

Fedeltà di un dato, affidabilità di una definizione operativa

di Alberto Marradi

1. L'attendibilità dal telescopio alla matrice dei dati

1.1. Il concetto di errore di osservazione, e il concetto speculare di precisione/accuratezza, nascono in astronomia, allorché diviene abituale osservare più volte lo stesso fenomeno per controllare la componente erratica di ciascuna singola osservazione. L'accuratezza è concepita come una proprietà della coppia costituita da uno strumento e dall'osservatore che lo utilizza. Essa può essere concepita come l'inverso della varianza di tutti i risultati di osservazioni relative allo stesso fenomeno — più precisamente: a stati (dati per) uguali dello stesso oggetto sulla stessa proprietà. Maggiore è tale varianza, minore è ritenuta l'accuratezza della coppia strumento-osservatore.

Secondo Cronbach (1947, 1-2), questa concezione riposa su due assunti:

- (a) lo stato dell'oggetto non cambia spontaneamente durante l'intervallo fra la prima e l'ultima osservazione;
- (b) lo stato dell'oggetto non è alterato dal processo di osservazione.

Dato che le osservazioni degli astronomi si riferiscono abitualmente alla posizione di un corpo celeste, l'assunto (b) è pacifico e poteva restare inespresso, mentre l'assunto (a), rigorosamente inteso, è infondato, e deve essere "liberalizzato" interpretandolo così:

- (a) il cambiamento spontaneo dello stato effettivo di un oggetto ha, sui risultati di un'osservazione effetti trascurabili rispetto a quelli dell'erraticità della coppia strumento-osservatore.

Per estendere la portata di questi concetti al di fuori dell'astronomia, cioè per considerare 'errore di osservazione' (attribuibile alla coppia strumento-operatore) ogni differenza fra lo stato effettivo di un oggetto su una proprietà e il risultato della corrispondente rilevazione, è necessario esplicitare altri due assunti che nel caso dell'astronomia potevano restare sullo sfondo in quanto anch'essi pacifici:

- (c) esiste uno stato effettivo dell'oggetto sulla proprietà in questione;
- (d) le differenze fra oggetti dello stesso tipo (ad esempio, fra atomi dello stesso elemento, oppure fra esseri umani) sono, se esistono, trascurabili al fine di stabilire l'accuratezza di una coppia strumento-operatore.

Anche ammesso che esista (assunto c), lo stato effettivo è per definizione inconoscibile al di fuori di un qualche processo di osservazione. Se si tratta della posizione di un corpo celeste, la robusta struttura teorica della loro disciplina permette agli astronomi di calcolarla sulla base di precedenti osservazioni, "aggiustate" alla luce delle teorie. Si deve al matematico Gauss la scoperta del fatto che le posizioni osservate assumevano una distribuzione campanulare attorno alla posizione ritenuta (in base a tali calcoli) quella effettiva, che tendeva quindi a costituire la moda, la mediana e la media della distribuzione. Il fatto che una distribuzione campanulare o gaussiana sia stata in seguito riscontrata per le osservazioni di altri fenomeni del mondo fisico ha autorizzato gli scienziati a ritenere, o comunque dare per scontato, che lo stato effettivo equivalesse in ogni caso alla mediana della distribuzione degli stati osservati.

1.2. In epoca di positivismo ruggente, questo meccanismo di accertamento dello stato effettivo è stato trasferito dal piano gnoseologico a quello ontologico. Sulla base di dati di natura e provenienza disparata (circonferenza toracica di soldati scozzesi, altezza di reclute francesi, tassi di criminalità, di nuzialità e di suicidio in vari paesi europei), l'astronomo e demografo belga Adolphe Quetelet

sostenne che le proprietà umane, sia individuali sia aggregate, presentavano sempre, così come gli “errori di osservazione”, una distribuzione campanulare, anzi una particolare forma di questa, dotata di utili proprietà matematiche, detta “curva normale” — tesi peraltro in seguito contestata, con dati più esaurienti e criteri più rigorosi, da vari statistici, fra i quali Pearson e Gini.

