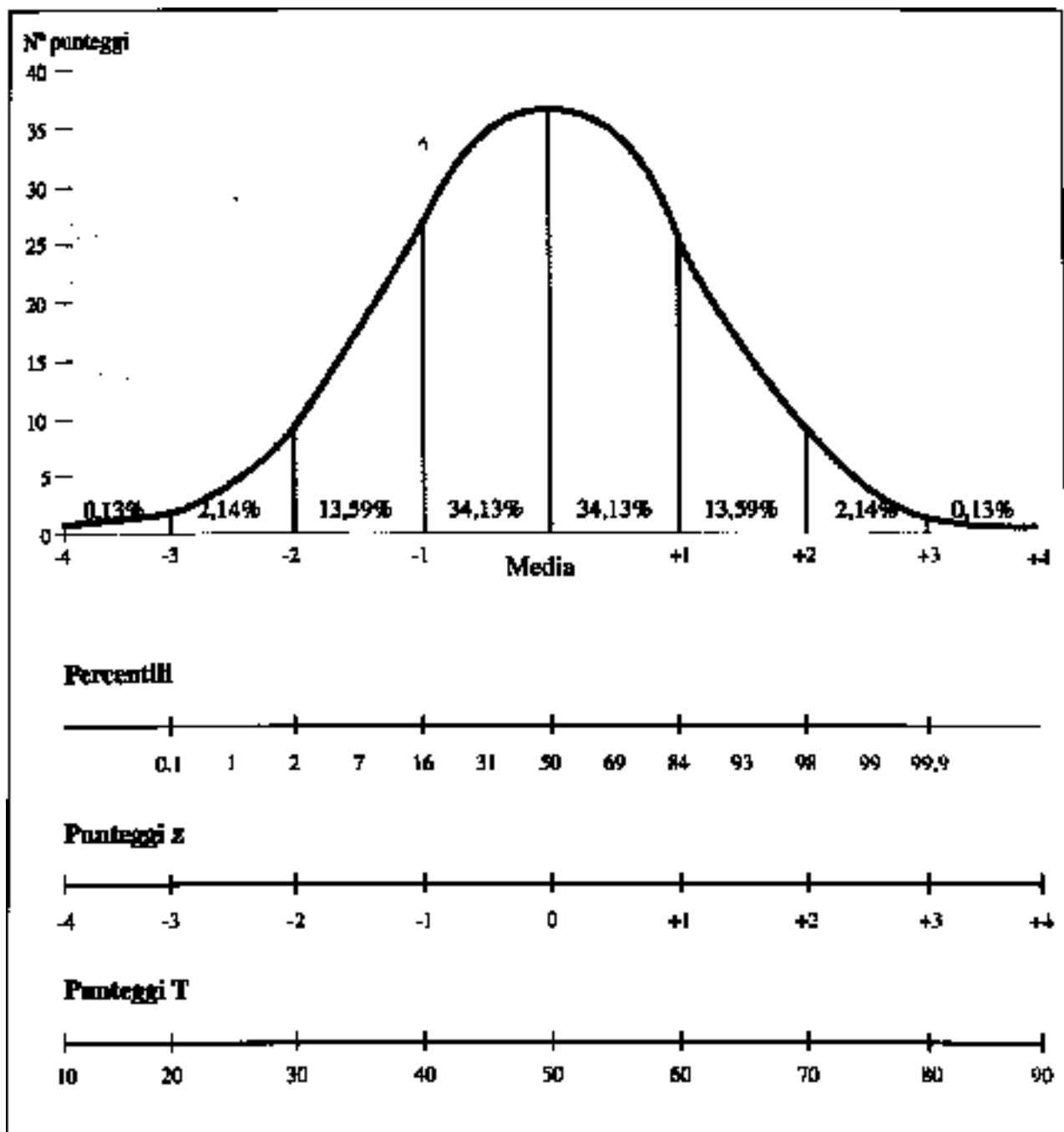
Il *range* (o intervallo) è la differenza tra il punteggio più basso e quello più alto.

La *varianza* è un indice che misura la tendenza dei punteggi a raggrupparsi intorno alla (o ad allontanarsi dalla) media dell'intera distribuzione. In termini matematici, la varianza è espressa dalla sommatoria (\sum) della differenza (elevata al quadrato) fra il punteggio di ciascuna osservazione (X) e la media del campione (M), divisa per il numero delle osservazioni (n) meno uno, secondo la formula:

$$\sum (X-M)^2 / (n-1)$$

La radice quadrata della varianza è la *deviazione standard* (standard deviation - SD) che esprime la deviazione dalla media, cioè di quanto i punteggi di una variabile si discostano dalla media. La SD può indicare anche la percentuale dei punteggi più alti o più bassi rispetto ad un determinato valore: in presenza di una distribuzione normale, rappresentata graficamente dalla curva di Gauss (Fig. 1.1), una metà del campione presenta valori al di sopra della media e l'altra metà al di sotto; il 68.26% del campione presenta valori compresi fra la media \pm una SD, il 95.45% fra la media \pm 2SD ed il 99.73% fra la media \pm 3SD.

I punteggi che si ottengono con i vari strumenti di valutazione possono essere di diverso tipo o possono essere variamente elaborati; quelli che raccogliamo quando facciamo delle valutazioni sono i *punteggi grezzi* (raw scores), cioè la misurazione immediata, diretta, della variabile. Quando si usano più RS, i cui item abbiano un diverso range di punteggio, i punteggi grezzi sono di scarsa utilità perché non consentono un confronto fra le diverse scale (avrebbe poco senso, infatti, confrontare un punteggio di 35 di una scala che ha un range da zero a 60 con uno di 50 di un'altra scala con range da 30 a 100). È necessario, quindi, trasformare in qualche modo i punteggi grezzi per poterli rendere confrontabili tra loro. Una delle trasformazioni più comuni (anche se forse non è quella che dà più informazioni) è quella in *percentili* (percentile ranks) (Fig. 1.1): i percentili indicano la percentuale dei punteggi più bassi di un determinato punteggio grezzo. Con la trasformazione in percentili, la mediana rappresenta il 50° percentile (il 50% dei punteggi del campione è, cioè, al di sotto di quel valore); i punteggi che si collocano, rispettivamente, a - 3, - 2 e -1 SD corrispondono al 1°, al 2° ed al 16° percentile, quelli che si trovano a +1, +2 e +3 SD, corrispondono all'84°, al 98° ed al 99,9° percentile. Questo tipo di punteggi può essere utile quando si confrontano valutazioni effettuate su popolazioni o a intervalli di tempo differenti; così, ed esempio, conoscendo la distribuzione normale del quoziente di intelligenza (QI) in una determinata popolazione, si può stabilire in quale posizione si colloca il QI di un determinato soggetto rispetto a quella popolazione, a quale percentuale di essa, cioè, egli appartiene. Utilizzando strumenti di valutazione con unità di misura diverse, con questa tecnica si può vedere comparativamente l'andamento rispetto a ciascuno strumento



Un confronto più diretto fra punteggi ottenuti con strumenti diversi si può comunque ottenere ricorrendo ai punteggi normalizzati o standardizzati (standard scores). I punteggi standardizzati di più comune impiego sono i punteggi z (z scores) ed i punteggi T (T scores) (Fig. 1.1). I punteggi z si calcolano portando a zero la media della distribuzione e facendo uguale ad 1 la SD; questi punteggi ci dicono, in sostanza, quanto il nostro punteggio si discosta dalla media. Con questa trasformazione i valori inferiori al punteggio medio danno numeri negativi, che possono essere fastidiosi da maneggiare nelle elaborazioni statistiche.

Per non avere punteggi negativi si può ricorrere ai *punteggi T* (T scores) che hanno media arbitraria di 50 e SD di 10. I punteggi T sono impiegati in numerosi strumenti di valutazione quali, ad esempio, il *Minnesota Multiphasic Personality Inventory - MMPI* (1967) ed il *Questionario per la Tipizzazione dell'Aggressività - QTA* (1957).

Come abbiamo già detto, un numero (e quindi anche il punteggio di una scala), per avere un significato, deve essere confrontato con altri numeri (o punteggi). Due sono i principali tipi di confronto, quello con un punteggio "normale" e quello con altre valutazioni dello stesso soggetto. Nel primo caso, si valuta di quanto il punteggio attuale si discosta da quello "normale", dal punteggio, cioè, idealmente rappresentativo della popolazione generale. L'impiego di strumenti che utilizzano questo tipo di confronto richiede spesso particolari cautele: è sempre opportuno accertarsi che gli strumenti siano stati testati su campioni sufficientemente rappresentativi, che il paziente faccia parte di quel tipo di popolazione (che, ad esempio, se il nostro paziente è un adolescente, lo strumento sia stato testato sugli adolescenti), che i dati "normali" siano abbastanza recenti, data la rapida evoluzione della società e delle sue norme socioculturali. Più semplice è il confronto con valutazioni precedenti del soggetto effettuate con lo stesso strumento. Questo sistema consente di valutare le variazioni del comportamento del soggetto nel corso del tempo, eventualmente in rapporto a specifici trattamenti, non richiede la trasformazione dei punteggi ed è sempre appropriato, essendo i punteggi utilizzati quelli del soggetto stesso e non di un campione che potrebbe essere non adeguato, non rilevante o cronologicamente non corrispondente.

Quando è stata raccolta in maniera standardizzata una serie di valutazioni di campioni sufficientemente ampi di pazienti, è possibile sottoporre i dati raccolti a più complesse analisi statistiche che possono essere variamente finalizzate. Ma questo è un tema che esula dallo scopo di questo lavoro e che lasciamo ai numerosi trattati di statistica.

BIBLIOGRAFIA

American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. Washington, DC, American Psychiatric Association, 1994.

Aitken R: *Measurement of feelings using visual analogue scales*. Proc Royal Soc Med, 62:989, 1969.

Buss A, Durkee A: *An inventory for assessing different kinds of hostility*. J Consult Psychol, 21:343, 1957.

Endicott J, Spitzer RL et al.: *The Global Assessment Scale: A procedure for measuring overall severity of psychiatric disturbance*. Arch Gen Psychiatry, 33:766, 1976.

Hathaway SR, McKinley JC: *Minnesota Multiphasic Personality Inventory Manual, Revised edition*. Psychological Corporation, New York, 1967.

LA VALUTAZIONE STANDARDIZZATA

GENERALITA'

Il colloquio psichiatrico, per quanto abbia insite anche valenze di altro tipo e significato, ha innanzitutto uno scopo informativo ponendosi, in sostanza, *quattro obiettivi*:

- *raccogliere informazioni* sui disturbi presenti al momento e sugli antecedenti dell'episodio attuale per arrivare alla formulazione di una diagnosi;
- *capire gli atteggiamenti ed i sentimenti* del paziente circa le sue condizioni ed i suoi sintomi;
- *cogliere il comportamento non verbale* per meglio definire la natura dei problemi psichici;
- *raccogliere, attraverso il comportamento nel corso dell'intervista, informazioni sulle caratteristiche del rapporto interpersonale*, spesso disturbato nel corso della patologia psichica.

Il successo di un colloquio psichiatrico si valuta in base alla sua capacità di fornire delle risposte esaurienti a questi 4 punti.

È del tutto verosimile che uno psichiatra ragionevolmente esperto sia in grado, nella sua pratica clinica, di ricavare dal colloquio psichiatrico tutte le informazioni necessarie per porre la diagnosi, impostare un trattamento e seguire adeguatamente l'evoluzione del quadro clinico del suo paziente. In ambito di ricerca, però, il clinico non sempre rappresenta un punto di riferimento attendibile, poiché le sue rilevazioni non sempre sono accurate, precise e sistematiche, non sempre tiene conto, nella formulazione della diagnosi, di tutte le possibili diagnosi differenziali e dei criteri di inclusione, di esclusione e di gravità. È stato perciò necessario far ricorso a strumenti, le RS, che lo obbligassero ad essere accurato nell'esplorazione della sintomatologia, che gli fornissero le necessarie informazioni sui criteri di inclusione e di esclusione facilitando così la classificazione diagnostica. In altri termini, è stato necessario far ricorso a dei sistemi di valutazione standardizzata che limitassero, per quanto possibile, la variabilità insita nell'approccio clinico.

La "valutazione standardizzata" è la tecnica che si propone di esplorare in maniera sistematica, ed il più possibile omogenea, settori più o meno vasti della psicopatologia in modo da ottenere dai diversi pazienti risposte confrontabili tra loro.

In senso assoluto, nella valutazione standardizzata il tipo di domande, le circostanze ambientali, la tecnica di valutazione e quant'altro possa influire sulla elicitazione delle risposte dovrebbe essere altamente standardizzato in modo da evitare l'introduzione di ogni elemento di variabilità. In pratica, però, con il termine di valutazione standardizzata si designano modalità diverse di raccolta delle informazioni in funzione di ciò che appare più appropriato nelle diverse circostanze.

Esistono perciò numerosi strumenti (RS, questionari, checklist, eccetera) per la valutazione standardizzata che consentono di raccogliere le informazioni da punti di vista diversi; essi vanno dall'autovalutazione da parte del paziente (p. es., *Symptoms Check List-90 - SCL-90*, *Beck Depression Inventory - BDI...*), all'eterovalutazione da parte dello psichiatra attraverso un'intervista più o meno rigidamente strutturata (p. es., *Structured Clinical Interview for DSM - SCID*, *Social Adjustment Scale - SAS...*) o attraverso un colloquio più o meno informale e l'osservazione del comportamento (p. es., *Hamilton Rating Scale for Depression - HAM-D*, *Brief Psychiatric Rating Scale - BPRS...*), dalla valutazione da parte di altre figure professionali (p. es., *Nurses' Observation Scale for Inpatient Evaluation - NOSIE*, *Psychotic Inpatient Profile - PIP...*) a quella da parte dei familiari (p. es., *Katz Adjustment Scale - KAS...*). Gli strumenti disponibili

hanno varia lunghezza (da un solo item a 100 e più item) e diversa difficoltà di applicazione e richiedono perciò tempi più o meno lunghi di somministrazione.

I vari strumenti possono avere un diverso grado di complessità valutando, ora semplicemente dei sintomi, ora la capacità del soggetto di mettere in atto stili di vita diversi, ora modalità comportamentali o affettive più o meno complesse.

Al fine di lasciare quanto meno spazio possibile all'interpretazione soggettiva, è necessario che, per ogni strumento:

- le domande siano sempre le stesse per tutti i soggetti intervistati e per lo stesso soggetto, se viene valutato più volte a diversi intervalli di tempo;
- le modalità di somministrazione siano il più possibile costanti;
- siano chiare le modalità di assegnazione dei punteggi che consentono la valutazione dell'importanza o della gravità del problema indagato da ciascun item;
- siano ben definite le unità di misura;
- sia specificato il periodo di tempo a cui si riferisce la valutazione (il momento dell'osservazione o il giorno precedente, l'ultima settimana o l'ultimo mese o tutta la vita);
- le variabili prese in considerazione si riferiscano a comportamenti, a sintomi, a disturbi piuttosto che a costrutti psicologici astratti.

Generalmente (ma non sempre!) gli Autori forniscono informazioni circa le proprietà psicometriche degli strumenti che hanno messo a punto, le finalità che si sono proposti ed i mezzi con cui hanno ritenuto di raggiungerle, le modalità di somministrazione e di assegnazione dei punteggi, i dati circa l'affidabilità e la validità e, quando appropriato, la norma di riferimento. La non disponibilità di queste informazioni non rende meno "standardizzato" uno strumento, ma, semplicemente, non fornisce all'utente tutti i criteri per scegliere gli strumenti da usare per gli scopi che si propone.

Lo scopo della valutazione standardizzata è sostanzialmente lo stesso di quello del colloquio clinico informale e consiste nel raccogliere informazioni sulla psicopatologia del soggetto: l'approccio quantitativo non dovrebbe andare a scapito della ricchezza e dell'individualità del colloquio clinico informale, poiché la valutazione clinica più efficace ed utile è quella che ha come fuoco dell'attenzione il singolo individuo ed il valutatore migliore è colui che è capace di interpretare i dati quantitativi in una prospettiva umanistica. I numeri non devono mai sostituire la valutazione personalizzata, ma essere, semmai, punto di partenza per una indagine più accurata ed approfondita e confluire nella descrizione del paziente.

Gli strumenti standardizzati di valutazione oggi disponibili coprono praticamente ogni area della psichiatria: comportamento e cognitività, sentimenti ed emozioni, ideazione e comportamenti d'abuso, sessualità e personalità, eccetera. Nella maggior parte dei casi prendono in esame, ciascuno nel suo specifico campo di indagine (più o meno vasto), la sintomatologia di superficie ed il comportamento che, generalmente, possono essere valutati in maniera abbastanza semplice, affidabile e quantificabile, senza bisogno di particolari interpretazioni.

I PRINCIPI DELLA VALUTAZIONE

La valutazione clinica standardizzata deve rispondere ad alcune regole fondamentali ben espresse

per la medicina in generale da Deyo e Carter (1992), ma ampiamente valide anche per la psichiatria:

1 - Gli strumenti di valutazione devono essere brevi, di facile impiego e clinicamente significativi poiché nella pratica clinica non sono proponibili strumenti che, per numero di item e/o per difficoltà di applicazione, richiedano un tempo eccessivo. Per essere più accettabili per il medico, questi strumenti dovrebbero fornire anche alcune informazioni relative alla diagnosi ed allo stato psichico, che sono il nucleo essenziale della valutazione psichiatrica.

2 - Gli strumenti di valutazione devono essere generali e completi in modo da esplorare l'intero arco della psicopatologia poiché i pazienti psichiatrici presentano spesso condizioni di comorbidità, o una patologia cosiddetta "di spettro", o quadri clinici complessi; infatti, in queste condizioni, sarebbe difficile e/o limitativo l'impiego di strumenti settoriali.

3 - Gli strumenti di valutazione devono essere standardizzati e metodologicamente validi in modo che le informazioni raccolte possano essere confrontabili con quelle raccolte da altri ricercatori. Frequente è la tentazione di creare strumenti di valutazione "personalizzati", tentazione alla quale è bene cercare di resistere poiché, in caso contrario, si perderebbe la possibilità di confrontare i propri dati con quelli degli altri, a meno di non affrontare il lungo e difficile cammino della validazione del nuovo strumento, ritenendo che esso stesso possa proporsi come standard di riferimento.

4 - Gli strumenti di valutazione devono essere tali da non richiedere un training lungo e faticoso per poterli utilizzare nella pratica clinica; per migliorarne l'affidabilità sarebbe opportuno che il punteggio dei singoli item avesse dei chiari "anchor point" riferiti ad un comportamento il più possibile obiettivo ed univocamente interpretabile.

5 - Almeno una parte degli item deve essere sensibile tanto da rilevare i cambiamenti della sintomatologia nel breve periodo, mentre altri item, cui è devoluto il compito di fornire una descrizione più globale del paziente ed indicazioni circa i suoi bisogni e di predire l'evoluzione della malattia, devono essere più stabili.

6 - Le variabili prese in considerazione non devono avere un eccessivo grado di sovrapposizione tra di loro, anche se è inevitabile un certo grado di correlazione (e quindi di ridondanza).

7 - Gli strumenti di valutazione devono soddisfare le necessità dello psichiatra in termini di informazione e la loro somministrazione non deve risultare faticosa, pesante, tanto per il paziente quanto per lo psichiatra stesso.

CENNI STORICI

La valutazione standardizzata dei disturbi psichici era affidata, in passato, agli psicologi i quali utilizzavano batterie di test psicologici, in particolare psicometrici e psicodiagnostici, che richiedevano, per la loro somministrazione ed interpretazione, una particolare preparazione e tempi di solito abbastanza lunghi: questo ha rappresentato un serio ostacolo ad una più ampia diffusione della valutazione standardizzata della patologia psichiatrica. Questo tipo di valutazione, peraltro, era in un certo senso funzionale alla realtà clinico-terapeutica del momento, caratterizzata dalla pressoché totale mancanza di strumenti terapeutici efficaci e specifici e fortemente connotata dalla teoria e dalla prassi psicodinamica. La valutazione psicologica aveva perciò come oggetto le dinamiche, le difese, le fantasie dei pazienti piuttosto che un accurato inquadramento diagnostico, una precisa definizione qualitativa e quantitativa del quadro psicopatologico, un'attenta misurazione dei cambiamenti del quadro clinico a seguito degli interventi terapeutici.

La scoperta di trattamenti efficaci per la cura delle malattie mentali ha cambiato in maniera radicale la teoria e la prassi psichiatrica. Il cambiamento più importante può essere individuato nella significativa modificazione dell'evoluzione e del decorso di molte malattie che ha drasticamente ridimensionato la durata dei ricoveri, quando non li ha resi non necessari. La psichiatria deve oggi impedire, per quanto possibile, l'allontanamento del paziente dal suo contesto sociale e, quando l'allontanamento si rende necessario, cercare di renderlo il più breve possibile. Questa tendenza è stata amplificata dalla necessità di contenere i costi sociali (e quelli della sanità in primo luogo) per cui, anche per i trattamenti a livello ambulatoriale, si tende a preferire trattamenti di più breve durata e minor costo. In questo contesto la psicoterapia è stata relegata, in molti casi, in un ruolo secondario, accessorio, focalizzata su problemi specifici, privilegiando i trattamenti di breve durata.

Con l'avvento di terapie sempre più efficaci e mirate, anche il linguaggio della psichiatria è cambiato, il sistema diagnostico, ampiamente rivisitato, è stato reso più descrittivo, più basato sui sintomi, più standardizzato e, in questa ottica, più adatto alla valutazione mediante strumenti standardizzati di valutazione. Anche il parallelo sviluppo e l'affermarsi delle terapie cognitive e comportamentistiche ha favorito la messa a punto di strumenti di valutazione capaci di mettere bene a fuoco l'oggetto dell'intervento e verificarne l'effetto.

E così la valutazione psicologica ha ampiamente mutato i suoi scopi e le sue strategie:

- mettendo in primo piano la misurazione della gravità dei sintomi, si è passati dall'interesse verso la dinamica e le motivazioni a quello verso i disturbi e le prestazioni;
- gli strumenti di valutazione sono diventati più specifici per consentire lo studio di particolari aree di interesse e più orientati verso la documentazione dell'obiettività;
- la necessità di valutare rapidamente un numero consistente di pazienti in modo da avere le necessarie informazioni per impostare un trattamento o per modificare quello in atto, ha favorito lo sviluppo delle scale di autovalutazione;
- le RS, per le loro caratteristiche, si prestano in maniera eccellente all'impiego dei computer per la somministrazione, ma soprattutto per la loro elaborazione e per la produzione di rapporti narrativi;
- spostando l'attenzione dal soggetto alla sua relazione con l'ambiente, sono state portate all'attenzione aree in precedenza quasi totalmente ignorate come, ad esempio, l'adattamento sociale;
- grazie alle loro caratteristiche, infine, le RS hanno offerto allo psichiatra la possibilità di valutare direttamente i pazienti con un modesto investimento di tempo per la conoscenza degli strumenti e per il loro impiego routinario.

La data di nascita delle scale di valutazione di variabili di tipo soggettivo è collocata da più Autori in tempi diversi. Già nel II secolo d.C. Galeno propose una scala per la valutazione soggettiva della dimensione caldo-freddo articolata su 9 punti, 4 per la sensazione di freddo e 4 per quella di caldo in rapporto ad una condizione basale considerata come standard di confronto: questa scala, per quanto ideata per la misurazione di una variabile fisica, era in effetti una vera e propria RS in quanto l'attribuzione di un valore numerico alla variabile indagata dipendeva da un giudizio psicologico (basato su di una sensopercezione).

Certamente più vicina alle RS moderne è la scala ideata nel 1692 da Thomasius, un filosofo tedesco che, ipotizzando che la personalità fosse la risultante di 4 dimensioni principali, la sensualità, l'attitudine ad acquisire, l'ambizione sociale e l'amore razionale, creò una scala per misurare

queste "inclinazioni" nei soggetti.

Altri esempi di scale di valutazione sono riportati in letteratura, ma fu certamente Francis Galton che introdusse la metodologia delle scale di valutazione nell'ambito della psicologia: nel 1880, infatti, egli pubblicò una scala per la valutazione della chiarezza delle immagini mentali. Ma, cosa ancora più importante, egli introdusse il concetto fondamentale che i punti della scala dovessero essere separati da intervalli uguali.

La RS che ha avuto maggiore applicazione nella pratica clinica è comunque quella elaborata da Kempf nel 1915 per la valutazione dei tratti psicopatologici.

Tuttavia è solo negli anni Cinquanta, con l'avvento dei primi farmaci psicotropi, che si è sentita pressantemente l'esigenza di ottenere valutazioni attendibili dello stato psichico dei pazienti e delle variazioni indotte su di esso dai trattamenti e che, nell'impossibilità di ricorrere a marker bio-umoral, si è fatto ricorso a strumenti che descrivessero, in maniera sistematica ed estensiva, le diverse aree del comportamento psicopatologico e che si prestassero alla sua misurazione. Sono nate così le RS, per prime quelle per la valutazione della psicopatologia generale dei disturbi psicotici e, successivamente, quelle per la valutazione di più limitati settori della psicopatologia, in stretto parallelismo con l'avvento di farmaci attivi su specifiche patologie o con l'estendersi delle indicazioni dei farmaci già in commercio a settori emergenti della psicopatologia.

L'entusiasmo acceso dall'avvento degli psicofarmaci, che offrivano per la prima volta la possibilità di curare malattie fino ad allora considerate incurabili, è stato probabilmente la molla che ha suscitato in molti un parallelo entusiasmo per quegli strumenti (le RS, appunto) che degli effetti degli psicofarmaci davano documentazione, al punto da considerarli come capaci di dare "scientificità" alla psichiatria. Questo atteggiamento ha fornito un comodo bersaglio a quanti, da posizioni anche contrapposte (quali quella della psicoanalisi – o più genericamente della psicoterapia – da un lato, e, dall'altro, quella della cosiddetta "antipsichiatria", che a partire dalla fine degli anni Sessanta ha attraversato, scuotendolo dalle radici, il mondo della psichiatria) vedevano nella psicomelia un modo per obiettivare il paziente e la sua malattia, scotomizzando le complesse interrelazioni tra il paziente ed il suo ambiente familiare e sociale, impedendo l'incontro empatico con lui e contribuendo alla sua ulteriore emarginazione attraverso l'attribuzione, secondo un procedimento paleo-positivista, di un valore assoluto a fenomeni che sono invece espressione di dinamiche personali e sociali più complesse. Il fronte delle critiche era variegato, spaziando da quello più estremo, che vedeva nelle RS "soltanto uno degli strumenti utilizzati fraudolentemente dagli psichiatri per la cinica conservazione del loro potere" (Faravelli, 1983), a quello più moderato, che sottolineava che le RS, tenendo conto soltanto dell'aspetto sintomatologico della malattia, finivano per isolarla dal suo contesto impedendone, in definitiva, la comprensione.

Com'è facile intuire, entusiasmi eccessivi e critiche feroci sono, ognuno per la sua parte, totalmente fuori misura, in quanto attribuiscono a questi strumenti proprietà e capacità che non hanno; la "portata teorica delle RS deve essere ricondotta entro i limiti sanciti dalla semplicità concettuale di questi strumenti, che non rappresentano certamente un contributo essenziale alla "scientificità" della psichiatria, né un attentato ad una comprensione più globale della devianza psichica, e che sono, invece, più semplicemente, un comune schema di riferimento che consente la trasposizione numerica di alcuni elementi obiettivabili della sintomatologia clinica, rendendoli così confrontabili ed analizzabili con tecniche statistiche" (Conti et al., 1993). L'impiego di questi strumenti, infatti, non può e non deve sostituirsi all'approccio clinico, ma deve essere limitato ad ambiti di ricerca specifici che richiedono l'impiego di tecniche statistiche ed informatiche che solo la trasposizione numerica delle informazioni cliniche rende possibile.

È doveroso comunque sottolineare come la ricerca psicofarmacologica clinica scientificamente fondata, abbia stimolato lo sviluppo di strumenti sempre più sensibili, capaci di valutare

accuratamente l'omogeneità dei campioni sperimentali, di cogliere modificazioni del quadro clinico anche modeste, di individuare con precisione gli aspetti psicopatologici sui quali maggiormente si esplica l'attività terapeutica; per contro, la più ampia disponibilità di farmaci efficaci ha reso necessario un più accurato inquadramento diagnostico ed una più precisa definizione delle caratteristiche psicopatologiche del quadro clinico ai fini di un intervento clinico-terapeutico più mirato e, di conseguenza, più efficace: ed è dunque in questa prospettiva che l'impiego delle RS diventa un necessario complemento dell'approccio clinico tradizionale.

Vorremmo, comunque, concludere questo rapido excursus con un'affermazione "dissacrante" di Max Hamilton, il "padre" di due RS per la valutazione della depressione e dell'ansia così famose e così diffusamente impiegate da essere considerate il termine di paragone per tutte le RS successive: "Le scale di valutazione sono oggi così diffusamente impiegate, soprattutto per la valutazione dei trattamenti nelle ricerche psicofarmacologiche cliniche, che è difficile capire che esse sono relativamente nuove e che, d'altra parte, esse non sono indispensabili neppure nelle ricerche terapeutiche. Ci sono altri mezzi per valutare le condizioni del paziente, come la durata del ricovero o dell'assenza dal lavoro per malattia. Una scala di valutazione non è niente di più che un modo particolare per registrare un giudizio clinico. Il clinico esprime la sua opinione sulla presenza o assenza di un sintomo, o sulla sua gravità, ma che egli la esprima con parole o con numeri, il giudizio non cambia. Se il giudizio è basato su informazioni inadeguate, o se il clinico non ha abbastanza esperienza, il giudizio non ha alcun valore, tanto che sia espresso con parole che con numeri. Le ulteriori analisi statistiche stupiranno soltanto coloro che non capiscono o non vogliono capire queste cose. Ciò nonostante, le scale di valutazione sono popolari poiché permettono di ottenere un'informazione standardizzata: gli item sono gli stessi per tutti i pazienti ed in tutte le occasioni, e sono definiti in modo che la loro graduazione ed il loro modo di impiego siano uniformi. Inoltre, sono facili da quantificare e consentono di ottenere dei numeri agevolmente utilizzabili per le analisi statistiche. Molti clinici hanno ancora seri dubbi sulla quantificazione, ma i metodi attuali hanno una solida base teorica. Ma, cosa più importante, [le scale di valutazione] sono soddisfacenti nella pratica clinica" (Hamilton, 1981).

LE BASI TEORICHE DELLA VALUTAZIONE

La ricerca scientifica si avvale sostanzialmente di due strategie di ricerca che possiamo definire una *intensiva* e l'altra *estensiva*. La prima, che viene effettuata su campioni limitati o anche su casi singoli, non risponde ad una metodologia rigidamente formalizzata e prevede la raccolta di numerose informazioni che, nel loro complesso articolarsi tra loro, possono fornire al ricercatore elementi per formulare nuove ipotesi. La seconda, quella estensiva, si basa sullo studio della distribuzione di specifiche variabili in campioni sufficientemente ampi ed opportunamente selezionati, e consente la verifica delle ipotesi formulate nell'ambito della ricerca intensiva. In definitiva, quindi, la ricerca intensiva appare più adatta per la formulazione di nuove ipotesi, e quella estensiva per la verifica della "universalità" di quelle ipotesi che, se dimostrata, trasforma le ipotesi in scoperte.

Le RS sono, almeno per la psichiatria, gli strumenti più idonei per la ricerca estensiva poiché consentono di effettuare numerose osservazioni, perché le osservazioni su individui diversi, trattandosi di strumenti standardizzati, sono comparabili tra di loro e perché, infine, la rilevante quantità di materiale che forniscono è espresso in forma numerica ed è perciò elaborabile mediante le tecniche statistiche.

Si potrebbe obiettare che, per quanto numerose possano essere le variabili esplorate da una RS, si tratta sempre di un approccio riduttivo rispetto alla globalità del fenomeno indagato; non dobbiamo tuttavia dimenticare il grande contributo, sul piano operativo, fornito alla ricerca dal riduzionismo scientifico: l'importante è essere consapevoli della riduzione operata. D'altronde, qualsiasi approccio pre-teorizzato è, in varia misura, riduzionistico, così come lo sono, del resto,

anche gli approcci programmaticamente ateoretici. Dobbiamo anche dire che, se ciascun fenomeno ha degli aspetti suoi peculiari che lo caratterizzano rispetto agli altri fenomeni dello stesso tipo, è anche vero che, per essere definiti "fenomeni dello stesso tipo", devono avere necessariamente degli aspetti che li accomunano. Sono proprio questi aspetti a comune che costituiscono la base della prassi scientifica: se ogni fenomeno costituisse un caso a sé, non ci sarebbero né la scienza né la cultura, l'esperienza non avrebbe alcun valore e non sarebbe possibile alcuna predizione. In questa ottica, quindi, l'uso delle RS non è altro che un aspetto del riduzionismo scientifico ed è perciò ampiamente giustificato (Faravelli, 1983).

Se, dunque, in questa prospettiva la pre-selezione delle variabili può essere accettabile, ciò che invece potrebbe proporsi come oggetto di discussione è la standardizzazione dei concetti psicopatologici che sono, di per sé, complessi e difficilmente standardizzabili.

È anche vero che è criticabile tanto la scomposizione del quadro psicopatologico in un mosaico di segni e sintomi, quanto la valutazione della gravità e/o del miglioramento in base al punteggio totale, cioè alla somma dei punteggi dei singoli item, poiché entrambe le operazioni non rispettano la realtà clinica; tuttavia dobbiamo riconoscere che, fino ad oggi, la valutazione standardizzata mediante RS e questionari è risultata il metro di giudizio più obiettivo di cui disponga la psichiatria. Del resto, l'ampia diffusione che questi strumenti hanno raggiunto testimonia efficacemente di quale entità fosse l'esigenza di misurazione e di standardizzazione in questo campo. Quindi, per quanto, mediante le RS, si operi una ipersemplificazione di un'esperienza umana complessa ed irripetibile, qual è la malattia mentale o, più in generale, la vita psico-emotiva di un soggetto, è solo attraverso questa inevitabile forzatura che è possibile raggiungere un giudizio obiettivabile, comunicabile e comparabile fra osservatori o in tempi diversi. D'altro canto, come già detto, è bene ricordare che la valutazione standardizzata non può e non deve essere posta in alternativa al colloquio psichiatrico inteso nella più completa accezione del termine.

BIBLIOGRAFIA

Conti L, Ruggeri M, Faravelli C: *Strumenti di valutazione e di misura*. In: Cassano GB et al. (Eds.) *"Trattato Italiano di Psichiatria"*, Masson, Milano, 1993.

Deyo RA, Carter WB: *Strategies for improving and expanding the application of health status measures in clinical settings*. Medical Care, 30:176, 1992.

Faravelli C: *Le scale di valutazione in psichiatria*. Quaderni Italiani di Psichiatria, 2:97, 1983.

Galton F: *Statistics of mental imagery*. Mind, 19:21, 1880.

Hamilton M: *Les échelles d'appréciation dans la depression: état actuel*. In: Pichot P, Pull C (Eds.) *"La symptomatologie dépressive: enregistrement et évaluation"*, Editions Spire, Paris, 1981.

Kempf EJ: *The behavior chart in mental disease*. Am J Insan, 71:761, 1915.

FUNZIONI E PROPRIETA' DELLE RS

GENERALITA'

Le RS, come abbiamo visto, sono strumenti costituiti da una serie di item che definiscono gli elementi psico-comportamentali da osservare ed indicano, con diverso grado di precisione e formalizzazione, i livelli di gravità a cui fare riferimento. Possono essere impiegate come *checklist di riferimento personale*, come *strumenti di rilevazione standardizzata* o, più compiutamente, come *strumenti di misura*.

a - Al livello più elementare, le RS possono essere considerate delle *checklist di riferimento personale*, cioè delle liste di variabili che consentono al valutatore di esplorare tutti gli aspetti rilevanti dell'argomento da indagare con la certezza di non dimenticarne qualcuno per disattenzione o per dimenticanza.

b - Poiché la scelta delle variabili che compongono una RS è operata, a priori, dall'Autore dello strumento, tutti coloro che utilizzano una determinata RS si uniformeranno, di necessità, nella raccolta delle informazioni operando di fatto una *standardizzazione della valutazione*. La raccolta delle stesse informazioni da parte di valutatori operanti anche in contesti diversi, rappresenta soltanto il primo passo per la standardizzazione della valutazione; livelli ulteriori di standardizzazione possono essere raggiunti se le RS specificano, oltre alle variabili da esplorare, anche i criteri da utilizzare per effettuare la valutazione e le modalità di assegnazione dei punteggi. È bene ricordare, tuttavia, che nonostante l'elevato grado di standardizzazione raggiunto dalle RS, fattori diversi possono influire sull'attendibilità di questi strumenti e che di essi è necessario tener conto nel loro impiego. Come chiaramente mostra la tabella 3.I, i fattori che possono ridurre l'attendibilità delle RS sono sostanzialmente di *tre tipi*, quelli *legati al soggetto che effettua la valutazione*, quelli *socioculturali* e quelli *semantici*; la loro influenza è diversa se la valutazione è fatta da un osservatore esterno o dal paziente stesso. Dobbiamo comunque dire che l'influenza di questi fattori può manifestarsi su qualsiasi modalità di rilevazione del vissuto psicopatologico e che, semmai, grazie agli studi di taratura, nelle RS questa influenza può essere in parte eliminata e, per quella parte non eliminabile, può essere "pesata". Per una più esauriente trattazione di questo tema si rimanda al lavoro di Poli e Collaboratori (1981).

c - Le RS sono, infine, *strumenti di misura* e, in quanto tali, devono rispondere sostanzialmente a *due principi*: il primo, specifico per la misurazione dei singoli item, è quello dell'*omomorfismo tra il sistema numerico e quello empirico*, il secondo, che riguarda le scale composte da più item, è quello del *rapporto tra i punteggi della scala e la realtà empirica*.

Il primo caso, che è quello della *valutazione di gravità unilineare*, riguarda la distinzione di livelli diversi di gravità nell'ambito di un singolo item. È evidente che, rispetto all'item in questione, esiste omomorfismo tra i punteggi e la realtà oggettiva se i punteggi di due soggetti (A e B) sono uguali solo se il soggetto A è grave quanto B o, se il punteggio di A è maggiore (o minore) di quello di B, A è più (o meno) grave di B. Detto in questo modo, sembrerebbe che il problema potesse essere risolto semplicemente mediante un confronto con un criterio esterno; in realtà questo confronto non è necessario se il sistema empirico mantiene lo stesso rapporto ordinale del sistema numerico, se, cioè, non è messa in discussione la transitività dell'ordinamento. In altre parole, se A è più grave di B, B non può essere più grave di A. È possibile, infatti, che la differenza tra A e B e quella tra B e C sia inferiore alla soglia di discriminazione, per cui è corretto dire che $A=B$ e $B=C$; ma la differenza fra A e C potrebbe superare tale soglia per cui sarà $A < C$ in apparente contrasto con la proprietà transitiva secondo la quale, se $A=B$ e $B=C$, $A=C$. Se, ad esempio, su di una scala da 0 a 100, abbiamo stabilito che differenze fino a 5 punti non sono discriminanti, possiamo avere che il soggetto A, che totalizza 60 punti, sia da considerarsi uguale al soggetto B, che totalizza 63 punti, e che questi sia uguale al soggetto C che ha 67 punti, ma A e C non potranno essere

considerati uguali poiché la differenza tra i loro punteggi è superiore alla soglia di discriminazione. Incongruenze di questo tipo sono tollerabili poiché non compromettono l'ordinalità del sistema; altri tipi di incongruenza (p. es., $A > B$, $B = C$, $C > A$) sono del tutto improbabili, potendo verificarsi solo quando le specifiche per l'attribuzione del punteggio sono confuse o mal definite. Per evitare questi rischi è perciò opportuno che i criteri di attribuzione del punteggio siano il più possibile discreti ed univoci (Faravelli, 1983).

TAB. 3.1 - FATTORI CHE INFLUENZANO L'ATTENDIBILITÀ DELLE RS*		
	Eterovalutazione	Autovalutazione
Fattori individuali (legati al valutatore)	<i>Fattori sistematici</i>	
	▪ tendenza ai valori medi (evitamento dei valori estremi)	▪ stato morboso (impossibilità o inattendibilità dell'autovalutazione)
	▪ effetto alone (influenza della valutazione degli altri item)	▪ livello culturale
	▪ tendenza ad ipervalutare i comportamenti più espressivi	▪ capacità di introspezione
	▪ distorsioni cosce e inconscie (utilizzo della scala come strumento di comunicazione)	▪ fattori di personalità
	<i>Fattori non sistematici</i>	
	▪ addestramento alla valutazione ▪ esperienza clinica ▪ background culturale ▪ qualità del rapporto col paziente	▪ qualità del rapporto con il medico
Fattori socioculturali	▪ cultura psichiatrica di chi ha costruito la scala	▪ inaccessibilità al p. di determinati item (p.es., l'introspezione)
	▪ valori sociali (maggiore peso ai comportamenti socialmente indesiderabili)	▪ valori sociali (desiderabilità/ indesiderabilità sociale dei diversi aspetti sintomatologici)
	▪ fattori transculturali (uso della RS in contesti culturali diversi da quello per il quale è stata creata)	▪ fattori transculturali (diversa percezione della malattia nelle diverse culture)
Fattori semantici	▪ chiarezza di formulazione	▪ chiarezza di formulazione (maggiore variabilità culturale tra pazienti che tra gli psichiatri)
	▪ specificazione sul come attribuire i punteggi	▪ specificazione sul come attribuire i punteggi
	▪ traduzione (per scale concepite in altra lingua)	▪ traduzione (per scale concepite in altra lingua)

Il secondo caso è quello della *valutazione di gravità multilineare*, in cui i criteri di gravità sono desunti da scale composte da più item. Generalmente la somma dei punteggi dei singoli item viene assunta come criterio di gravità della sindrome esplorata anche se, in realtà, i valori numerici che si ottengono non sono, generalmente, strettamente correlati con la situazione oggettiva. Se assumiamo come elemento di confronto un criterio esterno come, ad esempio, il giudizio clinico, è

facile che si possano mettere in evidenza incongruenze anche rilevanti fra il giudizio del clinico ed i punteggi delle RS, ma anche fra i punteggi di RS differenti. Il giudizio clinico, infatti, tiene conto sia della gerarchia dei sintomi (p. es., sintomi nucleari e sintomi accessori) che delle relazioni reciproche tra i sintomi (p. es., un sintomo può avere valore in presenza di un altro ma non da solo), mentre il punteggio della RS non è altro che la somma algebrica dei valori dei singoli item e rappresenterà perciò una via di mezzo tra la gravità e la pervasività della sintomatologia potendo uno stesso punteggio esprimere tanto una marcata gravità di un numero limitato di sintomi che una moderata gravità di un numero maggiore di sintomi. Ancora più rilevante può essere la discrepanza tra la gravità oggettiva e quella espressa mediante il punteggio delle scale di autovalutazione che, in assenza del filtro operato dal clinico, è ancora di più influenzato da fattori marginali, non ultimo la maggiore o minore tendenza del soggetto all'istrionismo ed alla teatralità.

Un'altra critica che viene comunemente sollevata alle RS è quella relativa alla loro reale "ordinalità". È evidente che il livello di ordinalità è in rapporto alla sensibilità della scala, ma è comunque opinabile che differenze modeste di punteggio totale corrispondano a reali differenze di gravità: è infatti difficile sostenere che, alla scala di Hamilton per la depressione (*Hamilton Depression Rating Scale - HAM-D o HDRS*), un punteggio totale di 30 esprima una gravità realmente maggiore rispetto ad uno di 29 o 28, mentre è probabile che ciò sia vero se si confronta un punteggio di 30 con uno di 22 o 23. Le RS sono dunque da considerare, almeno per quanto riguarda i punteggi grezzi, scale imperfettamente ordinali che corrispondono solo in misura più o meno approssimativa a ciò che si propongono di misurare. Ciò non toglie che, all'interno di un campione, siano ampiamente accettabili le inferenze statistiche; più problematica è l'accettazione dei risultati dei test statistici quando si vogliono applicare i risultati relativi al campione in esame, assunto come rappresentativo del fenomeno, all'universo dei fenomeni considerati (Faravelli, 1983).