In varie redazioni successive del suo *Sur l'homme*, ribattezzato nella versione definitiva *Physique sociale* (1869), Quetelet affermò che la natura mirava sempre a un obiettivo (la moda di ciascuna distribuzione), mancandola spesso di poco, raramente di molto. Questa “legge degli errori” governava quindi non solo la distribuzione degli errori di osservazione, ma anche quella degli stessi stati effettivi su ogni proprietà.

Nel secondo volume della *Physique*, Quetelet aggiunse che anche la società, come la natura, mirava a produrre un *homme moyen*, e che solo la difficoltà di misurare alcune qualità intellettuali e morali impediva di mostrare come la “legge degli errori” valesse per ogni genere concepibile di fenomeno. Difficoltà che peraltro non spaventarono il biologo e statistico inglese Francis Galton, che esibì distribuzioni normali di “abilità naturale”, di “valore civico” e di “genio” (1869), e in seguito (1909) le sovrappose alla distribuzione delle categorie economiche che Booth aveva riscontato nel suo *social survey* della popolazione londinese (1891-1903).

Galton ebbe almeno il merito di evidenziare un'antinomia implicita nella santificazione dell'uomo medio: dato che studiava di preferenza qualità intellettuali ed era un accanito fautore del miglioramento eugenetico dell'umanità, gli riusciva più difficile che a Quetelet sostenere che la natura mirasse idealmente a produrre un uomo di qualità medie anziché eccellenti. Per questo motivo spostò l'accento dalla media all'intera distribuzione, sottolineò l'importanza di valori distribuzionali come i quartili, e introdusse gradatamente nei suoi scritti le espressioni ‘curva normale’ (al posto di ‘legge degli errori’) e *standard deviation* (al posto di *probable error*), aprendo la strada a un'impostazione meno determinista della statistica (Mackenzie 1981, 11-59).

E' interessante notare che, mentre Durkheim ereditò alcuni concetti e impostazioni di Quetelet, uno dei durkheimiani più dotati di autonomia e spirito critico attaccò frontalmente il mito dell'uomo medio, e l'analisi di distribuzioni semplicemente monovariate su cui era basato, con considerazioni assai moderne sull'interconnessione sincronica e diacronica fra le proprietà, e quindi sull'assurdità dell'andare in cerca di leggi probabilistiche monovariate (Halbwachs 1912). Già in precedenza alcuni matematici avevano criticato il trasferimento di concetti ed elaborazioni gaussiane in settori ove ne mancavano i presupposti (Bertrand 1889). Ma il prestigio di Quetelet è rimasto praticamente immutato fra gli statistici (si veda ad es. il classico manuale di Jones 1921) fino a tempi recentissimi, malgrado autorità del calibro di Keynes abbiano rilevato la sua propensione a generalizzare sulla base di dati insufficienti e qualitativamente inadeguati, a proclamare “leggi” ad ogni pie' sospinto e ad estenderne la portata con analogie discutibili (Keynes 1921/1973, 367, 429 e *passim*).

1.3. Nelle scienze sociali, il concetto di errore di osservazione elaborato dagli astronomi, e le proprietà della relativa distribuzione, scoperte e formulate in termini matematici da Gauss, fecero quindi un indebito ingresso sul piano ontologico assai prima di diffondersi come strumenti gnoseologici. A quanto ci consta, fu lo psicologo tedesco Hermann Ebbinghaus ad utilizzare per primo (1885), nelle sue ricerche sulla memoria, la distribuzione di osservazioni ripetute al fine di stimarne la precisione. Con lo sviluppo della ricerca empirica in psicologia sulle due sponde dell'Atlantico, la famiglia di concetti errore/precisione/accuratezza, con le relative procedure di accertamento, prese piede fra gli psicologi, e quindi fra i sociologi. Nelle scienze sociali anglo-sassoni diventò centrale e si stabilizzò il termine *reliability* (affidabilità), mentre in Francia *fiabilité* coesisteva con *fidélité*.