Le RS, come abbiamo detto, misurano concetti clinici complessi mediante la loro scomposizione in elementi più semplici che sono ritenuti di più immediata valutazione e misurazione. Così, ad esempio, per misurare la depressione, noi valutiamo l'inibizione psicomotoria, la perdita di interessi, l'insonnia, le idee di colpa, l'ideazione suicidaria, eccetera. Quando, per misurare la gravità dell'entità morbosa in esame, sommiamo i punteggi dei singoli item è come se volessimo ridurre lo spazio multidimensionale individuato dalle singole componenti ad una retta; questo comporta una perdita di informazione tanto maggiore quanto più numerose sono le componenti in cui abbiamo scomposto il fenomeno e quanto minore è la correlazione tra di esse. Se, ad esempio, il fenomeno clinico A fosse scomposto in due soli item che esplorano le componenti B e C, che empiricamente abbiamo osservato presentarsi generalmente correlate tra loro, potremmo rappresentare A come un piano cartesiano delimitato dagli assi B e C; il singolo soggetto (a) sarebbe così individuato, rispetto al fenomeno A, dal punto in cui si incrociano i valori di B e di C (Fig. 3.1). I soggetti di un campione ($a_1, a_2, a_3 \dots a_n$) si distribuiranno nel piano A lungo un asse obliquo: la dispersione attorno a questo asse sarà tanto minore quanto maggiore è la correlazione tra B e C e la perdita di informazione è direttamente proporzionale all'entità della dispersione.

Nella realtà, noi generalmente scomponiamo un fenomeno psicomorbo in numerose componenti; ognuna di queste, nella correlazione con ciascuna delle altre, individua una serie di piani intersecantisi tra di loro. Una valutazione di gravità secondo un unico asse presupporrebbe una così stretta correlazione tra le diverse componenti che i punteggi finirebbero per disporsi su di un ellissoide molto affusolato ad n dimensioni (dove n corrisponde al numero delle componenti). Questa condizione (peraltro già riduttiva) raramente si verifica ed allora, nella pratica, si ricorre, mediante tecniche statistiche (di solito di tipo fattoriale), all'identificazione di un asse più significativo; questo comporta una più o meno marcata perdita di informazione, al punto che il fattore che misuriamo può, in certi casi, avere scarsa correlazione con la realtà clinica.

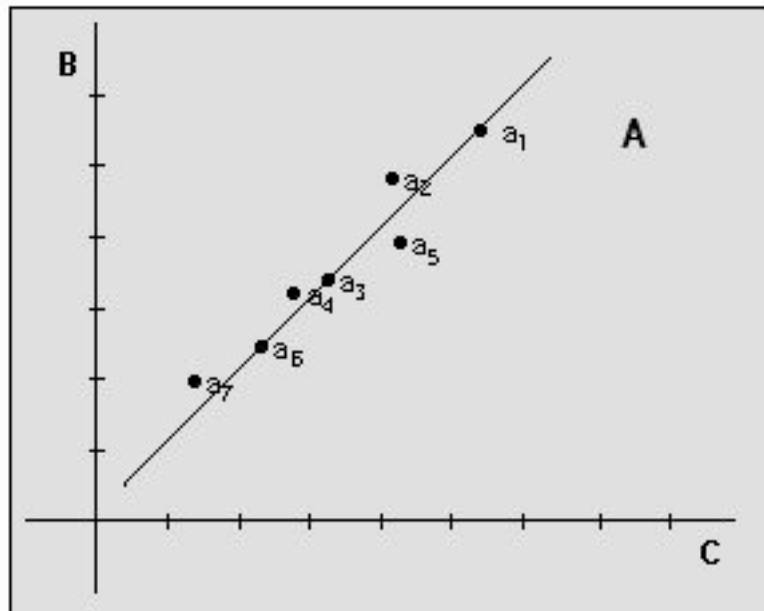


Fig. 3.1 - Rappresentazione del concetto di misurazione di una ipotetica entità clinica (A) attraverso la valutazione delle sue due componenti (B e C).

La mancanza di conoscenze intrinseche sul come e sul perché i singoli aspetti di un quadro psicopatologico sono tra loro collegati rende impossibile stabilire se e come raggruppare i valori delle diverse componenti in un punteggio unico che abbia un reale significato clinico. Si potrebbe obiettare che un paziente con un punteggio di 30 alla HAM-D ha, con molta probabilità, una depressione più grave rispetto ad un altro con punteggio di 15, ma questo non ci indica che cosa rappresenta quel "più", se cioè sono più gravi i sintomi nucleari o se è maggiore la componente ansiosa o se hanno un peso rilevante i sintomi accessori (ossessivi, paranoidei, di depersonalizzazione, eccetera).

Il punteggio totale non appare quindi il modo migliore e più appropriato per la definizione della gravità del quadro psicopatologico; più appropriata ed accettabile appare invece la quantificazione mediante un profilo, ricorrendo eventualmente, dato il gran numero di item che spesso compongono le RS, alla riduzione del numero delle variabili mediante procedimenti matematici (fattori, componenti principali) o empirici (cluster).

PROPRIETA' DELLE RS

Una misura può essere considerata "buona" se è affidabile, sensibile e valida (Tab. 3.II); l'affidabilità (reliability) dipende dalla consistenza dello strumento (i cui item devono misurare sempre ed allo stesso modo la stessa entità), la sensibilità è la sua capacità di discriminare fra fenomeni diversi, la validità (validity) dipende dall'accuratezza con cui misura ciò che si presume debba misurare. La mancanza di affidabilità e di validità è responsabile, rispettivamente, di errori casuali e di errori sistematici.

TAB. 3.II - PROPRIETÀ GENERALI DELLE RS	
Validità	Capacità dello strumento di misurare ciò che ci si propone di misurare
Sensibilità	Capacità dello strumento di discriminare tra fenomeni diversi
Affidabilità	Possibilità di ripetere la misurazione dello stesso fenomeno in tempi successivi o da parte di osservatori diversi con risultati simili

Validità

TAB. 3.III - CARATTERISTICHE DELLE RS: VALIDITÀ	
Per contenuto	Esplorazione di tutti i sintomi che concorrono a definire una specifica sindrome
Per criterio (o empirica)	Correlazione tra lo strumento ed altre misure riconosciute come affidabili
Concorrente	Relazione tra i risultati forniti dallo strumento ed uno o più criteri esterni
Predittiva	Capacità di predire l'evoluzione del fenomeno in esame
Discriminante (o di gruppo)	Capacità di discriminare soggetti che sappiamo avere i problemi che esplora lo strumento da soggetti che non hanno tali problemi
Per costrutto	Riferimento ad un modello teorico del concetto esplorato
Fattoriale	Verifica della validità concorrente e discriminante mediante l'analisi fattoriale
Nominale (o face validity)	Corrispondenza degli item che compongono lo strumento ad aspetti reali dell'entità clinica in esame

La *validità* è la capacità di una RS di misurare effettivamente ciò che l'Autore si proponeva di misurare. Si distinguono diversi tipi di validità: per contenuto, per criterio, per costrutto, concorrente, predittiva, discriminante, fattoriale, nominale (Tab. 3.III).

Validità per contenuto

La *validità per contenuto* si riferisce al fatto che gli item esplorino completamente l'area che si intende indagare. Data la grande variabilità del vissuto umano, normale e patologico, non sarebbe possibile, in pratica, esplorare tutti gli aspetti di un determinato problema per ciascun paziente; per tale motivo una RS è considerata valida per contenuto quando i suoi item coprono un'area sufficientemente rappresentativa del settore psicopatologico che si intende misurare. La validità per contenuto dovrebbe essere valutata sulla base dei criteri adottati dall'Autore della scala per la scelta degli item e del loro contenuto al fine di esplorare compiutamente il problema specifico. In alcuni casi queste informazioni sono riportate nel manuale della scala o negli articoli in cui lo strumento viene presentato, il più delle volte, però, queste informazioni non sono disponibili e dobbiamo perciò ripiegare su di una valutazione empirica, soggettiva, basata sull'esame del contenuto della scala. Quando si tratta di giudicare la validità per contenuto di strumenti che misurano, ad esempio, capacità specifiche, come l'acquisizione di determinate nozioni (matematica, fisica, eccetera), il giudizio può essere relativamente facile, ma quando si tratta di RS che valutano sindromi cliniche, il giudizio è più difficile e complesso poiché gli item che la compongono possono misurare tanto un sintomo quanto una sindrome o sintomi che sono abbastanza caratteristici quando sono presenti, ma la cui presenza non è particolarmente frequente. È necessario perciò che gli item che compongono la scala esplorino sintomi che si manifestano con buona frequenza nel quadro clinico esplorato per poter dire che essi esprimono pienamente la

sindrome che si intende esplorare e che, pertanto, la scala è valida.

Il problema, in psichiatria, è che, non esistendo un accordo unanime sulle diverse entità cliniche, è possibile descrivere la stessa sindrome usando termini e contenuti diversi e dando un peso relativo differente ai singoli aspetti. Tipico è, ad esempio, il caso della depressione per la valutazione della quale esistono numerose scale i cui item esplorano, come Faravelli (1983) ha ben documentato, aspetti sindromici diversi. Nasce allora il problema di che cosa realmente esplorino le RS, se la sindrome di per sé o la concezione che di quella sindrome hanno gli Autori delle diverse scale. Il problema non ha, ovviamente, soluzioni; generalmente si fa riferimento ad una RS assunta come standard e si calcola il coefficiente di correlazione fra questa e le altre scale impiegate nella stessa popolazione. Il grado di correlazione è tanto maggiore quanto più il coefficiente si avvicina all'unità e questo significa che l'ambito esplorato è abbastanza simile.

Dobbiamo anche ricordare che, quando l'entità che vogliamo valutare non è completamente definibile in base ad elementi osservabili e verificabili, come nel caso dei concetti psicodinamici, è praticamente impossibile avere una validità per contenuto, in quanto tali concetti sono esprimibili solo attraverso comportamenti che l'Autore presume collegati al concetto psicodinamico che egli vuole misurare; in questi casi si dovrà fare riferimento alla validità per costrutto (Faravelli, 1983).

Validità per criterio (o empirica)

La *validità per criterio*, detta anche validità empirica, si riferisce alla correlazione esistente tra lo strumento in esame ed altre variabili già considerate come misure affidabili. Questo tipo di approccio è più sofisticato rispetto alla validità per contenuto poiché presuppone la valutazione comparativa dello strumento con altre misure riconosciute come efficaci e con le quali ci si aspetta vi sia una correlazione. Il riscontro di questo tipo di validità consente di affermare che lo strumento in esame misura realmente ciò che si propone di misurare, anche se, come abbiamo detto, in psichiatria sono tutt'altro che comuni misure di confronto pienamente affidabili. Della validità per criterio fanno parte la *validità concorrente*, la *validità predittiva* e la *validità di gruppo*.

Validità concorrente

La *validità concorrente* indica la relazione che esiste fra i risultati forniti dalla RS ed uno o più criteri esterni (quali la diagnosi, il giudizio clinico, altre RS, eccetera), generalmente accettati come misure adeguate del fenomeno in esame, applicati contemporaneamente allo stesso paziente: così, ad esempio, un paziente con punteggio elevato alla *Brief Psychiatric Rating Scale* - BPRS, una scala di valutazione della psicopatologia generale, non dovrebbe avere un giudizio clinico di gravità "lieve" o "moderata". Generalmente il criterio esterno viene scelto ipotizzando che sia correlato con la RS.

Validità predittiva

La *validità predittiva* indica invece la capacità della scala di predire l'evoluzione del fenomeno che intende misurare. Poche sono le RS create per avere una validità predittiva e tra queste ricordiamo alcune scale prognostiche proposte per prevedere l'esito della schizofrenia. Il criterio della validità predittiva è molto importante per gli strumenti che mirano a individuare le specifiche

capacità e competenze del soggetto ai fini della scelta di un corso di studio di una carriera professionale piuttosto che di un'altra. In ambito clinico è difficile andare al di là di semplici indicazioni quali, ad esempio, un punteggio totale che indichi la necessità di instaurare un trattamento e, in alcuni casi, che il trattamento avrà probabilmente successo. Così, ad esempio, per molte RS è indicato il cosiddetto "cutoff point" (o score) che segna il limite fra la normalità e la patologia e che può essere usato come uno dei criteri per decidere se iniziare un trattamento. In altri casi, invece, il punteggio può essere usato come indice predittivo: nel caso di soggetti con disturbo d'ansia sottoposti a trattamento e valutati con una scala che esplora tanto l'ansia-tratto che l'ansia-stato, è prevedibile che, a parità di punteggio nell'ansia-stato, i pazienti con punteggi più elevati nell'ansia-tratto, vadano più facilmente incontro a ricadute. Il concetto su cui si basano tanto la validità concorrente che quella predittiva è analogo, essendo sempre un criterio esterno quello che definisce la bontà della scala, con la differenza che nella prima il criterio è contemporaneo, nella seconda è posposto nel tempo.

Validità per costrutto

La *validità per costrutto* si applica allorché l'entità da valutare non è definibile interamente in base a concetti verificabili empiricamente e ci si deve riferire perciò ad un modello teorico del concetto esplorato. È un tipo di procedimento che trova applicazione soprattutto nelle scale che valutano concetti psicodinamici, che non possono esaurirsi in una serie definita di comportamenti verificabili come, ad esempio, i meccanismi di difesa, l'oralità, la solitudine, eccetera. La validità per costrutto deriva dalla verifica della validità concorrente (se cioè il nostro strumento correla significativamente con altre variabili esterne di rilievo) e di quella discriminante (se cioè variabili esterne irrilevanti e teoricamente senza alcun rapporto con il concetto che vogliamo misurare non correlano significativamente con il nostro strumento). La validità della scala sarà tanto maggiore quanto più saranno soddisfatte le ipotesi generate dal concetto di partenza e quante meno saranno le ipotesi non verificate dai fatti (Faravelli, 1983).

Validità fattoriale

La validità concorrente e discriminante possono essere verificate anche mediante l'analisi fattoriale (*validità fattoriale*): le variabili simili, infatti, correleranno tra loro in un unico fattore e quelle dissimili andranno a costituire fattori diversi. L'impiego di questa tecnica richiede che il campione analizzato sia abbastanza ampio, che vengano scelte, fra le numerose disponibili, le procedure di calcolo più adatte e che i valori impiegati per stabilire l'accettabilità o meno della fattorializzazione (i cosiddetti "eigenvalues") siano sufficientemente rigorosi da garantire che i fattori siano significativi. Un'altra tecnica statistica che può essere utilmente impiegata è l'analisi della regressione mediante la quale si può verificare il grado di correlazione di ciascun item con il punteggio totale della scala e con altre variabili che teoricamente non hanno con loro alcun rapporto: questa procedura ci dice se il nostro strumento correla con variabili attinenti e diverge da quelle non attinenti.

Validità nominale

La *validità nominale* (o *face validity*) si riferisce alla reale corrispondenza fra gli item della scala e gli aspetti del quadro clinico che si intende valutare. Gli item della scala, in altri termini, devono esplorare effettivamente quelle manifestazioni che contribuiscono a definire i diversi aspetti dell'entità clinica che è oggetto del nostro interesse.

SENSIBILITA'

Il termine "*sensibilità*" (*sensitivity*) può essere inteso, in psicometria, sia come la capacità di uno strumento di distinguere i pazienti dai non pazienti, sia come la capacità di cogliere variazioni quantitative della sintomatologia esplorata.

Nella prima accezione, quindi, la sensibilità di una scala è la sua capacità di discriminare tra fenomeni diversi ed ha *due significati*: uno, di *derivazione epidemiologica*, che indica la percentuale dei casi reali di una malattia individuati, ed uno, che può essere più correttamente definito come "*specificità*", che indica la percentuale di persone normali che risulta negativa alla scala. Una terza componente della sensibilità è il tasso di errore (o *misclassification rate*), cioè la somma dei falsi negativi e dei falsi positivi, cioè la percentuale di soggetti normali classificati come pazienti più quella dei pazienti classificati come normali. La valutazione della sensibilità e della specificità di una scala è una procedura lunga e complessa in rapporto alle caratteristiche dello strumento, alla prevalenza della malattia o del disturbo che si propone di studiare ed alla metodologia di studio impiegata. È certamente più semplice discriminare i casi estremi piuttosto che quelli più lievi, che sono di più comune riscontro nella popolazione generale. Una scala sarà tanto più sensibile quanto più basso sarà il *misclassification rate*.

La seconda accezione del termine è di più recente introduzione ed è in rapporto alla valutazione dell'efficacia dei trattamenti e, in particolare, alla capacità di discriminare fra gli effetti di un farmaco e quelli di un placebo. La definizione generalmente accettata è: *la sensibilità è la caratteristica di una scala di discriminare a livello significativo fra gli effetti di uno psicofarmaco e di un placebo (o fra gli effetti di due farmaci o di due o più livelli di dose dello stesso farmaco) in una ricerca psicofarmacologica clinica in doppia cecità*.

Si può dire che una scala è tanto più sensibile quanto più piccole sono le differenze che riesce a discriminare. La sensibilità di una scala varia anche in rapporto a fattori esterni e, tra questi, importante è il *range* dei punteggi possibili per ciascun item: in linea di massima si può dire che l'ampiezza del range è correlata alla sua sensibilità. Una scala SI/NO è meno sensibile di una a 5 punti e questa, a sua volta, è meno sensibile di una a 10 punti. Non sempre è necessaria una sensibilità elevata: nelle scale di valutazione diagnostica, ad esempio, è sufficiente la presenza/assenza del sintomo; quando si debbano cogliere, invece, variazioni anche modeste del fenomeno che vogliamo misurare, come nel caso della valutazione dell'effetto terapeutico dei farmaci, è necessaria una più elevata sensibilità. Un giudizio sulla sensibilità può derivare dall'esame della distribuzione dei punteggi: se il campione è sufficientemente ampio, si osserva, di solito, che una scala poco sensibile tende ad utilizzare prevalentemente alcuni valori (di solito quelli centrali) mentre una sensibile utilizza più omogeneamente tutti i valori. È da tener presente, però, che un'eccessiva sensibilità tende a ridurre la riproducibilità della scala ed a limitarne così le possibilità di impiego.

In psichiatria è necessaria molta cautela nella scelta delle scale da usare negli studi di psicofarmacologia clinica poiché scale altamente correlate tra loro possono avere sensibilità molto diversa; se, ad esempio, una scala misura un "tratto" ed una il corrispondente "stato", la correlazione fra le due sarà molto elevata, ma la sensibilità al cambiamento sarà quasi nulla nella prima ed elevata nella seconda. Per essere sensibile una scala deve essere anche affidabile: se non è affidabile, se contiene, cioè, molti errori ed è soggetta a fluttuazioni casuali, non potrà essere considerata sensibile perché non sarà possibile stabilire se le variazioni eventualmente registrate sono imputabili al caso o all'effetto del trattamento. Il grado di sensibilità di una RS ha un'importanza che è inversamente proporzionale all'ampiezza del campione in esame; in una ricerca multicentrica mirata al confronto tra un farmaco attivo ed il placebo e che coinvolga

numerosi pazienti, non è necessaria una scala particolarmente sensibile per discriminare significativamente tra i due trattamenti. Se invece il campione è piccolo, o se il confronto è tra farmaci attivi, o tra livelli diversi di dose dello stesso farmaco, è necessario ricorrere a scale più sensibili per poter cogliere le eventuali differenze.

AFFIDABILITA'

Se un test che misura il livello di intelligenza fornisce, a distanza di pochi giorni (o se somministrato da valutatori diversi) ed a parità di condizioni, valori molto diversi, è evidente che i risultati di quel test saranno poco credibili ed il test stesso sarà poco affidabile. L'*affidabilità* (*reliability*) (Tab. 3.IV) è la possibilità che una scala, usata da osservatori diversi, in contesti o in tempi diversi, consenta di ripetere la misurazione dello stesso fenomeno con risultati sostanzialmente non dissimili. Per valutare l'affidabilità di uno strumento è necessario stabilire se i singoli item che esplorano una dimensione sono coerenti tra di loro (*consistenza interna*), se i punteggi sono stabili nel tempo (*test-retest reliability*), se osservatori diversi che valutano indipendentemente la stessa serie di pazienti danno punteggi sostanzialmente simili (*inter-rater reliability*) e se due forme diverse dello stesso strumento danno risultati sovrapponibili (*parallel form reliability*). Si può dire che l'affidabilità è il livello più alto di verifica della validità: se una RS non è in grado di fornire risultati riproducibili in tempi diversi, fra valutatori diversi o fra i suoi stessi item, i risultati con essa ottenuti sono inconsistenti e non è possibile definire che cosa misura quella scala.

L'affidabilità può essere valutata a vari livelli, a livello di sintomo, di sindrome o di diagnosi (Tab. 3.IV). Il livello di analisi a cui deve essere valutata l'affidabilità è in funzione degli obiettivi della ricerca; se, per esempio, siamo interessati più all'inquadramento diagnostico che alle modificazioni dei sintomi nel tempo, un'affidabilità a livello di diagnosi è senz'altro più importante di quella a livello di sintomi.

TAB. 3.IV - CARATTERISTICHE DELLE RS: AFFIDABILITÀ	
Consistenza interna	Coerenza tra gli item che esplorano una stessa dimensione
Test-retest reliability	Stabilità dei punteggi in somministrazioni successive dello strumento
Inter-rater reliability	Concordanza fra più valutatori che valutano lo stesso gruppo di soggetti
Intra-rater reliability	Concordanza fra valutazioni successive effettuate dallo stesso valutatore sullo stesso soggetto (effettuate generalmente su videotape)
Parallel forms reliability	Concordanza fra due versioni diverse dello stesso strumento
Error standard	Stima dell'entità dell'errore legato allo strumento

Consistenza interna

Per misurare efficacemente un quadro psicopatologico è necessario che gli item ne esplorino i diversi aspetti e siano quindi coerenti tra loro; se questa coerenza manca, è probabile che essi misurino cose diverse e non contribuiscano perciò alla misurazione del fenomeno in studio. Questa coerenza tra gli item è la *consistenza interna*, che può essere valutata mediante analisi statistiche che misurano il grado di correlazione tra le variabili; fra quelle più largamente impiegate ricordiamo il *coefficiente alfa di Cronbach* (un valore di alfa superiore a .80 indica generalmente una buona consistenza interna), il *test di Kuder-Richarson*, particolarmente indicato per gli

strumenti con item dicotomi, e la *split-half reliability*, generalmente usata per le scale di autovalutazione e che consiste nel calcolare la correlazione fra una metà degli item con l'altra metà, scegliendo variamente come suddividere gli item in due gruppi (la prima metà contro la seconda, quelli pari contro quelli dispari, creando due gruppi mediante assegnazione randomizzata, eccetera). Un'elevata consistenza interna indica che gli item esplorano le diverse facce di uno stesso fenomeno o che sono modalità diverse di descriverlo.