La concezione prevalente dell'attendibilità ricalcava fedelmente quella elaborata dagli astronomi: per il sociologo operazionista Stuart Dodd, “l'attendibilità può esser descritta come un indice che misura il grado di accordo (...) fra osservazioni ripetute dello stesso fenomeno” (1942, 484). Per lo psicologo sociale William Scott, “la precisione con cui le posizioni dei soggetti sono determinate può in linea di principio essere stabilita mediante misurazioni ripetute. Nella misura in cui si ottengono punteggi identici, le posizioni dei soggetti sono precise, e lo strumento è attendibile” (1968, 256). Malgrado il concetto di attendibilità abbia subito importanti modifiche per le ragioni tecniche che vedremo, definizioni del termine *reliability* basate su rilevazioni ripetute nel tempo sono state e sono tuttora

offerte da metodologi (Jahoda, Deutsch e Cook 1951, 100; Kaplan 1964, 195; Siegel e Hodge 1968, 56), sociologi (Zetterberg 1954, 124), psicologi (Fleshman 1968, 372; Upshaw 1968, 65; Ingram 1977, 16; Clark (1977, 109), politologi (Frey 1970, 244; Macrae 1970, 275; Hill 1980, 407).

Anche gli esempi sono sovente presi dalla realtà fisica: una sbarra che misura più volte la lunghezza di un tavolo (Spector 1981, 13-14) o la statura di una persona (Carmines e Zeller 1979, 13); una bilancia che pesa un oggetto (Phillips 1966, 85), e così via. Com'è abituale nelle scienze sociali, concetti e relative procedure sono stati mutuati dalle scienze fisiche senza preoccuparsi della loro trasferibilità, e più specificamente di quanto gli assunti incorporati in tali concetti e procedure restino plausibili allorché riferiti alla realtà umana e sociale.

In particolare, facendo riferimento all'elenco del par. 1.1, l'assunto (d) — irrilevanza delle differenze individuali — non è sostenibile: se lo fosse, potremmo far ricerca su un vicino di casa e attribuire quello che troviamo a tutto il genere umano, così come il fisico attribuisce a tutti gli atomi di azoto dell'universo quello che riscontra relativamente a un qualsiasi, magari piccolissimo, insieme di atomi di azoto che si trova sottomano. Questa, sia detto per inciso, è la differenza epistemologica cruciale fra le scienze fisiche e le scienze umane, e la sua portata va ben oltre la questione in esame.

Anche l'assunto (a) — lo stato dell'oggetto non cambia spontaneamente nell'intervallo fra le rilevazioni — ha sollevato perplessità. Cronbach lo definisce “altamente improbabile” (1947, 4); secondo Deutscher, “l'assunto che il pensiero e il comportamento umano siano statici... è semplicemente antitetico alle scienze sociali” (1966b, 240). E l'assunto (b) — lo stato dell'oggetto non è alterato dal processo di osservazione — è scarsamente difendibile per molte delle correnti procedure di rilevazione, e specialmente per quelle cui, per la natura del loro oggetto, devono ricorrere gli psicologi, che per primi nelle scienze umane adottarono l'idea degli astronomi di stimare l'accuratezza di un'osservazione riprendendola.

Sviluppando l'idea di Ebbinghaus, nel primo quarto di questo secolo gli psicologi anglosassoni hanno infatti codificato una procedura che “misura” l'attendibilità di uno strumento mediante la ripetizione delle osservazioni. Questa procedura, detta *test-retest*, prevede le seguenti fasi (Guttman 1946; Carmines e Zeller 1979, 37):

- (i) un certo strumento (di solito un test con più *items*) è somministrato al tempo t , nel luogo u e nei modi m a un insieme di soggetti (di solito, gli studenti di un corso universitario di psicologia);
- (ii) le risposte sono codificate e così trasformate in un vettore di cifre;
- (iii) le fasi (i) e (ii) sono ripetute al tempo t' ma con gli stessi soggetti e — nei limiti del possibile — nello stesso luogo e stessi modi;
- (iv) si calcola un coefficiente di correlazione fra i due vettori di cifre ottenute nelle fasi (ii) e (iii);
- (v) tale coefficiente (o il suo quadrato, secondo McKennell 1970, 228-9) viene denominato *test-retest reliability coefficient* o *stability coefficient* dello strumento.

Per quanto strettamente ispirata al concetto di accuratezza delle scienze fisiche, la procedura *test-retest* presenta una variante cospicua, rilevata a nostra conoscenza solo da Carmines e Zeller (1979, 31).