Test-retest reliability

La *test-retest reliability* valuta la stabilità dei punteggi in somministrazioni successive dello strumento; se la sintomatologia esplorata non è cambiata nell'intervallo fra le due valutazioni, il punteggio della scala dovrebbe essere sostanzialmente sovrapponibile e quindi la correlazione fra le due valutazioni dovrebbe essere elevata. La stabilità al test-retest è particolarmente importante per gli strumenti che valutano le caratteristiche o i tratti di personalità, per i quali bassi coefficienti di correlazione possono indicare che la scala è instabile o per la presenza di item che esplorano stati o per difficoltà di comprensione degli item. Per le patologie acute la test-retest reliability è inappropriata. La stabilità di una scala è importante quando questa venga impiegata più volte, a varia distanza di tempo, come nel caso della ricerca psicofarmacologica clinica, poiché solo una buona affidabilità al test-retest può consentire l'attribuzione di eventuali variazioni di punteggio all'azione del trattamento e non all'instabilità dello strumento. Il grado di correlazione al test-retest varia in funzione della distanza di tempo intercorrente fra le successive valutazioni; per intervalli di tempo limitati (1-2 settimane) il coefficiente dovrebbe essere superiore a .80, mentre sono accettabili anche valori di .69 o più per tempi intorno al mese. Data la variabilità nel tempo della sintomatologia psichiatrica, per la valutazione della test-retest reliability si fa spesso ricorso alla registrazione di interviste (con pazienti reali o simulati) su videocassette che vengono poi riproposte a tempi diversi.

Inter-rater reliability

Particolarmente importante è l'*inter-rater reliability*, cioè il grado di correlazione fra i punteggi assegnati allo stesso campione di pazienti da parte di due o più valutatori indipendenti che usano la stessa RS. Generalmente la valutazione viene fatta indipendentemente dai due (o più) valutatori presenti alla stessa intervista, anche se sarebbe metodologicamente più corretto che i valutatori intervistassero separatamente il paziente.

La stima dell'affidabilità di una scala richiede che il campione esaminato possieda una discreta variabilità in modo che si possa evidenziare la capacità dello strumento di discriminare i soggetti tra di loro: la buona affidabilità, infatti, presuppone che i valutatori siano concordi non solo nella valutazione ma anche nella discriminazione. L'*inter-rater reliability*, perciò, non si esprime soltanto in termini di accordo percentuale tra i valutatori, ma in termini di correlazione fra un insieme di fenomeni. Si deve tener conto, infatti, che una parte della concordanza è dovuta al caso: in una scala a 4 livelli (da 1 a 4), per esempio, i valutatori hanno il 25% di probabilità di concordare solo in funzione della casualità. Uno dei metodi più usati è il coefficiente *r di Pearson*, il quale però non corregge per gli errori sistematici (p. es., la costante sottostima di un valutatore rispetto ad un altro). Più affidabile è il *K di Cohen* e, soprattutto, il *weighted K (wk) di Hall*, che è stato sviluppato appositamente per la valutazione dell'accordo fra valutatori diversi di una RS.

Molto spesso l'*inter-rater reliability* e l'*accordo tra valutatori* sono considerati due concetti equivalenti ed intercambiabili: si tratta di una convinzione del tutto erronea poiché l'essenza della reliability è la condivisione, è il condividere la discriminazione fra soggetti, mentre l'accordo tra

valutatori riguarda il singolo soggetto e, come è ovvio, con un solo soggetto non può esserci discriminazione.

Nel caso, del tutto improbabile, che due valutatori concordino al 100% che l'entità del miglioramento in risposta ad un certo trattamento, in un campione di pazienti, è esattamente sovrapponibile, ci troveremmo di fronte ad un perfetto accordo tra valutatori ma la reliability sarebbe indeterminata.

Intra-rater reliability

Le capacità del clinico di valutare i pazienti possono cambiare (migliorare!) mano a mano che il valutatore acquisisce esperienza e pertanto la correlazione fra valutazioni effettuate a distanza di tempo può ridursi. Per valutare questo tipo di affidabilità (*intra-rater reliability*), essendo impossibile avere a disposizione a distanza di tempo lo stesso paziente con la stessa sintomatologia, si ricorre alla valutazione di casi registrati su videotape e si calcola quindi il grado di correlazione tra i punteggi ottenuti nelle diverse sessioni.

Parallel forms reliability

La *parallel forms reliability* viene utilizzata quando esistono due versioni diverse di una stessa scala (p. es., una versione estesa ed una breve) e si deve valutare se entrambe misurano la stessa entità; in questo caso è richiesto un elevato coefficiente di correlazione (in genere superiore a .80) perché le due versioni possano essere considerate corrispondenti.

Errore standard

I diversi approcci all'affidabilità che abbiamo descritto, servono a documentare l'assenza di errori nella valutazione mediante RS; c'è però un modo diretto per calcolare l'entità dell'errore legato allo strumento ed è l'*errore standard* (o standard error - SE) che è, in sostanza, una stima della deviazione standard dell'errore. Questo indice consente di valutare in che misura le variazioni di punteggio sono dovute all'errore. Così, se lo SE di una scala è pari a 3, solo variazioni del punteggio, tra una valutazione e l'altra, superiori a 3 potranno essere considerate cambiamenti reali. In linea generale, una scala è tanto più affidabile quanto minore è lo SE. Poiché il valore dello SE varia in funzione del range di punteggio della scala, scale con range diverso non sono confrontabili direttamente; il confronto è tuttavia possibile trasformando il valore dello SE in percentuale.

In realtà l'affidabilità è una caratteristica non tanto di una RS di per sé quanto della scala applicata ad uno specifico campione, da un particolare valutatore ed in un determinato contesto. Così, ad esempio, una RS risulterà "non affidabile" se è usata in un campione che per una qualche ragione non mostra alcuna variabilità nelle dimensioni esplorate da quella scala, e potrà ancora risultare "non affidabile" se è applicata da un valutatore inesperto piuttosto che da uno ben addestrato. In teoria, quindi, sarebbe necessario che ogni ricercatore procedesse alla verifica dell'affidabilità dello strumento che intende usare in condizioni il più possibile simili a quelle della ricerca che intende effettuare (campione, valutatori, setting, eccetera). Nella pratica la maggior parte dei ricercatori accetta i dati relativi agli studi di affidabilità riportati in letteratura, studi di solito effettuati dall'Autore dello strumento. È bene ricordare, in questo caso, che coefficienti di affidabilità inferiori a .70 rendono molto aleatoria l'affidabilità dello strumento in condizioni

diverse da quelle originali poiché non forniscono sufficienti garanzie contro gli errori di tipo 1 (rifiutare l'ipotesi che non esistano differenze quando le differenze non esistono) e di tipo 2 (accettare l'ipotesi che non esistano differenze quando invece le differenze esistono).

BIBLIOGRAFIA

Faravelli C: *Le scale di valutazione in psichiatria*. Quaderni Italiani di Psichiatria, 2:97, 1983.

Poli E, Faravelli C, Marchetti G: *La valutazione dei sintomi psichiatrici*. Rassegna di Studi Psichiatrici, 70:10, 1981.

CARATTERISTICHE DELLE RS

I più importanti fra i numerosi elementi che caratterizzano le RS sono:

- l'area esplorata
- il numero degli item e le modalità di attribuzione del punteggio
- il tempo da prendere in considerazione nella valutazione
- chi effettua la valutazione
- il grado di complessità della valutazione

AREA ESPLORATA

Come già abbiamo avuto occasione di dire, gli obiettivi della valutazione standardizzata, nell'ambito della clinica, sono essenzialmente quelli di un più preciso inquadramento diagnostico, di una chiara definizione del quadro psicopatologico, di una misurazione della gravità dei sintomi, di un monitoraggio degli effetti dei trattamenti. A fianco di questi, si collocano, poi, strumenti atti a valutare le risorse del paziente (intelligenza, personalità, deterioramento, eccetera), a studiare la prevalenza e l'incidenza nella popolazione della patologia psichiatrica in genere o di specifici disturbi in particolare, a valutare l'adattamento sociale ed altro ancora.

Per le diverse finalità della valutazione è necessario ricorrere a "strumenti" con caratteristiche diverse; è evidente che per una valutazione diagnostica può essere sufficiente stabilire la presenza o l'assenza di un sintomo, mentre per misurare la gravità del quadro clinico e del suo modificarsi nel corso del trattamento sarà necessario che, per i vari sintomi che lo caratterizzano, siano chiaramente specificati i diversi livelli di gravità. Analogamente, saranno necessari strumenti con caratteristiche diverse se vogliamo studiare i rapporti della sintomatologia specifica di un determinato quadro clinico con la psicopatologia nella sua globalità (come nel caso, ad esempio, dello studio della patologia di spettro o della comorbidità), oppure se vogliamo indagare aspetti limitati di uno specifico quadro clinico (come potrebbe essere, ad esempio, l'approfondimento di sintomi quali la perdita di speranza, i sentimenti di inutilità e di impotenza, considerati come elementi essenziali per la comprensione della depressione).

Nel caso della valutazione dell'efficacia di un trattamento, è necessario che siano prese in considerazione quelle dimensioni che sono capaci di rilevare i cambiamenti. La mancanza di queste dimensioni (o di un loro numero sufficiente) può far giungere alla conclusione di una mancanza di efficacia terapeutica, quando invece è presente. Bisogna anche evitare l'eccesso di variabili, eliminando quelle per le quali non ci sono indizi che siano influenzabili dal trattamento in studio, poiché la loro ridondanza può rendere difficile l'interpretazione dei risultati della ricerca. In alcuni casi, queste variabili aspecifiche possono addirittura raggiungere la significatività statistica in seguito alle modificazioni di alcuni parametri solo per effetto della casualità.

Così, ad esempio, nella valutazione degli effetti del trattamento del paziente schizofrenico acuto, si dovranno valutare le modificazioni del giudizio di realtà, dei disturbi del pensiero e delle sensopercezioni, della disorganizzazione del comportamento, del ritiro sociale e dei comportamenti auto ed eteroaggressivi, mentre nel trattamento di quello cronico l'attenzione sarà più spostata sul versante dell'adattamento sociale. Anche nei quadri depressivi si terrà conto del giudizio di realtà e dell'adattamento sociale, ma l'attenzione sarà maggiormente focalizzata sui disturbi dell'umore,

sulla sofferenza soggettiva, sui sintomi somatici, sul rischio suicidario.

Vedremo questi aspetti più in dettaglio nei singoli capitoli; vogliamo sottolineare qui che non per tutte le aree il consenso è unanime. Esso è certamente molto più ampio quando si tratta di esplorare i disturbi psicopatologici che fanno ormai parte dell'esame psichico tradizionale, come i disturbi dell'umore, della psicomotricità, dell'orientamento, della memoria, del pensiero, del giudizio di realtà, eccetera, poiché è generalmente accettato il loro valore come indici di gravità del disturbo e sono concetti ormai ben definiti e valutabili con buona affidabilità. Meno unanime è il consenso su altri concetti, come, per esempio, quello di "adattamento sociale", di "qualità della vita", eccetera, poiché i valori su cui si basano concetti quali "buon funzionamento", "soddisfazione", eccetera sono estremamente variabili non soltanto a livello soggettivo, ma anche a livello socioculturale. Basti pensare al diverso significato che questi due termini, "buon funzionamento" e "soddisfazione", possono avere se riferiti a condizioni (vita coniugale, lavoro, tempo libero, attività sociali, eccetera) ed a persone diverse (un giovane ed un anziano, un laureato ed uno appena alfabetizzato, di una classe sociale abbiente o non abbiente, un uomo ed una donna e così via). Ma anche a parità di altre condizioni, il giudizio sull'adattamento sociale è complesso, derivando dal giudizio sul livello di attività prestazionali e di soddisfazione giudicato accettabile dalla famiglia e dalla società e quello attuale del paziente. La valutazione dell'adattamento sociale è anche in rapporto al "setting": ben diverso è il giudizio in ambito di ricovero, in una comunità per lungodegenti o nel normale ambiente di vita. Inoltre modalità diverse di trattamento possono influenzare in maniera diversa, e talora senza apparenti correlazioni, il quadro psicopatologico e l'adattamento sociale.

Altri settori controversi sono quelli relativi alla valutazione della personalità (per la mancanza di consenso su quali siano le sue dimensioni fondamentali) e dei concetti psicoanalitici, perché esiste disaccordo sia su quali siano realmente importanti, sia sulla loro stessa definizione, sia, infine, sulla loro indipendenza rispetto ai sintomi psicopatologici.

Alcuni studiosi hanno criticato le RS poiché prendono in considerazione soltanto i disturbi psicopatologici ed ignorano gli aspetti positivi dello psichismo, quelli che potremmo indicare come segni di salute mentale. Se in passato è prevalsa l'opinione che gli aspetti positivi dello psichismo potevano essere considerati l'immagine speculare di quelli negativi, oggi si tende a considerare i due aspetti come relativamente indipendenti. Alcuni strumenti di valutazione, come la *Global Assessment Scale* (Spitzer et al., 1973) o, ancor più, scale di valutazione dell'adattamento sociale, come la *Structured and Scaled Interview to Assess Maladjustment* (Gurland et al., 1972) e la *Social Adjustment Scale* (Weissman et al., 1971) hanno recepito questa istanza inserendo item che esprimono aspetti positivi, come la cooperatività, la contentezza, l'avere il senso dell'umor, l'essere responsabili, eccetera. Anche per gli aspetti positivi del funzionamento mentale manca comunque, così come per l'adattamento sociale o per le dimensioni della personalità, un vasto consenso tra i diversi Autori.

Sulla base dell'area esplorata si possono sostanzialmente distinguere le scale di valutazione diagnostica e quelle di valutazione psicopatologica e, fra queste, quelle di valutazione globale e quelle di valutazione settoriale; ci sono poi le scale di valutazione dell'adattamento sociale, del rischio di suicidio, della qualità della vita, eccetera. Questo criterio di classificazione sarà seguito nella parte speciale, dove saranno descritte una buona parte delle RS che meritano, per qualche ragione, di essere ricordate e saranno riportate quelle più importanti o di maggiore diffusione.

NUMERO DEGLI ITEM E MODALITA' DI ATTRIBUZIONE DEL PUNTEGGIO

Il numero ottimale degli item di una scala dipende da molteplici fattori quali il tipo di area esplorata, le finalità (valutazione diagnostica o quantitativa), il valutatore (auto o eterovalutazione), la frequenza con cui viene applicata allo stesso soggetto, il tipo di ricerca, ed altro ancora. In alcuni casi può essere più utile una scala più ampia, che esplora più diffusamente

l'area in studio, altre volte (e soprattutto nell'autovalutazione) è consigliabile l'impiego di scale con un numero limitato di item. Quando si esplora un'area psicopatologica specifica mediante l'autovalutazione, è bene non superare i 20 item sia per evitare che il paziente perda o riduca le sue capacità attentive e la sua concentrazione, sia perché può essere difficile trovare un numero adeguato di sostantivi e di aggettivi che possano essere compresi anche da una persona di modesto livello culturale (e per di più ammalata!).

Le modalità in base alle quali può essere attribuito il punteggio agli item delle varie RS sono diverse. Molte RS contengono item dicotomi (Sì-No; Vero-Falso; Presente-Assente); il vantaggio delle scale dicotome è rappresentato dalla rapidità di compilazione (l'MMPI, ad esempio, può essere compilato in un tempo relativamente breve pur essendo composto da oltre 500 item). Alcuni soggetti hanno, invece, notevole difficoltà con questo tipo di strumenti perché le dimensioni ed i tratti che essi devono dicotomizzare generalmente sono, come del resto la maggior parte delle dimensioni e dei tratti psico(pato)logici, di tipo continuo e difficilmente sono totalmente assenti. È evidente, quindi, che per decidere della presenza/assenza di una di tali dimensioni è necessario far ricorso ad una sorta di "cutoff soggettivo" che faccia da discriminare. Questo cutoff è variabile da individuo ad individuo e può variare in rapporto a molteplici fattori (personalità, situazione psico(pato)logica del momento, compresenza o meno di altri problemi bio-psico-sociali, eccetera); già basamente, quindi, ci può essere una certa variabilità che può raggiungere livelli estremi, da un lato nei soggetti con cutoff molto basso o assente (come può essere il caso, ad es., degli ossessivi) e, dall'altro, nei soggetti con cutoff molto elevato (come, ad es., negli ipertimici o, ancor più, ipomaniaci). Tutto questo può risultare penalizzante per la categoria "presente" che include diversi gradi di gravità che non vengono valutati, con conseguente perdita di informazione.

Le scale dicotome sono più indicate in tutti quei casi, come nelle valutazioni diagnostiche in cui ci interessa più la presenza o l'assenza dei sintomi piuttosto che la loro gravità. Nella maggior parte dei casi, ed in particolare quando si voglia misurare il livello di psicopatologia ed i suoi cambiamenti nel tempo, si preferiscono scale i cui item prevedono un maggiore o minore numero di livelli di gravità. La quantificazione di quest'ultima è affidata ad aggettivi o ad avverbi ordinati secondo una scala crescente di gravità (ad esempio: assente, lieve, moderato, grave, gravissimo; oppure: mai, quasi mai, qualche volta, talora, spesso, quasi sempre, sempre; o ancora: come al solito, un po' meno, meno, molto meno, per niente, eccetera). Si parla in questo caso di *scale semidefinite*. Nonostante l'ambiguità insita in questi livelli di graduazione della gravità, i clinici, purché usino le stesse definizioni dei concetti ed abbiano un'esperienza abbastanza simile, sono generalmente in grado di fornire un giudizio con un buon livello di accordo tra di loro. Se però l'esperienza dei valutatori è diversa, è necessario, per raggiungere un elevato livello di standardizzazione e di affidabilità, ricorrere ad un *training* in modo da fornire dei riferimenti comuni.

Per migliorare la validità e l'affidabilità, senza dover ricorrere a speciali training, sono state costruite *scale altamente formalizzate* nelle quali i comportamenti corrispondenti ai singoli punteggi di ciascun item sono accuratamente e dettagliatamente descritti. Esempi tipici di scale altamente formalizzate sono la HAM-D (1960) o la *Comprehensive Psychopathological Rating Scale - CPRS* (1978) (Tab. 4.I) nelle quali ad ogni punteggio è associata una precisa descrizione.

Queste scale, riducendo al minimo la variabilità tra osservatori diversi, consentono di raggiungere la massima standardizzazione della valutazione ed un'altissima affidabilità, anche se a prezzo di una ridotta flessibilità, costringendo il valutatore a forzature quando il comportamento osservato non sia esplicitamente descritto nelle istruzioni, cosa peraltro non eccezionale dato che le descrizioni riportate non possono coprire tutte le possibilità espressive di un determinato sintomo.

□

TAB. 4.I - ESEMPI DI RILATIVAMENTE FORMALIZZATE	
<i>HAMILTON DEPRESSION RATING SCALE (HAM-D)</i>	
Item n° 1 - <i>Umore depresso</i> (sentimento di tristezza, mancanza di speranza, sentimento di incapacità e di inutilità)	0 = Assente
	1 = Manifesta questi sentimenti solo se interrogato
	2 = Esprime spontaneamente questi sentimenti
	3 = Comunica questi sentimenti con messaggi non verbali, cioè attraverso l'espressione del volto, la posizione del corpo, la voce e la tendenza al pianto
	4 = Il paziente manifesta quasi esclusivamente questi sentimenti mediante messaggi sia verbali che non verbali
<i>COMPREHENSIVE PSYCHOPATHOLOGICAL RATING SCALE (CPRS)</i>	
Item n° 1 - <i>Tristezza</i> Esperienza soggettiva di umore triste, indipendentemente dal fatto che tale stato d'animo sia evidente o meno all'esterno. Comprende l'umore depresso, l'essere giù di morale, scoraggiati, il sentirsi senza possibilità di essere aiutati e senza speranza; valutata secondo l'intensità, la durata e il grado in cui l'umore è influenzato dalle circostanze. L'umore euforico è valutato zero in questo item	0 = Umore abitualmente stabile. Occasionalmente triste in rapporto alle circostanze
	1 = Predomina la tristezza, ma possono verificarsi momenti di gaiezza
	2 = Tristezza o melanconia persistente. Umore è influenzato molto difficilmente da eventi esterni
	3 = Esperienza continua di infelicità o di estremo scontento

Sul versante opposto rispetto alle scale altamente formalizzate, si pongono le *scale analogiche*, rappresentate da un segmento lungo 10 cm, alle cui estremità stanno a sinistra, l'assenza del sintomo ed a destra la sua massima gravità (Tab. 4.II); il valutatore è invitato a collocare il paziente (o il paziente è invitato a collocare se stesso) fra questi due estremi, nella posizione che, a suo giudizio, meglio esprime la gravità del sintomo; il punteggio è espresso dai millimetri che intercorrono fra lo zero (assenza del sintomo, all'estremità sinistra) ed il punto in cui è stato collocato il paziente. Il livello di riproducibilità di queste scale è limitato, essendo elevata la soggettività del valutatore, mentre sono di buon livello la flessibilità e la sensibilità. Una soluzione che potrebbe unire i vantaggi delle scale analogiche con quelli delle scale formalizzate è quella delle scale analogiche "con punti di ancoraggio" (*anchor point*), scale analogiche, cioè, sulle quali sono riportate, a determinati intervalli, descrizioni esemplificative dei comportamenti corrispondenti, ai quali il valutatore può fare riferimento nell'assegnazione dei punteggi (Tab. 4.III).

TEMPO DELLA VALUTAZIONE

La valutazione deve sempre coprire un arco di tempo definito, appropriato a ciò che ci proponiamo di valutare; esso può variare dal momento della valutazione a tutta la vita del paziente, ma per essere confrontabile deve essere costante per lo stesso strumento e nell'ambito della stessa ricerca. La scelta dell'intervallo di tempo da prendere in considerazione dipende da ciò che vogliamo misurare e per quale scopo. I tratti di personalità si riferiscono, per definizione, a pattern comportamentali stabili o comunque di lunghissima durata; l'umore può essere valutato sia sul breve che sul medio termine; l'ansia, se è un "tratto", è una caratteristica stabile del soggetto, se è uno "stato", è una caratteristica di breve-medio periodo. Quando si valuta il quadro psicopatologico il riferimento è, generalmente, variabile da una settimana ad un mese in rapporto alla maggiore o minore rapidità con cui la specifica sintomatologia generalmente si modifica spontaneamente o in funzione del trattamento; se si effettua una valutazione diagnostica, i tempi



difficile comprensione generale per numerosi pazienti, possono prestarsi ad interpretazioni semantiche diverse, o possono essere usate come strumento di manipolazione da parte del paziente. Accanto, quindi, ad Autori come Zung, Biggs, Kellner, McNair, Heiman, Lipman, Covi, eccetera, che sostengono la validità e l'attendibilità delle SRS nella valutazione della sintomatologia e del suo cambiamento e nella discriminazione tra farmaco attivo e placebo nelle prove cliniche, ce ne sono molti altri (Carrol, Hamilton, Williams, Pichot, Rickels, Prusoff, Aitken, Angst, Faravelli, eccetera) che mettono più o meno decisamente in dubbio l'utilità delle SRS avendo riscontrato, con queste:

- una bassa concordanza con il giudizio clinico;
- la difficoltà per il malato di interpretare i diversi livelli di gravità o di capire il significato di alcune espressioni;
- l'impossibilità di cogliere aspetti obiettivi rilevanti per il clinico, come l'espressione mimica, il tono della voce, la psicomotricità, eccetera;
- la difficoltà che il paziente può avere nel rilevare cambiamenti modesti che, pur non incidendo sull'essenza del disturbo, sono importanti per il clinico che può ricavare da questi indicazioni circa l'opportunità di proseguire il trattamento, incrementarne le dosi o sospenderlo;
- la possibilità che il paziente amplifichi miglioramenti irrilevanti per gratificare il medico o per ottenere vantaggi quali il trasferimento in un reparto migliore, una minore sorveglianza o la dimissione (magari per mettere in atto gesti autolesivi);
- la possibilità, al contrario, che il paziente minimizzi i miglioramenti o amplifichi la sintomatologia per richiamare l'attenzione del medico, ottenere maggiore attenzione, o altro.

Le ORS si prestano ad essere utilizzate in varia maniera. Una prima modalità è quella dell'osservazione effettuata da un informatore (ed esistono per questo appositi questionari) il quale può fornire notizie sul comportamento del paziente nel suo ambiente naturale. Il campo di osservazione migliore è, in questo caso, il comportamento direttamente osservabile, mentre i sentimenti ed i pensieri possono più facilmente sfuggire o esser tenuti nascosti dal paziente. Non sempre, purtroppo, è disponibile un informatore e quelli possibili possono non prestarsi o essere incapaci di effettuare le rilevazioni.

Le RS per l'osservazione naturalistica da parte del personale di assistenza in ambito di ricovero, sono assai numerose e sono state fra le prime RS ad essere state sviluppate. L'attenzione di questi strumenti è per lo più focalizzata sul comportamento manifesto e sono perciò più adatti a quei tipi di patologia in cui il comportamento è significativo e di facile rilevazione. Sono particolarmente indicati per valutare la frequenza dei comportamenti abnormi o disturbanti, soprattutto nei pazienti molto gravi, con i quali è difficile comunicare. L'uso di questi strumenti ha avuto la massima diffusione quando la psichiatria era svolta prevalentemente negli ospedali (psichiatrici e non) ed i ricoveri erano protratti; oggi, con lo spostamento dell'assistenza psichiatrica nella comunità e con la marcata riduzione della frequenza dei ricoveri e della loro durata, hanno perso molto del loro valore.

La modalità più comune, rimane comunque la valutazione diretta del paziente da parte del medico il quale, con un'intervista ben condotta, può ottenere informazioni non ottenibili per altre vie; combinando le informazioni così raccolte con quanto osservato nel corso dell'intervista e con altre informazioni eventualmente raccolte da altre fonti, egli può formulare giudizi molto più appropriati e pertinenti. Attraverso un buon colloquio psichiatrico è possibile migliorare la qualità e la quantità delle informazioni fornite dal paziente o, comunque, evidenziare la tendenza a mentire o a sotto o sovrastimare i sintomi e dare così un giudizio ponderato. Il limite maggiore

delle ORS deriva dal fatto che ogni clinico ha una propria modalità di effettuare l'intervista, di porre le domande, di esplorare più o meno ampiamente le diverse aree della psicopatologia. Lo stesso intervistatore, inoltre, spesso cambia il proprio modo di intervistare non solo i diversi pazienti, ma anche lo stesso paziente in momenti successivi. Ne deriva che, con le ORS, la variabilità fra valutatori può essere maggiore di quella fra i pazienti. Per ovviare a questo, sono state proposte delle interviste standardizzate, più o meno rigidamente strutturate, che migliorano nettamente l'accordo fra valutatori: parleremo più avanti di questo tipo di interviste.

È generalmente accettato il principio che le informazioni utilizzabili in ambito di ricerca siano soltanto quelle ottenute direttamente dal paziente, in modo che ogni valutatore disponga delle medesime informazioni e che, quindi, la concordanza fra valutatori sia maggiore. Questa maggiore affidabilità è ottenuta, però, a scapito della validità, che sarà certamente minore di quella raggiungibile se si potessero utilizzare anche altre fonti di informazione (familiari, personale di assistenza, eccetera) dato che, nel corso dell'intervista, possono non emergere alterazioni perché troppo sporadiche o per il fatto di essere rilevabili soltanto dall'osservatore, ma non dal soggetto, o perché il paziente tende a negare, o per altri motivi ancora; è il caso, ad esempio, del comportamento motorio, dei disturbi del pensiero e delle allucinazioni, dei comportamenti bizzarri o di abuso o aggressivi, eccetera, che non sono costanti nel tempo o che il paziente può tendere a dissimulare nel corso dell'intervista.

L'impiego di fonti di informazioni diverse dall'intervista con il paziente, crea comunque ulteriori problemi: il valutatore può non essere sicuro della qualità delle informazioni derivanti da altre fonti, o fonti diverse possono fornire informazioni differenti o addirittura contrastanti; i diversi valutatori possono avere quantità o tipi diversi di queste informazioni e tutto questo può influenzare anche pesantemente il risultato finale. Com'è facile intuire, non esistono soluzioni ottimali per risolvere questi problemi.

Prima di concludere, è necessario tornare un attimo sul problema delle SRS e delle ORS, di cui Faravelli (1983) ha fatto un'ampia disamina, della quale cercheremo di fare qui una sintesi. A ben vedere, quelli indicati come gli svantaggi delle ORS, il dispendio di tempo da parte del medico e l'introduzione di una fonte di varianza, sono in realtà irrilevanti. Nel contesto di un normale rapporto clinico, il colloquio con il paziente è (o dovrebbe essere) tale da fornire le informazioni necessarie per rispondere praticamente a tutti gli item di una comune RS; l'attenta osservazione, contestuale al colloquio, dovrebbe fornire la risposta agli item di tipo comportamentale; in genere, solo per pochi item può essere necessario porre domande specifiche. È evidente, quindi, che la compilazione di una ORS non richiede, di solito, che pochissimi minuti in più rispetto al tempo normale di consultazione. Quanto, poi, alla variabilità introdotta dall'osservazione esterna, è stato ampiamente dimostrato in numerose ricerche che l'accordo fra valutatori diversi è generalmente molto elevato, poiché il giudizio degli psichiatri, anche se di impostazione dottrina diversa, è, per l'omogeneità della cultura di base, per la formazione medica, per l'esperienza clinica e per l'abitudine all'osservazione, molto più coerente e confrontabile rispetto all'autovalutazione, che è molto più influenzabile da una serie di fattori personali, quali la capacità di introspezione, il livello culturale, le conoscenze linguistiche, eccetera. È stato dimostrato, infatti, che la correlazione fra l'auto e l'eterovalutazione:

- è direttamente proporzionale al livello culturale dei pazienti, probabilmente perché il contenuto semantico degli item è il prodotto dell'evoluzione culturale generale e quindi l'accordo è tanto maggiore quanto più lo sviluppo culturale del paziente è vicino a quello del medico;
- è maggiore per quegli item per la cui valutazione è necessario interrogare il paziente (p. es., i sentimenti di colpa) che per quelli per i quali la valutazione è più legata all'osservazione (p. es., l'agitazione psicomotoria);
- è maggiore nelle forme medio-lievi di psicopatologia che in quelle gravi.

È un dato di fatto che i pazienti tendono a dividersi in tre categorie: quelli che valutano criticamente le proprie sensazioni ed il proprio stato (raggiungendo, di solito, un buon accordo con la valutazione del clinico), quelli che amplificano al massimo la propria condizione e quelli che, al contrario, la minimizzano; questi ultimi due tipi di pazienti tendono a dare valutazioni nei valori estremi della scala, lasciando perciò poco margine per eventuali ulteriori peggioramenti o, rispettivamente, miglioramenti.

È bene, poi, ribadire il concetto che il paziente (ma non, ovviamente, il clinico) attribuisce alla propria valutazione un significato di comunicazione e che, pertanto, più che dare una valutazione oggettiva dei propri disturbi, egli tende ad esprimere attraverso la SRS una richiesta di aiuto e può essere perciò portato a rafforzare il messaggio mediante un'amplificazione funzionale.

Queste osservazioni ci portano al nocciolo del problema, quali siano, cioè, i criteri attraverso i quali lo psichiatra ed il paziente giungono alla valutazione. È evidente che il clinico, di fronte al malato, opera un confronto tra il quadro clinico attuale e quello di pazienti con patologia analoga di cui, sulla base della sua formazione professionale e della sua esperienza personale, si è creato in mente uno stereotipo. Il punteggio delle ORS ha perciò, di solito, una spiccata tendenza centrale, essendo appunto "centrale" lo stereotipo ed oscillando intorno a questo i punteggi dei pazienti reali. Il paziente, invece, ha come punto di riferimento solo se stesso ed il proprio vissuto attuale confrontato a quello precedente e perciò, quale che sia la gravità "oggettiva" del vissuto attuale, essa sarà comunque soggettivamente grave; è per questo che, con le SRS, la distribuzione dei punteggi tende verso l'estremo, che rappresenta la maggiore gravità.

Anche nella valutazione del cambiamento, che è poi uno degli scopi principali delle RS, il clinico ed il paziente adottano criteri diversi. Il valutatore esterno, che avrà certamente difficoltà a ricordare esattamente le condizioni precedenti del paziente, soprattutto se è trascorso un certo lasso di tempo, tenderà a fare, anche in questo caso, una valutazione trasversale, facendo riferimento soprattutto allo stereotipo che egli ha in testa, più che alle osservazioni precedenti dello stesso paziente. Il paziente, per i motivi sopra esposti, tenderà invece a fare il confronto con le sue condizioni precedenti e soprattutto con quelle della massima acuzie. Se questo è vero, il miglioramento espresso mediante le RS dovrebbe essere espresso in percentuale con le SRS e mediante i punteggi grezzi con le ORS. Con queste ultime il punteggio esprime il cambiamento assoluto della sintomatologia ed è perciò maggiormente confrontabile tra pazienti diversi. Il valutatore esterno, inoltre, è in grado di cogliere variazioni anche modeste della sintomatologia, mentre il paziente tende generalmente a valutare le sue condizioni nell'ottica più sfavorevole (del resto, quali disturbi sono più gravi di quelli in atto? Quelli passati, anche se più gravi di quelli attuali, non ci sono più!) ed ha bisogno perciò di variazioni sensibili per prendere coscienza dell'avvenuto cambiamento. Tutto questo sembra confermare l'opinione di molti Autori che auto ed eterovalutazione misurano qualcosa che è sostanzialmente diverso, tanto dal punto di vista concettuale che da quello semantico. Non si può negare, infatti, che il paziente tende a valorizzare la sofferenza soggettiva così come è psicologicamente vissuta, mentre gli aspetti più "esterni" della malattia hanno meno rilievo e minor peso; lo psichiatra, invece, che non può penetrare più di tanto nell'intimo della sofferenza del paziente, riesce a cogliere gli aspetti più esterni, ma non meno importanti, della realtà clinica, quali il comportamento, la compromissione somatica, le problematiche socio-relazionali, eccetera.

Anche gli studi sulle RS sembrano confortare questa opinione: l'accordo fra ORS e SRS, espresso dal coefficiente di correlazione, generalmente non è molto elevato, a testimonianza del fatto che i due tipi di scale non misurano, nella sostanza, la stessa cosa; per contro la correlazione tra le SRS tra loro e tra le ORS tra loro è abitualmente più elevata, a documentare che i due tipi di scale sono sostanzialmente diversi tra loro. O, per meglio dire, misurano aspetti diversi di uno stesso fenomeno, che è apparentemente non definibile nella sua totalità, assumendo aspetti diversi in rapporto all'ottica secondo cui viene osservato.

Per concludere, le ORS sono più vicine all'idea che il medico ha della malattia, sono strumenti maneggevoli, più aderenti delle SRS al concetto di strumento di misura ed al compito per cui sono state concepite, e cioè la valutazione obiettiva della psicopatologia, della sua gravità e del suo cambiamento, e perciò più adatte a soddisfare le necessità della ricerca. Esse però non sono in grado di descrivere gli aspetti soggettivi, comunicativi e relazionali, che sono parte integrante della malattia: questi aspetti sono meglio esplorati dalle SRS, che consentono una più approfondita conoscenza delle problematiche del paziente. In definitiva, quindi, possiamo considerare le scale di auto e di eterovalutazione strumenti da utilizzare in maniera integrata, per meglio rispondere all'esigenze tanto della clinica che della ricerca.

COMPLESSITA' DELLA VALUTAZIONE

La valutazione può richiedere livelli diversi di complessità; ad un livello più elementare, al valutatore può essere richiesto di registrare semplicemente la presenza o assenza di una ben definita "unità" di comportamento (p. es., "Il p. piange", "Il p. lamenta insonnia", eccetera), oppure di fornire dei giudizi più complessi ed articolati che richiedono di prendere in considerazione anche il contesto in cui il comportamento si manifesta (p. es., "Il p. è delirante ed i suoi deliri influenzano il suo comportamento", "Esprime la convinzione di essere affetto da una grave malattia, anche se ciò non è vero", eccetera). I giudizi semplici sono certamente più attendibili di quelli complessi e sarebbe perciò preferibile che gli item indagassero i singoli comportamenti piuttosto che richiedere inferenze da parte del valutatore. In ogni caso, indipendentemente dalla complessità dell'item, sarebbe opportuno che fosse sempre specificato chiaramente che cosa deve essere preso in considerazione per fornire il giudizio e che cosa deve essere escluso; se si deve valutare l'umore depresso è necessario specificare se si deve tener conto solo di quello che il paziente riferisce o solo degli aspetti comportamentali o di entrambi.

BIBLIOGRAFIA

Asberg M, Montgomery SA et al.: *A Comprehensive Psychopathological Rating Scale*. Acta Psychiat Scand, 271(s):5, 1978.

Faravelli C: *Le scale di valutazione in psichiatria*. Quaderni Italiani di Psichiatria, 2:97, 1983.

Gurland BJ, Yorkston N et al.: *The Structured and Scaled Interview to Assess Maladjustment (SSIAM): I. Description, rationale, and development*. Arch Gen Psychiatry, 27:259, 1972.

Hamilton M: *A rating scale for depression*. J Neurol Neurosurg Psychiat, 23:56, 1960.

Spitzer RL, Gibbon M, Endicott J: *Global Assessment Scale*. Biometric Research, New York State Department of Mental Hygiene, New York, 1973.

Weissman MM, Paykel ES et al.: *The social role performance of depressed women: Comparison with a normal group*. Am J Orthopsychiat, 41:390, 1971.

PRINCIPI DI VALUTAZIONE

L'INTERVISTA

Per un corretto impiego delle RS è indispensabile una loro buona conoscenza ed un adeguato addestramento al loro uso. È pur vero che, nella maggior parte dei casi, gli Autori delle RS forniscono le indicazioni per l'impiego delle loro scale nella pubblicazione in cui le presentano o, per quelle più complesse, nel manuale di istruzioni che le accompagna, ma questo non è, in genere, sufficiente e se ne dovrà tenere conto, poiché la mancanza di un adeguato addestramento può limitare la validità e la generalizzabilità dei dati ottenuti.

Il problema dell'addestramento è particolarmente importante nel caso delle ricerche multicentriche, per le quali è una necessità assoluta che i valutatori dei diversi centri usino gli strumenti in maniera uniforme, in modo da rendere pienamente comparabili i risultati ottenuti. Questo vale tanto per gli strumenti nuovi o di impiego eccezionale, quanto per le scale più note e di uso più comune, per il principiante così come per lo psichiatra esperto, per evitare che la valutazione effettuata sulla base della formazione e dell'esperienza personale riduca la inter-rater reliability e quindi la comparabilità dei dati. È perciò prassi comune che, prima dell'inizio della ricerca multicentrica, coloro che effettueranno le valutazioni si sottopongano a sedute in cui, di solito, sulla base di registrazioni video di interviste psichiatriche, vengono valutati diversi pazienti e vengono discusse le eventuali differenze rilevate fra le diverse valutazioni, in modo da migliorare la inter-rater reliability.

Un problema molto dibattuto è quello del modo di condurre l'intervista ai fini della valutazione. Molti preferiscono il colloquio libero purché condotto, ovviamente, da un investigatore esperto. Agli inizi, peraltro, l'intervista era qualcosa di non strutturato, lasciata all'abilità dell'investigatore, alla sua capacità di far emergere segni e sintomi e di registrarli; al massimo, venivano suggerite ai valutatori delle linee-guida da seguire nel corso della valutazione. È chiaro, come abbiamo già detto, che l'intervista clinica trova dei limiti nel fatto che esiste, tra i valutatori, un'elevata variabilità nella modalità di proporsi al paziente, nel modo di porre le domande e, addirittura, nell'esplorazione dell'argomento in questione. Oltre alle differenze fra valutatori, si deve tener conto anche del fatto che lo stesso valutatore tende spesso a proporsi in modo diverso ai vari pazienti ed allo stesso paziente in tempi diversi. Ne consegue che il range di variabilità fra valutatori può risultare molto elevato.

Per ridurre questa fonte di variabilità, alcuni Autori hanno sostenuto la necessità di interviste più o meno rigidamente strutturate tali, cioè, da aumentare l'affidabilità e la riproducibilità della valutazione. L'introduzione del concetto di intervista strutturata risale agli anni Settanta in relazione alle problematiche legate all'accuratezza diagnostica ed il primo tentativo di intervista strutturata è il *Present State Examination - PSE* (Wing et al., 1974).

Le interviste strutturate, in effetti, consentono di ridurre la variabilità legata alla diversa modalità di condurre l'intervista ed alla sua ampiezza; consentono di coniugare la flessibilità dell'intervista clinica con la completezza della raccolta delle informazioni e la confrontabilità tipiche dell'inchiesta sistematica. Si tratta, in sostanza, di strumenti di ricerca per i quali è necessaria una certa selezione dei valutatori ed un loro training.

L'impiego delle interviste strutturate è strettamente indicato negli studi multicentrici condotti da gruppi di ricercatori indipendenti, nelle ricerche transculturali, negli studi a lungo termine, nei quali i pazienti vengono reclutati nel corso degli anni, per garantire l'omogeneità dei criteri diagnostici. Meno vincolante è il loro impiego nelle ricerche condotte da un numero limitato di valutatori, su di una sola popolazione di soggetti e per tempi limitati.

L'intervista strutturata consiste in una serie di domande e di definizioni che l'intervistatore usa per ottenere le informazioni dal paziente; se l'intervistatore è ben preparato e l'intervista è ben condotta, può non apparire sostanzialmente diversa da un'intervista clinica libera. Di solito l'intervista è suddivisa in argomenti e le domande sono tali da consentire di ottenere tutte le informazioni necessarie per dare una risposta agli item relativi a quell'argomento. Le domande sono spesso formulate in modo tale da stimolare il soggetto a parlare del problema piuttosto che a rispondere semplicemente con un "sì" o un "no" (p. es., "Che cosa mi dice del suo appetito?" piuttosto che "Il suo appetito è ridotto?"); vengono anche suggerite domande supplementari quando l'intervistatore voglia approfondire o chiarire un argomento (p. es., "Che cosa intende quando dice....?", oppure "Quando è accaduto tutto ciò?"). È fondamentale che le domande non siano poste in maniera tale che il paziente possa ritenere che ci si attende da lui una risposta piuttosto che un'altra o che, addirittura, forniscano delle risposte preconfezionate. Nella tabella 5.1 abbiamo riportato, a titolo di esempio, un brano dell'intervista strutturata dalla *Social Adjustment Scale II* - SAS-II (Schooler et al., 1979).

Se ben condotta, l'intervista strutturata, come abbiamo detto, può risultare sostanzialmente sovrapponibile ad un'intervista clinica libera; se, invece, il valutatore segue troppo rigidamente lo schema di domande che gli è stato proposto, passa in maniera goffa da un argomento all'altro, non tiene conto di risposte già fornite in precedenza, non adatta le domande alla situazione specifica del soggetto, allora l'intervista diviene pesante e noiosa. Se, al contrario, il valutatore usa l'intervista strutturata semplicemente come una guida generale, ponendo le domande in maniera diversa da come suggerito, si ha una riduzione della confrontabilità coi risultati degli altri valutatori. Un'intervista rigidamente strutturata è indispensabile soltanto quando si effettuino indagini (di solito di tipo epidemiologico) su campioni molto vasti, tali da richiedere l'impiego di numerosissimi valutatori che, non potendo avere tutti adeguate competenze cliniche, possono essere addestrati agevolmente all'uso di un'intervista già predisposta. Quando un rigore di questo genere non è necessario, possono essere utili le interviste semistructurate, che aiutano ad esplorare l'area in studio, suggerendo anche delle domande, senza richiedere però un'aderenza assoluta.

IL PERIODO DI RIFERIMENTO

L'impiego di una RS richiede che venga preliminarmente stabilito il periodo a cui si riferisce la valutazione. Questo può essere più o meno lungo in funzione del tipo di ricerca. Di solito, per valutazioni nel corso di un trattamento, si fa riferimento a periodi brevi (4-7 giorni), mentre per valutazioni tese ad esplorare fenomeni a più lenta evoluzione (p. es., l'adattamento sociale), si fa riferimento ad intervalli di tempo anche di 2-3 mesi. È evidente che nell'ambito di una stessa ricerca il periodo di tempo esplorato deve essere mantenuto costante.

L'intervallo di tempo entra in gioco anche nella definizione della gravità di un sintomo assieme ad altri *tre elementi*; l'*intensità del disturbo*, la sua *durata* e la *frequenza* con cui si manifesta. Se il periodo considerato è breve (qualche giorno) ha maggior peso l'intensità del sintomo, se invece il periodo è lungo (diverse settimane) entrano in gioco anche la durata e la frequenza, così, un disturbo persistente ma poco intenso sarà giudicato lieve, un sintomo intenso ma che si manifesta non molto frequentemente sarà considerato medio o lieve, uno intenso e frequente grave o gravissimo.

Solo eccezionalmente le RS contengono item specifici per la valutazione della soggettività e dell'obiettività di alcuni disturbi, cioè di quanto esperito dal paziente e di quanto osservato dall'investigatore.

TAB. 5.I - ESEMPIO DI INTERVISTA STRUTTURATA: SOCIALADJUSTMENT SCALE II - SAS-II (1979)	
B - Situazione sessuale coniugale	
Guida all'intervista	Item
<p><i>Lei e (nome del partner) avete messo in atto qualche pratica anticoncezionale o avete praticato l'astinenza periodica negli ultimi 2 mesi?</i></p> <p><i>Se sì: Quale provvedimento avete adottato?</i></p> <p><i>Se sì: Sono stati soddisfacenti per lei?</i></p> <p><i>[In che misura erano seri questi problemi?]</i></p>	<p>17 - <i>Controllo delle nascite</i></p> <p>1 = Legatura delle tube, vasectomia, spinale</p> <p>2 = La «pillola»</p> <p>3 = Diaframma, preservativo, schiume o creme contraccettive</p> <p>4 = Astinenza periodica o altri mezzi meno sicuri o qualsiasi altra forma di contraccezione che soggetto consideri come tale</p> <p>5 = Nessuna forma di contraccezione</p> <p>6 = Superata l'età fertile, sterilizzazione chirurgica, isterectomia</p> <p>9 = Non pertinente, il membro principale della famiglia non è il partner coniugale, oppure manca un membro principale della famiglia</p>
<p><i>Negli ultimi 2 mesi, aveva desiderio dei rapporti sessuali?</i></p> <p><i>Con quale frequenza, negli ultimi 2 mesi, ha avuto rapporti sessuali?</i></p> <p><i>[Questo è nelle sue abitudini? Quando lei stava male o non si sentiva bene, erano meno frequenti?]</i></p> <p><i>Per la voce "INTERESSE", valutare i sentimenti del soggetto. Per i soggetti del tipo istammi del periodo di esame, valutare solamente sulla base dell'interesse</i></p>	<p>18 - <i>Interesse</i></p> <p>1 = Piacevole e/o attivamente interessato</p> <p>2 = Spesso piacevole e/o spesso interessato</p> <p>3 = Passabile e/o disinteressato-indifferente</p> <p>4 = Spesso spiacevole e/o generalmente disinteressato</p> <p>5 = Spiacevole e/o nessun interesse</p> <p>9 = Il membro principale della famiglia non è il partner coniugale o manca il membro principale della famiglia</p>
<p><i>Se no: Secondo lei, qual è il problema?</i></p> <p><i>[Qualche problema nell'iniziare, nel provare interesse o nel suscitare interesse nel suo partner?]</i></p>	<p>19 - <i>Problemi</i></p> <p>1 = Non ha quasi mai problemi</p> <p>2 = Talvolta ha problemi e/o prova difficoltà</p> <p>3 = Spesso ha problemi e/o prova difficoltà</p> <p>4 = Molto spesso ha problemi e/o prova difficoltà</p> <p>5 = Quasi sempre ha problemi e/o prova difficoltà</p> <p>9 = Il membro principale della famiglia non è il partner coniugale o manca il membro principale della famiglia</p>
<p><i>Ci sono stati cambiamenti nella frequenza dei rapporti sessuali negli ultimi 2 mesi?</i></p>	<p>20 - <i>Frequenza</i></p> <p>1 = In media più di una volta la settimana</p> <p>2 = In media una volta la settimana</p> <p>3 = In media due volte al mese</p> <p>4 = In media non più di una volta al mese</p> <p>5 = Nessun rapporto</p> <p>9 = Il membro principale della famiglia non è il partner coniugale o manca il membro principale della famiglia</p>

Così, ad esempio, il paziente può lamentare disturbi della memoria non obiettivabili o disturbi del sonno che apparentemente non coincidono con quanto riferito dai familiari o dal personale infermieristico: a meno che non si abbia il fondato sospetto di una simulazione, tali disturbi dovranno essere presi in adeguata considerazione poiché rappresentano, con ogni evidenza, un vissuto del paziente.

NORMA INDIVIDUALE E INTERINDIVIDUALE

Nella valutazione della psicopatologia, la *norma interindividuale* avrà un peso maggiore rispetto alla *norma individuale*, fatta eccezione per alcuni aspetti particolari in cui è invece più rilevante la norma individuale, come nel caso, ad esempio, della sessualità, delle abitudini di sonno, della socievolezza. Salvo queste e poche altre eccezioni, quindi, la presenza e la gravità di un sintomo dovranno essere valutate rispetto alla norma prevalente in assoluto piuttosto che in rapporto al comportamento precedente del soggetto. Qualcuno ha posto il problema se i sintomi debbano essere valutati sempre e comunque o soltanto se superano un'ipotetica soglia che delimita il normale dal patologico; la risposta più logica è che i sintomi devono essere sempre valutati qualsiasi sia la loro gravità senza tener conto del giudizio di normalità o di patologia, giudizio che deriva non dalla presenza o meno di un singolo sintomo, ma dall'insieme dei sintomi, dal grado di distorsione della realtà che essi determinano e dal livello di compromissione bio-psico-sociale che essi provocano. Avere qualche item positivo, con gravità lieve o al più moderata, nella maggior parte delle RS, è del tutto normale, poiché è normale essere a volte un po' ansiosi o tristi, irritabili o euforici, distratti o indecisi.

IL PROCEDIMENTO DECISIONALE

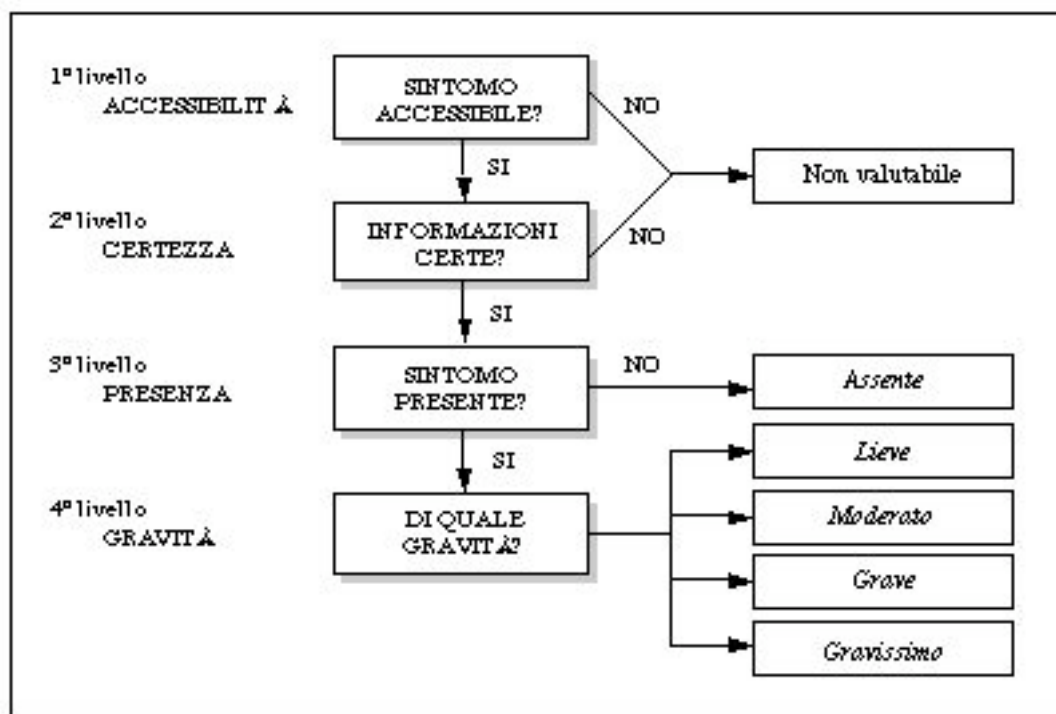


Fig. 5.1 - Albero decisionale per la valutazione di un sintomo.

Il procedimento di valutazione degli item di una scala può essere espresso mediante un albero decisionale che si sviluppa su 4 livelli: l'*accessibilità*, la *certezza*, la *presenza* e la *gravità* (Fig. 5.1).

□

Il primo quesito che il valutatore si pone è se quell'item è o meno *accessibile*, se dispone cioè delle informazioni necessarie per formulare un giudizio. Se quelle informazioni non sono, per qualsiasi ragione, accessibili, l'item è "non accessibile" e, se la RS lo prevede, deve essere segnato come "non valutabile". Se sono disponibili delle informazioni, è necessario valutare, in primo luogo, il

loro grado di affidabilità o di *certezza*; se il valutatore ritiene che la qualità delle informazioni ottenute è tale da non consentire di stabilire con certezza la presenza o meno del sintomo esplorato, anche quel sintomo sarà "non valutabile". La risposta "non valutabile", quindi, non può e non deve esprimere i dubbi soggettivi del valutatore, ma un giudizio sulla qualità dell'informazione ottenuta dopo aver indagato approfonditamente.

Una volta raggiunta una ragionevole certezza sulla qualità dell'informazione, il valutatore dispone degli elementi necessari per stabilire la *presenza* o meno di quel sintomo e, se presente, dovrà valutarne la gravità, tenendo conto, come abbiamo detto, dell'interazione fra intensità, durata e frequenza. La graduazione di un sintomo, soprattutto quando l'item corrispondente non è formalizzato, non riporta, cioè, una descrizione della sintomatologia corrispondente ai diversi livelli di gravità, può essere non facile. Può succedere così che un sintomo di notevole gravità ma di breve durata sia giudicato "lieve" ed un altro, di minore gravità ma più persistente, "grave". L'esperienza clinica dovrebbe fare da guida nei casi controversi, ma essendo necessario tendere alla massima standardizzazione, può essere ragionevole suggerire che un sintomo poco marcato ma persistente venga valutato "lieve", uno di maggiore intensità ma non costantemente presente, "moderato" ed uno intenso e spesso presente, "grave".

LA SCELTA DEGLI STRUMENTI DI VALUTAZIONE - Principi Generali

Nella scelta e nell'impiego degli strumenti di valutazione è necessario tenere presenti alcuni fattori importanti per i risultati che si vogliono ottenere.

Il primo punto da prendere in considerazione è l'*utilità*: uno strumento di valutazione può essere considerato utile solo se aiuta a migliorare il servizio reso ai pazienti e se fornisce un adeguato feedback dell'attività svolta. L'utilità di uno strumento è anche in funzione della sua applicabilità (tempo necessario per la sua somministrazione, semplicità di valutazione, facilità di interpretazione dei risultati, eccetera).

La scelta di uno strumento di valutazione deve essere preceduta dalla precisa identificazione dell'area da esplorare e degli aspetti, delle dimensioni, che meglio la caratterizzano e che, nel caso si vogliano valutare le modificazioni della psicopatologia sotto trattamento, abbiano un maggiore valore predittivo di risposta al trattamento.

Non è del tutto eccezionale, infatti, che una RS venga scelta perché è disponibile, perché si ha con essa una certa familiarità, perché è facile e/o rapida da applicare, senza tener conto degli obiettivi specifici della ricerca in cui deve essere impiegata. Nel contesto di una ricerca psicofarmacologica clinica, ad esempio, ha poco senso impiegare, per valutare l'effetto del trattamento, una RS che esplora dei tratti di personalità, che sono relativamente stabili; analogamente, se uno strumento non esplora completamente l'area che dobbiamo valutare, si rischia di non trovare ciò che effettivamente c'è ma che i nostri strumenti non ci consentono di vedere. Ma anche un'eccessiva estensione dell'area esplorata rischia di compromettere i risultati della ricerca, perché possono essere evidenziati sintomi aspecifici, non strettamente pertinenti alla psicopatologia indagata o, nel caso di valutazione dell'effetto di un trattamento, possono essere valorizzati (fino anche alla significatività statistica) cambiamenti marginali, aspecifici o casuali.

Poiché gli strumenti di valutazione sono generalmente impiegati per verificare i cambiamenti nel tempo della sintomatologia, la *sensibilità* ai cambiamenti è un altro elemento da tenere in considerazione. La sensibilità di una scala, comunque, deve avere come fondamento la *stabilità* della scala stessa (verificabile mediante la test-retest reliability), per non correre il rischio di misurare ogni volta cose diverse, scambiandole per variazioni della sintomatologia esplorata. Una scala, quindi, deve essere sufficientemente stabile da non modificarsi se non si verificano

cambiamenti della sintomatologia ed al tempo stesso abbastanza sensibile da cogliere variazioni anche modeste.

Un altro fattore di cui tenere conto è la *modalità di esplorazione* del comportamento, dei pensieri, dei sentimenti, eccetera, del paziente. Alcuni strumenti valutano il problema in maniera *indiretta* e l'esempio più classico è il *test di Rorschach*; altri, invece, lo esplorano in maniera *diretta*, come avviene con gli strumenti che valutano il comportamento. La maggior parte delle scale si colloca in genere fra questi due estremi. Questo problema deve essere tenuto presente nella scelta degli strumenti: nelle scale di eterovalutazione, un certo grado di valutazione indiretta può essere accettabile anche se può comportare tempi più lunghi per l'assegnazione dei punteggi; è bene ricordare, comunque, che, aumentando la quantità della valutazione indiretta, si riduce l'affidabilità della scala. Nelle scale di autovalutazione è necessario, invece, ridurre al massimo la valutazione indiretta poiché entra in gioco anche la *comprensibilità* da parte del paziente. Un linguaggio troppo astratto, un vocabolario troppo ricercato, frasi troppo complesse, possono rendere il test di difficile comprensione per soggetti scarsamente acculturati, o troppo gravemente disturbati, o emotivamente immaturi; se lo strumento non rispecchia fedelmente e comprensibilmente i problemi del paziente è di scarsa utilità.

Questi strumenti devono essere anche percepiti come *accettabili* dal paziente, sia nel senso che il loro contenuto è percepito come rispondente ai suoi reali problemi e quindi la loro compilazione è ritenuta utile ai fini della comprensione da parte del medico e dell'impostazione del trattamento, sia nel senso che il contenuto non deve essere percepito come intrusivo o addirittura come offensivo, come può accadere per i problemi inerenti alla sessualità in certe subculture.

Uno strumento di autovalutazione, infine, deve essere il più possibile *non-reattivo*, non deve agire, cioè, modificando ciò che si vuole misurare. È questo il caso, ad esempio, delle scale che esplorano comportamenti quali il bere o il fumare; in questi casi la sola valutazione agisce come freno, producendo di per sé un miglioramento. Qualcuno potrebbe obiettare che, se questo è vero, si potrebbe evitare la somministrazione di farmaci semplicemente valutando il paziente con una certa frequenza: in realtà, questo tipo di miglioramento è generalmente di breve durata e quindi di scarsa utilità, ed è spesso sostituito dalla tendenza a minimizzare (e quindi a falsare i risultati). È meglio, perciò, usare strumenti non-reattivi, che sono più stabili e consentono un monitoraggio più efficace ed affidabile.

Un elemento importante, ma spesso trascurato, è quello relativo al tempo necessario per la valutazione. Nella routine clinica le scale sia di auto che di eterosomministrazione non dovrebbero richiedere più di 10-15 minuti per la compilazione e la valutazione clinica, nell'ambito della quale vengono usate le RS, non dovrebbe durare più di 25-30 minuti. Diverso è il caso di strumenti più lunghi e complessi, come le interviste diagnostiche strutturate, che vengono utilizzati di solito una sola volta, per i quali può essere necessario un tempo molto più lungo.

Nel decidere la scelta di uno strumento piuttosto che di un altro, è necessario valutare l'ampiezza e la profondità delle informazioni richieste, la disponibilità del paziente a sottoporsi a lunghe sedute di valutazione, la frequenza delle valutazioni e, per il valutatore, quante valutazioni può ragionevolmente effettuare in un giorno. Pochi pazienti sopportano la somministrazione ripetuta di scale di autovalutazione che richiedano più di 15-20 minuti; se una seduta di valutazione dura più di 45 minuti, anche il valutatore più motivato non può effettuare più di 4 valutazioni al giorno se non si vuole correre il rischio di una perdita notevole di accuratezza.

In conclusione, la scelta di uno strumento di valutazione, soprattutto per un impiego routinario, deve essere tale per cui lo strumento risulti *appropriato*, richieda, cioè, un tempo limitato di compilazione e di interpretazione, sia affidabile e valido, stabile e sensibile, accettabile da parte del paziente ed appropriato per ciò che vogliamo esplorare, il più possibile diretto, ma non reattivo. Strumenti più sofisticati, con tempi più lunghi di applicazione, capaci di fornire

informazioni più approfondite e dettagliate, possono essere utilizzati efficacemente in ricerche specifiche su campioni più limitati di soggetti e secondo protocolli di ricerca specifici.

Una batteria di strumenti di valutazione dovrebbe tener conto dello stato psichico e funzionale del paziente, della sua sensazione di benessere e del suo grado di soddisfazione; lo stato psichico e quello funzionale sono meglio esplorabili mediante scale di eterovalutazione, mentre il benessere ed il grado di soddisfazione, che richiedono la valutazione di sensazioni e percezioni, sono meglio coglibili mediante le scale di autovalutazione. La valutazione del paziente psichiatrico, inoltre, dovrebbe prevedere sia item sensibili al cambiamento, sia item più stabili, capaci di fornire una descrizione più globale del paziente ed indicazioni circa i suoi bisogni, e di predire l'evoluzione della malattia.

Vale la pena di ricordare che la patologia psichiatrica è spesso anche patologia delle emozioni e, pertanto, la valutazione del benessere può confondersi (o covariare) con lo stato psichico (in particolare questo rischio è elevato per i disturbi dell'umore e d'ansia). Della valutazione dello stato psichico, infine, dovrebbero far parte, oltre alla valutazione psicopatologica, una valutazione diagnostica ed una valutazione della comorbidità somatica, ancora oggi troppo spesso trascurata.

LA SCELTA DEGLI STRUMENTI DI VALUTAZIONE - Elementi Pratici

Le RS sono strumenti imperfetti, grossolani, incapaci di fornire una risposta soddisfacente per lo psichiatra; ne è una dimostrazione anche il fatto che le RS messe a punto negli ultimi 20-30 anni sono talmente numerose da creare non poche difficoltà a chi debba scegliere gli strumenti più idonei per le sue finalità di studio e di ricerca. Anche se molte di queste RS hanno avuto vita effimera o diffusione solo locale, non poche sono quelle che hanno avuto un'ampia diffusione ed una lunga vita.

Esistono certamente scale più o meno buone, ma non è possibile farne una graduatoria assoluta poiché, almeno in parte, la bontà di una scala dipende anche dal contesto in cui viene impiegata e dalle finalità e dalle modalità del suo impiego. Pertanto, chi deve scegliere degli strumenti da utilizzare nella propria pratica clinica o per la ricerca, deve avere chiari alcuni parametri essenziali:

- quali sono le caratteristiche generali della scala;
- quale tipo di psicopatologia si deve esplorare;
- che tipo di intervista richiede (strutturata, semistrutturata o libera);
- chi effettua l'intervista (psichiatra, psicologo, assistente sociale, infermiere, familiare, paziente stesso...);
- qual è la qualificazione necessaria per l'intervistatore (esperienza clinica, addestramento...);
- quali sono i criteri per l'attribuzione dei punteggi;
- chi è l'intervistato (il paziente o un informatore-chiave come, ad esempio, un familiare significativo o un operatore che conosce bene il paziente) e qual è il suo livello culturale;
- quanto deve essere numeroso il campione per poter ottenere risultati affidabili;
- in quale setting si svolge la ricerca;

- quale livello di sensibilità della scala è necessario;
- quanto tempo richiede l'intervista;
- se le interviste devono essere ripetute e con quale frequenza.

In questo repertorio verranno fornite quante più possibili informazioni sulle caratteristiche delle RS. Saranno presi in considerazione solo gli strumenti più rilevanti, quelli, cioè, più noti, più diffusi o che abbiamo ritenuto più validi da un punto di vista metodologico, e quelli realmente applicabili, quelli cioè già in uso nel nostro contesto socioculturale o che sono sembrati più adattabili a tale contesto. Per completezza, tuttavia, verrà fatto cenno anche di strumenti non più in uso, ma importanti dal punto di vista storico.

Abbastanza modesto è il numero delle RS originali italiane; la maggior parte è stata creata in lingue diverse dall'italiano (soprattutto in inglese) ed è stata quindi tradotta. La traduzione di una RS pone non pochi problemi, tuttora irrisolti. La procedura classica prevede che una scala, scritta, ad esempio, in inglese e che voglia essere usata, come nel nostro caso, nella lingua italiana, debba essere tradotta in italiano e quindi sottoposta alla cosiddetta "back-translation", cioè alla ri-traduzione in inglese da parte di un traduttore esperto, diverso dal primo, al fine di verificare il rispetto dei significati originali; una volta verificata la fedeltà della traduzione, questa deve essere validata su di un campione italiano. Questo procedimento, cui è stata data una notevole importanza, di fatto si rende necessario soltanto quando si esplorano dimensioni soggettive della psicopatologia e l'uso di un termine al posto di un altro può, quindi, cambiare più o meno ampiamente il significato di un item; quando si impiegano, invece, termini tecnici, di chiara ed inequivocabile comprensione per lo psichiatra, la back-translation può essere omessa. Quando poi la RS è costruita in funzione di uno specifico ambiente socioculturale, oltre alla traduzione è necessario anche un adattamento alla situazione socioculturale diversa, ed allora la back-translation è non soltanto inutile ma anche, al limite, assurda.

In linea generale, chi intende utilizzare le RS ed i questionari, dovrebbe acquisire in primo luogo delle nozioni generali di psicometria e, quindi, prendere piena padronanza degli aspetti teorici e pratici degli strumenti che intende utilizzare (e, se è il caso, sottoporsi ad un apposito addestramento). Dovrebbe poi accertarsi della validità e dell'applicabilità degli strumenti che egli intende utilizzare in funzione del contesto specifico e, se vengono impiegati più valutatori, procedere a delle sedute di inter-rater reliability.

Per concludere, lo psichiatra che intenda usare le RS nell'ambito di una ricerca, deve valutare gli strumenti disponibili in base agli obiettivi che si prefigge ed alle possibilità concrete di utilizzarli. Avrà scelto gli strumenti ottimali se avrà dato la giusta risposta ad una serie di quesiti:

- che tipo di pazienti devo valutare?
- che tipo di collaborazione posso aspettarmi da loro in base alla loro patologia ed al loro livello culturale?
- quanti pazienti devo valutare e con quale frequenza dovrò fare le valutazioni?
- sono pazienti ricoverati o ambulatoriali?
- quanti sono e che livello di qualificazione hanno coloro che effettueranno le valutazioni?
- quale livello di sensibilità è necessario per descrivere al meglio il fenomeno che voglio studiare e/o le sue variazioni?

Esistono alcune regole generali che dovrebbero servire da guida:

- l'autovalutazione è probabilmente più attendibile nei pazienti depressi o ansiosi piuttosto che in quelli psicotici;
- i pazienti ricoverati si prestano meglio a valutazioni frequenti;
- nel caso di valutazioni frequenti, è bene scegliere RS di eterovalutazione semplici e di facile applicazione, riservando quelle più lunghe e complesse alle valutazioni più importanti (prima ed ultima o prima, intermedia ed ultima);
- se i valutatori hanno una buona esperienza clinica e/o hanno una buona familiarità con le RS, ci si può avvalere di strumenti compilabili nel corso di un colloquio libero, ma se questi hanno scarsa esperienza clinica e/o scarso addestramento specifico, è meglio optare per interviste più o meno rigidamente strutturate;
- se vogliamo rilevare fenomeni molto sfumati (o le loro variazioni), si devono scegliere RS molto analitiche, che sarebbero inutili per la rilevazione di fenomeni più grossolani, per i quali sono sufficienti scale più globali;
- se il personale di cui si dispone per la valutazione è scarso, è necessario dare più spazio all'autovalutazione.

Poiché difficilmente un solo strumento è in grado di fornire un quadro esauriente del problema che vogliamo indagare, è preferibile associare, per quanto possibile, più strumenti di valutazione concettualmente diversi, ma che si integrino tra loro; questo porterà ad un aumento della quantità delle informazioni che potranno essere utilmente impiegate nell'analisi dei risultati, ricorrendo, eventualmente, alla consulenza di un buon statistico. Fra le RS scelte, è opportuno che ve ne sia sempre almeno una di buon livello internazionale, in modo da rendere confrontabili i dati della nostra ricerca con quelli di altre ricerche e da favorire la comprensibilità e l'accettabilità della nostra ricerca da parte degli altri ricercatori.

È comunque importante che sia sempre chiaro che le RS ed i questionari sono soltanto degli strumenti e che sta a noi usarli nella maniera più opportuna e, soprattutto, interpretare i risultati ottenuti alla luce del buon senso e dell'esperienza clinica.

BIBLIOGRAFIA

Schooler N, Hogarty GE, Weissman MM: *Social Adjustment Scale II (SAS-II)*. In: Hargreaves WA et al. (Eds.) *"Resource materials for Community Mental Health Program Evaluators"*, ADM, Washington D.C., 1979.

Wing JK, Cooper JE, Sartorius N: *Measurement and classification of psychiatric symptoms. An instruction manual for the PSE and Catego program*. Cambridge University Press, London, 1974.

ALTRE MODALITA' DI MISURAZIONE

Anche se il nostro specifico interesse è rivolto agli strumenti di valutazione standardizzata, di cui abbiamo ampiamente parlato nei capitoli precedenti, non possiamo non accennare anche ad altre e diverse modalità con cui possono essere misurati i problemi dei pazienti psichiatrici, modalità che pure esistono e che possono essere utili in molti casi. La rassegna che qui faremo di queste modalità di misurazione, sarà necessariamente sintetica, ma dovrebbe convincere anche i più scettici che, nonostante tutto, è possibile misurare il comportamento, i sentimenti, i pensieri, l'attività e la qualità della vita.

L'impiego di questi strumenti, così come di quelli di valutazione standardizzata di cui abbiamo detto nei capitoli precedenti, deve seguire alcune linee-guida essenziali: la scelta dovrà sempre orientarsi, fra quelli disponibili, verso lo strumento di misura più appropriato, quello cioè più diretto, valido, sensibile ed affidabile; è altrettanto consigliabile utilizzare più di uno strumento di misura, poiché i problemi con i quali dobbiamo confrontarci sono generalmente complessi e coinvolgono più aree del comportamento, per cui è comunque utile avere un quadro il più completo possibile dei problemi dei nostri pazienti, oltre che una loro visione multifocale che tenga conto, cioè, di punti di vista diversi.

LA VALUTAZIONE COMPORTAMENTALE

In questa sede useremo il termine "comportamentale" nel senso più ampio, riferendoci non soltanto a ciò che il soggetto fa in maniera chiaramente visibile (camminare, parlare, muoversi, sorridere, piangere, eccetera), ma anche a ciò che non è direttamente osservabile ed è invece indirettamente ricavabile da ciò che il soggetto riferisce (presenza di sentimenti di tristezza o di gioia, pensieri che si presentano in maniera ossessiva, eccetera). È indifferente, ai nostri fini, che il comportamento sia rilevato da un osservatore esterno o che sia il soggetto stesso a rilevarlo e quindi a riferirlo. La valutazione comportamentale è particolarmente importante perché è, in generale, l'espressione più diretta dei problemi del paziente, è anche abbastanza precisamente quantificabile e, quindi, sostanzialmente affidabile.

La valutazione comportamentale può essere fatta essenzialmente contando la frequenza, la durata o l'occorrenza di un evento in un determinato periodo di tempo del fenomeno. Si valuta la *frequenza* quando il problema che indaghiamo si verifica troppo spesso (p. es., attacchi di panico, numero di sigarette fumate, numero di volte che viene effettuato un certo rituale, eccetera) e deve essere limitato, o troppo di rado (p. es., numero dei contatti interpersonali, numero dei compiti portati a termine fra quelli assegnati, eccetera) e deve essere incrementato. Quando è più importante il fattore tempo, nel senso che il problema si verifica per tempi eccessivamente prolungati (p. es., crisi di panico, o di emicrania, o di pianto, eccetera) o eccessivamente brevi (p. es., tempo dedicato allo studio, durata dell'erezione peniena, eccetera), si ricorre alla misurazione della *durata*; il comportamento studiato deve avere un inizio ed una fine ben identificabili, in modo da poterlo misurare con precisione. Quando, infine, il comportamento che vogliamo indagare si verifica con una frequenza tale, o dura così a lungo, o è tale da renderne difficile la suddivisione in intervalli precisi e da rendere difficoltoso misurarne la durata o la frequenza, si ricorre allora alla valutazione dell'occorrenza del fenomeno in specificati intervalli di tempo: si stabilisce, cioè, un periodo di osservazione, lo si suddivide in intervalli di tempo uguali tra loro, e si registra se il fenomeno in questione si verifica in ciascuno degli intervalli di tempo stabiliti, indipendentemente dal numero di volte che si verifica e dalla durata.

Nella valutazione comportamentale è necessario individuare con estrema precisione il problema che vogliamo studiare e registrarlo accuratamente, facendo ricorso anche a strumenti di

misurazione e di registrazione (cronometri, segnatempo, registratori, checklist, eccetera).

SCALE DI VALUTAZIONE PERSONALIZZATE

Delle RS parleremo ampiamente più avanti, essendo l'oggetto del nostro specifico interesse. Qui vogliamo ricordare soltanto le *scale personalizzate*, quelle cioè "costruite" specificamente per il soggetto. Due sono, in sostanza, le RS di questo tipo, quelle che fanno riferimento alle modalità con cui il soggetto esperisce determinate esperienze o sensazioni e quelle mirate su specifici problemi.

Le prime, una volta identificata l'area da valutare, richiedono che il soggetto stesso definisca, su di una scala graduata (meglio se abbastanza ampia, 7-10 punti) i limiti estremi di gravità (nessun problema - massima gravità) e gli intervalli (il più possibile uguali) su cui distribuire i livelli intermedi di gravità. Quindi dovrà descrivere accuratamente le sensazioni, le esperienze, i sentimenti, eccetera, che associa a ciascun punto.

Per esempio, se il sintomo esplorato è l'ansia, il soggetto potrà descrivere il livello più basso come "assenza di percezione di sensazioni o sentimenti collegabili all'ansia, o percezioni che non eccedono quelle legate ai normali eventi della vita quotidiana senza particolari aspettative pessimistiche", quello più grave come "stato d'ansia continuo e intenso, non motivato da alcun evento o aspettativa, accompagnato da marcata componente somatica, irrequietezza e agitazione, non alleviabile da alcuna rassicurazione e tale da provocare estrema sofferenza e da impedire qualsiasi attività lavorativa o sociale" e quello intermedio come "stato di allarme pressoché continuo, che stimoli anche banali precipitano in ansia vera e propria, spiacevole anche per le manifestazioni somatiche che comporta, ma comunque non particolarmente intensa e/o di lunga durata, in cui le attività socio-lavorative sono possibili anche se richiedono un certo sforzo".

Le scale mirate a specifici problemi sono basate, invece, sulla descrizione che il soggetto fa del suo problema (o dei suoi problemi); questo deve essere definito accuratamente e dettagliatamente all'inizio e quindi, periodicamente, il soggetto è invitato ad esprimere un giudizio di miglioramento (o di peggioramento) rispetto alla situazione di partenza.

Questi due tipi di scale sono molto utili per la valutazione del singolo soggetto, ma non offrono alcuna possibilità di confronto con altri soggetti o con casistiche più ampie.

DIARIO

L'uso di un diario in cui il soggetto registra gli eventi che ritiene rilevanti per la sua situazione può essere molto utile perché aiuta a focalizzare meglio il problema, riduce le possibili distorsioni cui può andare soggetta l'esposizione del problema stesso da parte del paziente, rappresenta una buona misura dell'andamento del problema nel corso del trattamento, può fornire informazioni su eventi o situazioni che possono essere in rapporto al problema stesso.

Il diario può essere libero, ma è preferibile che abbia un formato standardizzato, come potrebbe essere un foglio giornaliero suddiviso in colonne nelle quali registrare l'ora in cui si è verificato l'evento e la sua natura, quali reazioni ha determinato, se era presente qualcuno (e, in caso affermativo, chi), eccetera. A seconda dei casi, si può chiedere al paziente di registrare tutto ciò che egli ritiene rilevante oppure di registrare eventi specifici e più o meno precisi dettagli di tali eventi.

Particolarmente consigliabile è l'uso del diario in alcuni tipi di disturbi come nei disturbi fobici

(fobie specifiche e, soprattutto, fobia sociale), ossessivo-compulsivi ed in quelli di panico con o senza agorafobia.

MISURE NON INTRUSIVE

Abbiamo detto che alcuni strumenti di valutazione tendono ad essere "reattivi", modificano cioè il comportamento che vogliamo indagare soltanto per il fatto che il soggetto sa di essere sottoposto a valutazione. In generale, questo effetto è da evitare e perciò si tende, in vario modo, a rendere per quanto possibile "non-reattivi" gli strumenti di valutazione standardizzata. Questo può essere ottenuto adottando particolari tecniche di valutazione, ma esistono anche le cosiddette *misure non intrusive*, che evitano la reattività da parte del soggetto, valutandolo senza che questi se ne renda conto. Questo può essere fatto misurando i risultati del suo comportamento (contando, ad esempio, i mozziconi di sigaretta, misurando la quantità di vino rimasta, valutando la quantità e/o la qualità del lavoro svolto, eccetera) od osservando direttamente il comportamento attraverso specchi unidirezionali. In alcuni casi possono essere usate informazioni raccolte in altre circostanze e per altri fini (il registro delle presenze scolastiche, ad esempio, può fornire informazioni sull'impegno scolastico, il registro del Pronto Soccorso ci può dire se il paziente ha avuto ricadute e quante e così via).

MISURAZIONI ELETTROFISIOLOGICHE

È questo un settore solitamente poco usato, ma che può dare, in molti casi, informazioni preziose sullo stato di un problema e sulle sue modificazioni nel tempo. Se in molti casi si tratta di esami che hanno solo un rapporto indiretto con il problema in studio, come possono essere l'elettrocardiogramma (ECG) o il riflesso psicogalvanico (skin resistance response - SRR) per la valutazione dell'ansia, in altri casi il rapporto è più diretto, come nel caso della pletismografia e della valutazione delle erezioni notturne per i problemi sessuali. Fra gli esami elettrofisiologici ricordiamo, oltre ai già citati, l'elettromiogramma (EMG), l'elettroencefalogramma (EEG), la resistenza elettrica cutanea (skin resistance level - SRL), lo studio dei potenziali evocati, eccetera.

GENERALITA'

La schizofrenia è certamente il nucleo centrale della patologia psichiatrica e, in certo modo, ne rappresenta l'aspetto più caratteristico. Eppure, nonostante il suo ruolo centrale, le RS costruite specificamente per la sua valutazione sono in numero limitato. Questo, in realtà, è vero solo in parte, dato che le scale di valutazione della psicopatologia generale, di cui abbiamo parlato nel capitolo precedente, sono state costruite tutte (o quasi) per la valutazione dei "pazienti psicotici" e quindi, di fatto, anche se non specificamente, per i pazienti schizofrenici o appartenenti, comunque, allo spettro schizofrenico. La sintomatologia che caratterizza questi disturbi, infatti, è talmente vasta ed eterogenea da comprendere la psicopatologia nella sua globalità e, pertanto, le scale più adatte alla sua valutazione finiscono per essere quelle che valutano la psicopatologia generale. D'altro canto l'esperienza ci insegna che i diversi aspetti sintomatologici della schizofrenia (e dei disturbi psicotici in genere), dai disturbi del comportamento a quelli del pensiero, dalle alterazioni psicosensoriali a quelle dell'affettività, dalla perdita della progettualità all'impoverimento ideativo, siano abbastanza autonomi e indipendenti l'uno dall'altro nell'evoluzione spontanea, tanto da caratterizzare diversi sottotipi del disturbo, e rispondano in misura e tempi diversi ai vari trattamenti. È per questo che le scale di valutazione generale meglio si prestano alla valutazione della sintomatologia psicotica e delle sue modificazioni sotto trattamento. Per contro, le scale costruite specificamente per la valutazione della schizofrenia sono in numero limitato e sono generalmente scale studiate per la valutazione di aspetti specifici, limitati, del quadro clinico, come, ad esempio, i sintomi positivi o quelli negativi. Certamente, come opportunamente rileva Pancheri (1995), la valutazione quantitativa della schizofrenia è oggettivamente difficile per **tre ragioni principali**:

- in *primo luogo* perché la schizofrenia è il disturbo psichiatrico meno facilmente definibile ed identificabile in termini operativi, e la misurazione della psicopatologia è tanto più facile quanto più definita e delimitata è l'entità che vogliamo misurare;
- in *secondo luogo* perché nella schizofrenia non esistono, in sostanza, sintomi patognomonicamente, ma un polimorfismo sintomatologico con variabilità di sintomi da un soggetto ad un altro e nello stesso soggetto nei vari momenti della malattia;
- *infine*, perché la schizofrenia è un disturbo caratterizzato da sintomi di stato e da sintomi di decorso e quelli di stato, ai quali è prevalentemente rivolta la valutazione psicometrica, variano notevolmente nelle varie fasi della malattia.

LE SCALE PER LA SCHIZOFRENIA

Per queste ragioni la BPRS ha avuto un ruolo di primo piano nella valutazione della schizofrenia, almeno fino a quando ha prevalso la concezione bleuleriana di questo disturbo. Per questo Autore i sintomi "fondamentali" della schizofrenia, quelli comuni, cioè, a tutte le forme cliniche, erano da individuarsi nella *disorganizzazione del pensiero*, nelle *alterazioni dell'affettività*, nell'*autismo* e nell'*ambivalenza*, mentre erano da considerarsi sintomi "accessori" tutti gli altri, quelli cioè non necessariamente presenti in ogni caso, come le allucinazioni, le variazioni dell'umore, le alterazioni dello stato di coscienza e le manifestazioni catatoniche. Negli anni Sessanta e Settanta c'era stata una certa rivalutazione dei deliri e delle allucinazioni al punto che i *sintomi di primo ordine di Schneider* (voci udibili, voci sotto forma di discorsi e repliche, voci che commentano l'operato del soggetto, esperienza di passività somatica, furto del pensiero ed altre esperienze di influenzamento del pensiero, diffusione del pensiero, percezioni deliranti e sensazioni di

costrizione del dominio della volizione, dei sentimenti e degli impulsi), per quanto non patognomonicamente, erano considerati come una guida infallibile per la diagnosi, tanto da avere un ruolo importante negli RDC, nel PSE e nel DSM-III. Ma i sintomi di primo ordine di Schneider mostrarono ben presto i loro limiti e si tornò, con qualche aggiustamento, ai concetti di Bleuler ed in particolare alla denominazione di "**sintomi negativi**" dei sintomi fondamentali ed alla denominazione di "**sintomi positivi**" di quelli accessori. La distinzione fra due sindromi schizofreniche, una con sintomi negativi ed una con sintomi positivi — espressione, i primi, di perdita o riduzione di alcune funzioni e, i secondi, di alterazione o distorsione di altre funzioni —, riproposta da Crow nel 1980, ha suscitato notevole interesse ed attenzione. Nella formulazione originale di Crow, la **schizofrenia di tipo 1**, caratterizzata da buon adattamento premorbo, funzione cognitiva normale, prevalenza di sintomi positivi e buona risposta al trattamento con neurolettici, sarebbe sostenuta da un'eccessiva trasmissione dopaminergica (da cui i buoni risultati del trattamento con bloccanti dopaminergici), mentre la **schizofrenia di tipo 2**, caratterizzata dalla predominanza di sintomi negativi, da scarso adattamento premorbo e compromissione cognitiva, avrebbe alla base un'anomalia strutturale cerebrale (ampliamento ventricolare evidenziabile alla TAC) e questo spiegherebbe la scarsa risposta al trattamento. La dicotomia proposta da Crow, per quanto lasci insoluiti, alla luce dell'esperienza clinica, alcuni problemi (di cui i principali sono rappresentati dalla presenza di forme cosiddette "miste", in cui coesistono sintomi delle due serie, e da quei quadri che, nel tempo, vanno incontro a drastiche variazioni) che non vogliamo certamente affrontare in questa sede, è stata il punto di partenza di numerose ricerche e, per quello che ci riguarda, anche per lo sviluppo di strumenti standardizzati di valutazione che a quella dicotomia fanno più o meno esplicitamente riferimento. Un primo contributo all'inquadramento di queste due forme di schizofrenia è stato fornito, sul piano psicometrico, dalla *Scale for the Assessment of Thought, Language, and Communication Disorders - TLC*, una scala metodologicamente ben costruita e con un solido substrato teorico, messa a punto dalla Andreasen (1979) per studiare specificamente i disturbi cognitivi degli schizofrenici attraverso l'esplorazione di 18 categorie di disturbo del pensiero. La TLC esplora, mediante un'intervista strutturata, le esperienze recenti e gli interessi personali dei soggetti senza prendere in considerazione l'insieme della sintomatologia. Dopo un breve colloquio informale, viene posta al paziente una serie di domande relative ad argomenti diversi (dall'astratto al concreto, dai problemi personali ai rapporti interpersonali); il punteggio è una sintesi della frequenza e della gravità del sintomo. L'analisi della TLC ha consentito di distinguere due tipi principali di disturbo del pensiero (positivo e negativo) e di isolare alcune forme di disturbo del pensiero, fra cui la "povertà di linguaggio" e la "povertà del contenuto del linguaggio" che caratterizzano pazienti con scarsa risposta al trattamento e con prognosi peggiore. La stessa Andreasen ha poi messo a punto due scale, una per la valutazione dei sintomi negativi, la *Scale for the Assessment of Negative Symptoms - SANS* (1982) ed una, la *Scale for the Assessment of Positive Symptoms - SAPS* (1982), per la valutazione di quelli positivi. La SANS e la SAPS hanno rappresentato un vero e proprio salto di qualità nella valutazione della schizofrenia, poiché hanno consentito la definizione e la valutazione quantitativa delle manifestazioni più specifiche del disturbo. La definizione precisa dei termini e l'ampiezza dell'area psicopatologica coperta rappresentano le caratteristiche migliori delle due scale della Andreasen, mentre un po' carente è la definizione dei punteggi. Per quanto di agevole applicazione, entrambe le scale richiedono una buona esperienza clinica da parte del valutatore, e questo vale ancora di più per la SANS che richiede una particolare abilità nel far emergere i sintomi e dar loro una quantificazione. La SANS è composta da 25 item suddivisi in 5 cluster (Piattezza o ottundimento affettivo, Alogia, Abulia, Apatia, Anedonia, Asocialità ed Attenzione) e per ogni cluster è richiesta una valutazione globale. Gli item della SAPS sono 34 e sono raggruppati in 4 cluster (Allucinazioni, Deliri, Anomalie del comportamento e Disturbi formali positivi del pensiero); anche per questa scala è richiesta una valutazione globale dei singoli cluster. Le due scale della Andreasen, tuttavia, non forniscono indicazioni sul livello generale della psicopatologia del paziente schizofrenico e, per contro, la scala di valutazione della psicopatologia generale più usata nella ricerca e nella pratica clinica con questi pazienti, la BPRS, non fornisce adeguate informazioni sui sintomi positivi e negativi che caratterizzano questi pazienti. Kay e collaboratori (1987), nel tentativo di superare i limiti di

questi strumenti, hanno elaborato la *Positive And Negative Schizophrenic Symptoms - PANSS* integrando i 18 item della BPRS con i 12 della *Psychopathology Rating Scale - PRS* (Singh e Kay, 1975) ed articolando i 30 nuovi item in tre distinti cluster, uno per i sintomi positivi (7 item), uno per quelli negativi (7 item) ed uno per i sintomi psicopatologici generali (16 item). Il manuale che accompagna la scala fornisce una dettagliata spiegazione dei singoli item e dei criteri di quantificazione dei sintomi (che sono valutati su una scala a 7 punti). L'intervista prevede una fase iniziale, non strutturata, durante la quale si cerca di stabilire un rapporto con il paziente, si raccolgono informazioni anamnestiche e si incominciano a valutare gli aspetti comportamentali ed interpersonali, ed una seconda fase, strutturata, in cui vengono esplorati gli altri ambiti psicopatologici, dall'umore alla cognitività, dall'ansia alle capacità di ragionamento astratto. Informazioni sul comportamento del paziente nel corso della settimana precedente l'intervista vengono raccolte dai familiari o, per i pazienti ricoverati, dal personale infermieristico. In questo modo, si può **valutare** in maniera più completa i diversi ambiti psicopatologici, dal comportamento ai rapporti interpersonali, dai processi cognitivi al contenuto del pensiero, **caratterizzare** i pazienti in termini di appartenenza ad uno dei due sottotipi di schizofrenia (a sintomi positivi o a sintomi negativi), **dare una misura affidabile** della gravità complessiva del quadro clinico. La PANSS, quindi, grazie alla descrizione analitica degli item, permette, al pari della SANS e SAPS, di cogliere gli elementi sintomatologici specifici della serie negativa e positiva, ma è operativamente più flessibile e, grazie all'introduzione di item esploranti la sintomatologia generale, consente di valutare anche quei sintomi che interagiscono con la sintomatologia positiva e negativa pur non facendone parte (Pancheri et al., 1995). Grazie all'analisi statistica (analisi delle componenti principali), la PANSS ha messo in evidenza, nella popolazione schizofrenica, 4 cluster di sintomi:

- una *Sindrome Positiva*, caratterizzata da deliri ed allucinazioni,
- una *Sindrome Negativa*, comprendente l'isolamento e la mancanza di spontaneità;
- una *Sindrome Depressiva*, di cui fanno parte la colpa, l'ansia e la depressione;
- una *Sindrome di Eccitamento* comprendente l'ostilità ed uno scarso controllo sugli impulsi.

Queste sindromi si collocano spazialmente a formare una piramide che ha per base un triangolo rettangolo che ha agli estremi dell'ipotenusa le sindromi Negativa e Positiva ed all'angolo retto la Depressione; l'Eccitamento rappresenta l'apice della piramide e si colloca perciò al di fuori del piano in cui si trova la base. La distribuzione di queste sindromi nei pazienti mostra che queste non sono dimensioni separate ma fanno parte di un continuum in cui:

- il lato che unisce la sindrome Positiva a quella Negativa descrive i pazienti con *disturbi cognitivi* quali disorganizzazione concettuale e ridotta capacità di critica, classificabili nella categoria diagnostica della *schizofrenia di tipo disorganizzato*;
- quello che unisce la sindrome Positiva a quella Depressiva include i pazienti con *sospettosità ed idee persecutorie*, classificabili nella *schizofrenia di tipo paranoide*;
- quello, infine, che unisce la sindrome Negativa a quella Depressiva è caratterizzato dal *rallentamento motorio* e dal *negativismo* e raccoglie pazienti classificabili nella *schizofrenia di tipo catatonico*.

Il modello a piramide non implica quindi che le sindromi isolate siano mutuamente escludentisi, ma sottolinea soltanto che alcune manifestazioni della sintomatologia schizofrenica sono distinte rispetto alle altre. Così, la sindrome Positiva non è correlata con la schizofrenia di tipo catatonico (asse Negativo-Depressivo), quella Negativa non è correlata con la schizofrenia paranoide (asse Positivo-Depressivo) e la sindrome Depressiva non è correlata con la schizofrenia disorganizzata

(asse Positivo-Negativo). Studi successivi, mediante l'analisi delle componenti principali della PANSS (Lepine, 1991; Lindström e von Knorring, 1993; von Knorring e Lindström, 1995), hanno messo in evidenza l'importanza di un quinto fattore, il fattore "cognitivo" per un più completo ed articolato inquadramento del disturbo schizofrenico. Recentemente Pancheri ed i suoi collaboratori hanno proposto due scale per la valutazione della schizofrenia, la *3TRE* (1995), e la *Scala per la valutazione della Disorganizzazione del pensiero - SCADIS* (1996). La 3TRE è stata studiata in funzione di una valutazione rapida e sintetica della sintomatologia schizofrenica. Si basa sul presupposto teorico della dicotomia tra sintomi positivi e negativi e si propone di superare i limiti delle scale precedenti (BPRS, SANS, SAPS e PANSS), limiti che sono stati individuati nel fatto che, quelle scale:

- contengono sintomi per i quali non vi è accordo unanime per l'attribuzione all'una o all'altra delle due categorie sintomatologiche;

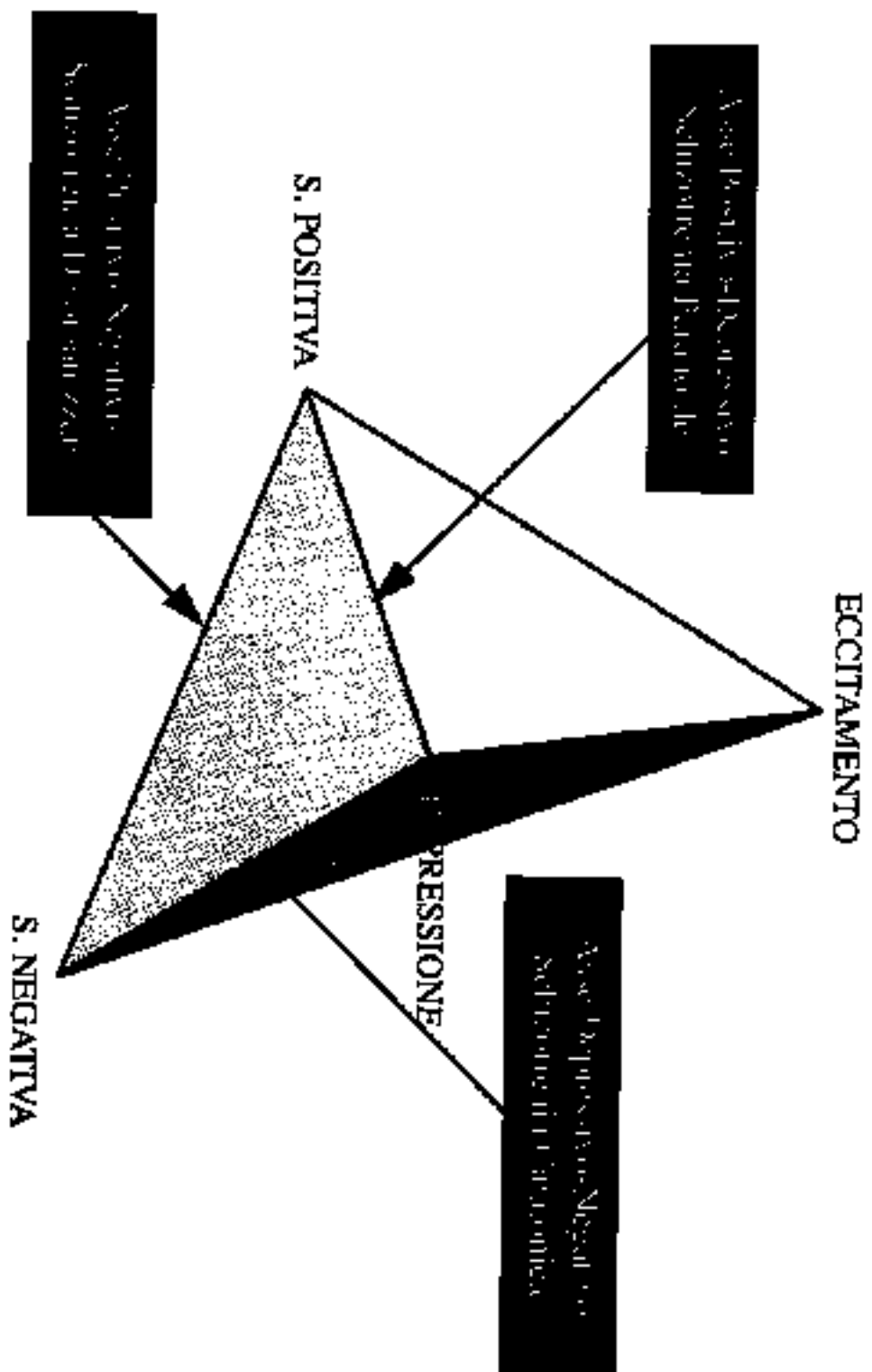


Fig. 9.1 - Rappresentazione spaziale dei rapporti fra le 4 sindromi ricavate dall'analisi delle componenti principali effettuata sulla PANSS.

- presentano, fra i sintomi negativi, item che riguardano il comportamento e le interazioni sociali e che dovrebbero essere considerati, in realtà, una conseguenza dei sintomi sia positivi che negativi;
- hanno un'attribuzione del punteggio alquanto aleatoria, mancando una descrizione precisa dei diversi livelli, e questo lascia uno spazio eccessivo alla soggettività del valutatore;
- richiedono, per la loro utilizzazione, un training specifico e per la loro compilazione sono necessari, in genere, tempi abbastanza lunghi.

La 3TRE è stata formulata in modo tale da:

- consentire la valutazione dei sintomi "essenziali" della sindrome positiva e di quella negativa della schizofrenia, di quei sintomi, cioè, che sono unanimemente attribuiti ad una delle due sindromi;
- poter valutare separatamente i sintomi positivi da quelli negativi, ed entrambi da quelli comportamentali che, oltre ad essere spesso conseguenza dei primi due, non sono rilevabili mediante il colloquio clinico;
- fornire precise definizioni dei livelli di gravità in modo da rendere più uniforme l'attribuzione dei punteggi;
- dare al valutatore criteri operativi per la valutazione psicopatologica e semeiologica di routine attraverso la formulazione e la definizione dei singoli item.

La 3TRE consta di tre sezioni (*Sintomi positivi*, *Sintomi negativi* e *Sintomi comportamentali*) ognuna delle quali è costituita da 3 item considerati "essenziali" per l'esplorazione di quell'area; la valutazione è fatta su di una scala da 1 a 5 ed ogni livello di gravità è adeguatamente descritto. Gli studi di validazione hanno documentato che la scala possiede proprietà psicometriche non inferiori a quelle delle scale concorrenti, rispetto alle quali è di più semplice e rapida somministrazione (Pancheri, 1995; Pancheri et al., 1995). La SCADIS è stata messa a punto per coprire un importante settore della sintomatologia schizofrenica trascurato dagli ormai ben validati strumenti per la misura della sindrome positiva e negativa, il settore, cioè, della disorganizzazione ideativa, che può essere considerato il terzo pilastro su cui si fonda il disturbo schizofrenico. Questo strumento si propone di "valutare quantitativamente e qualitativamente la destrutturazione delle modalità formali della comunicazione associate alla dimensione psicopatologica della "disorganizzazione"" (Pancheri et al., 1996) e non prende in considerazione, quindi, le tradizionali "alterazioni formali del pensiero", ma i meccanismi che le sottendono in funzione comunicativa. La SCADIS è composta da 7 item: i primi cinque — *Finalizzazione*, *Capacità di astrazione*, *Sistema simbolico di riferimento*, *Logica di riferimento* e *Ridondanza procedurale* — fanno riferimento a componenti del processo comunicativo, mentre gli altri due — *Procedura di controllo* e *Qualità della comunicazione* — hanno un carattere generale. Gli item sono accuratamente descritti e sono fornite adeguate indicazioni per la quantificazione della sintomatologia. La scala è ancora in fase di validazione. Per concludere, passeremo rapidamente in rassegna altri strumenti di valutazione dei disturbi schizofrenici (o, più in generale, psicotici) meno noti.

La *Level of Functioning Scale* - *LFS* di Strauss e Carpenter (1972) è una scala di 9 item messa a punto per la valutazione dell'esito della schizofrenia negli studi a lungo termine. Esplora la frequenza e la qualità dei contatti sociali e dell'attività lavorativa utile, la presenza di sintomi, la capacità di far fronte alle necessità di base ed il livello generale di funzionamento. Avendo come obiettivo prevalente lo studio delle relazioni sociali, ne accenneremo nel capitolo dedicato alla valutazione dell'adattamento sociale (Cap. 18). La *Magical Ideation Scale* - *MIS* (Eckblad e Chapman, 1983) e la *Perceptual Aberration Scale* - *PAS* (Chapman e Chapman, 1987) sono due

scaie messe a punto per la valutazione della predisposizione alla schizofrenia e/o al disturbo schizotipico di personalità. La MIS, composta da 30 item, è basata sull'ipotesi di Meehl che il pensiero magico caratterizzi i alle percezioni strane, bizzarre, che spesso hanno, secondo Krepelin e Bleuler, gli schizofrenici. Le due scaie, usate spesso assieme, hanno dimostrato una buona capacità di discriminare soggetti normali e pazienti non schizofrenici da soggetti schizofrenici e schizotipici. Nell'ambito degli strumenti che esplorano aspetti del carattere considerati contigui, o predisponenti, o marginali, rispetto ai disturbi psicotici, si colloca anche la: *Raulin Intense Ambivalence Scale* - *RIAS* (Raulin, 1984), una scaia messa a punto per la valutazione dell'ambivalenza, cioè della presenza contemporanea (o in rapida sequenza) di forti sentimenti positivi e negativi nei confronti dello stesso oggetto o della stessa situazione, che è considerata un'importante caratteristica della schizofrenia. La *RIAS*, composta da 45 item dicotomi (vero/falso), richiede tempi modesti per la compilazione ed ha mostrato ottime caratteristiche psicometriche. Dello stesso Autore è anche la *Social Fear Scale* - *SFS* (Raulin e Wee, 1984), una scaia di autovalutazione mirata allo studio dell'evitamento e dell'inadeguatezza sociale e dei problemi di rapporto interpersonale, comportamenti ed atteggiamenti considerati caratteristici dei soggetti con disturbo schizotipico di personalità (e quindi, verosimilmente, con predisposizione alla schizofrenia). Di questa scaia diremo con maggior dettaglio nel capitolo 22 quando parleremo dei disturbi di personalità. Lo *Auditory Hallucinations Questionnaire* - *AHQ* (Husting e Hafner, 1990) è uno strumento che misura la presenza e le caratteristiche delle allucinazioni uditive, ma anche i deliri e l'umore. È formulato in forma di diario che il soggetto deve riempire tre volte al giorno (ore 8, 14 e 20) e dal quale si possono ricavare, oltre ad un punteggio totale, i punteggi nelle componenti delirante (item 1 e 2), allucinatoria (item 3-6) e dell'umore (item 7-9).

LA VALUTAZIONE DI ALTRI ASPETTI DELLA SCHIZOFRENIA

Per lo psichiatra è importante capire l'atteggiamento dello schizofrenico nei confronti della propria malattia e della propria salute ed un modo indiretto per ricavare informazioni di questo tipo è la valutazione del rapporto che il paziente ha con il trattamento psicofarmacologico. A questo scopo Hogan e collaboratori (1983) hanno sviluppato il *Drug Attitude Inventory* - *DAI* che si propone di misurare la risposta soggettiva al trattamento in modo da approfondire la conoscenza dei fattori che influenzano la *compliance* al trattamento. Lo strumento, di autovalutazione, era composto, nella versione originale, da 30 item (*DAI-30*) che sono stati ridotti a 10 (*DAI-10*) nelle revisioni successive. Gli item esplorano, alcuni, le sensazioni derivanti dal trattamento (rilassato, stanco, rallentato, eccetera), altri riflettono, invece, l'atteggiamento verso il trattamento e come è vissuto; lo strumento consente di identificare i pazienti a maggiore rischio di scarsa *compliance* per i quali è consigliabile ricorrere a farmaci meglio tollerati e/o ad interventi psicologici di supporto. Al di là degli aspetti psicopatologici specifici, la schizofrenia propone anche problemi di altro tipo, ma non per questo di minore importanza, dei quali è indispensabile tenere conto e dei quali fornire un'adeguata valutazione. Fra questi problemi, quelli dell'inserimento (o re-inserimento) e dell'adattamento sociale, meritano in modo particolare la nostra attenzione. La compromissione del ruolo socio-lavorativo è uno degli aspetti più caratteristici della patologia schizofrenica. Ma se il quadro psicopatologico può essere controllato in maniera spesso soddisfacente dal trattamento psicofarmacologico, i deficit sul piano sociale e lavorativo sono scarsamente sensibili a questo tipo di trattamento ed hanno, per contro, un grave impatto sulla qualità della vita, compromettendo in misura crescente i rapporti interpersonali a tutti i livelli e rendendo praticamente impossibile, alla maggior parte dei pazienti, una vita sociale accettabile. per il problema dell'adattamento sociale era già presente quando la psichiatria era ancora incentrata sull'ospedale psichiatrico, ma ha assunto un ruolo di primo piano da quando la psichiatria ha superato la fase "manicomiale" ed è diventata psichiatria della comunità. La corretta e completa valutazione dei pazienti psichiatrici, ed in maniera preminente di quelli schizofrenici, non può prescindere perciò dalla valutazione dell'adattamento sociale, del quale parleremo

diffusamente al capitolo 18.

BIBLIOGRAFIA

Andreasen NC: *Thought, Language, and Communication Disorders. I. Clinical assessment, definition of terms, and evaluation of their reliability.* Arch Gen Psychiatry, 36:1315, 1979.

Andreasen NC: *Thought, Language, and Communication Disorders. II. Diagnostic significance.* Arch Gen Psychiatry, 36:1325, 1979.

Andreasen NC, Olsen S: *Negative v. positive schizophrenia: Definition and validation.* Arch Gen Psychiatry, 39:789, 1982.

Andreasen NC: *Negative symptoms of schizophrenia: Definition and reliability.* Arch Gen Psychiatry, 39:784, 1982.

Chapman LJ, Chapman JP: *The search for symptoms predictive of schizophrenia.* Schizophrenia Bull 13:497, 1987.

Crow TJ: *Positive and negative schizophrenic symptoms and the role of dopamine.* Br J Psychiatry, 137:383, 1980.

Eckblad M, Chapman LJ: *Magical ideation as an indicator of schizotypy.* J Cons Clin Psychol, 51:215, 1983.

Hogan TP, Awad AG, Eastwood R: *A self-report scale predictive of drug compliance in schizophrenics: Reliability and discriminative validity.* Psychol Med, 13:177, 1983.

Husting HH, Hafner RJ: *Persistent auditory hallucinations and their relationship to delusion and mood.* J Nerv Ment Dis, 178:264, 1990.

Kay SR, Fiszbein A, Opler LA: *The Positive and Negative Syndrome Scale (PANSS) for schizophrenia.* Schizophrenia Bull, 13:261, 1987.

Lepine JP: *Dimensions positives et négatives dans les schizophrénies.* Les Cahiers de Prisme, 1:23, 1991.

Lindström E, von Knorring L: *Principal component analysis of the Swedish version of the Positive and Negative Syndrome Scale (PANSS) for schizophrenia.* Nordic J Psychiatry, 47:257, 1993.

Pancheri P: *La 3TRE: una scala per la valutazione rapida della sintomatologia schizofrenica. I: Basi teoriche, descrizione, campi di applicazione.* Rivista di Psichiatria, 30:205, 1995.

Pancheri P, Brugnoli R et al.: *Valutazione dimensionale della sintomatologia schizofrenica. Validazione della versione italiana della Scala per la valutazione dei Sintomi Positivi e Negativi (PANSS).* Giorn Ital Psicopat, 1-3:60, 1995.

Pancheri P, Marconi PL, Brugnoli R: *SCADIS: una nuova scala per la valutazione della disorganizzazione. 1. Costrutti teorici, principi organizzatori, descrizione della scala.* Giorn Ital Psicopat, 3:216, 1996.

Pancheri P., Romiti R et al.: *La 3TRE: una scala per la valutazione rapida della sintomatologia schizofrenica. II: Validazione*. Rivista di Psichiatria, 30:224, 1995.

Pancheri P, Romiti R et al.: *SCADIS: una nuova scala per la valutazione della disorganizzazione. 2. Omogeneità, struttura fattoriale, validità concorrente, attendibilità*. Giorn Ital Psicopat, 3:232, 1996.

Raulin ML: *Development of a scale to measure intense ambivalence*. J Consult Clin Psychol, 52:63, 1984.

Raulin ML, Wee JL: *The development and initial validation of a scale to measure social fear*. J Clin Psychol, 40:780, 1984.

Singh MM, Kay SR: *A comparative study of haloperidol and chlorpromazine in terms of clinical effects and therapeutic reversal with benztropine in schizophrenia: theoretical implication for potency differences among neuroleptics*. Psychopharmacologie, 43:103, 1975.

Strauss JS, Carpenter WT: *The prediction of outcome in schizophrenia*. Arch Gen Psychiatry, 27:739, 1972.

Von Knorring L, Lindström E: *Principal components and further possibilities with the PANSS*. Acta Psychiat Scand, 91(s):5, 1995.