Nelle scienze fisiche, in conseguenza dell'assunto (d), l'accuratezza di una coppia strumento-osservatore viene misurata sulla base di osservazioni relative allo stesso oggetto: i dati sono organizzati in un vettore i cui elementi riportano risultati di osservazioni effettuate in momenti diversi, e l'accuratezza è l'inverso della loro varianza; poiché la varianza di una distribuzione con pochissimi valori sarebbe poco robusta, astronomi e fisici hanno bisogno di un certo numero di osservazioni.

Ma l'assunto (d) — irrilevanza delle differenze interindividuali — non può essere sostenuto nelle scienze umane; le rilevazioni per il calcolo dell'attendibilità *test-retest* sono quindi eseguite su una pluralità di soggetti (anche se di solito si tratta di una pluralità assai moderata: gli studenti del corso di psicologia di chi ha ideato il test o ne controlla l'attendibilità); ogni rilevazione produce già un vettore (n soggetti, un solo momento); più rilevazioni producono una matrice (n soggetti, r momenti). Di conseguenza, mentre nelle scienze fisiche l'accuratezza di una coppia strumento-osservatore è una grandezza scalare, l'attendibilità calcolata con la procedura *test-retest* dovrebbe essere una grandezza vettoriale (una cifra per ciascun soggetto).

Ma è qui che l'armamentario statistico viene in soccorso degli psicologi: tra due vettori della stessa matrice, i cui elementi siano numeri cardinali, si può calcolare un coefficiente di correlazione, che è uno scalare. Gli psicometrici chiamano *test-retest reliability coefficient* il coefficiente di correlazione fra i due vettori di punteggi prodotti da due somministrazioni dello stesso test agli stessi soggetti. In tal modo essi eliminano le conseguenze negative della specificità epistemologica del loro oggetto di studio, perché riescono a condensare la "misura" della precisione del loro strumento in una sola cifra pur non potendo negare la rilevanza delle differenze interindividuali. Non solo; essi riescono anche ad estrarre dal suddetto limite epistemologico un vantaggio pratico: dato che ogni somministrazione dello strumento produce già un vettore di cifre, due sole somministrazioni bastano loro per calcolare un coefficiente di attendibilità — mentre, come si diceva, ad astronomi e fisici occorrono molte osservazioni per calcolare una varianza su un numero accettabile di cifre.

In tal modo, sfruttati i vantaggi computazionali derivanti dalle differenze inter-individuali, si nega ogni altro effetto e rilevanza a tali differenze. Il coefficiente di attendibilità calcolato nel modo visto viene infatti riferito non solo a tutti i membri del gruppo cui è stato somministrato il test, ma a tutti gli individui possibili in tutte le situazioni possibili, malgrado qualche raro manuale avverta che "può venir applicato soltanto a campioni simili a quello sulla base del quale è stato calcolato" (Anastasi 1954/1981, 187). Il coefficiente invece viene considerato un attributo permanente e definitivo del test X, e lo accompagna sempre, dovunque e qualunque sia la popolazione di soggetti cui capiti di somministrarlo.

Esistono infatti raccolte di *tests* o "scale" di uso corrente nelle scienze sociali, ciascuno accompagnato dal suo coefficiente di attendibilità (di tipo *test-retest* o di altri tipi che vedremo in seguito) calcolato una volta per tutte; solo in casi assai rari si riporta il risultato di un successivo controllo, che mostra discrepanze anche forti con il valore del coefficiente calcolato la prima volta (in Shaw e Wright, 1967, si riportano ad esempio casi in cui da .98 si passa a .71).

Ricostruiamo il retroterra epistemologico dell'accertamento dell'attendibilità mediante la procedura *test-retest*.

Degli assunti presentati nel par. 1.1 e relativi alle osservazioni delle scienze fisiche, i primi tre (esiste uno stato effettivo dell'oggetto sulla proprietà in questione, che non cambia spontaneamente nell'intervallo fra le osservazioni e non è da esse modificato) restano immutati; l'assunto (d) è invece insostenibile, ed è così modificato e integrato:

(d') data la rilevanza delle differenze inter-individuali, l'attendibilità di un strumento non può essere stabilita osservando un solo soggetto. E' necessaria una pluralità di soggetti (che peraltro può essere costituita senza rispettare alcun criterio; si può quindi ricorrere, per ragioni di comodità, agli studenti del proprio corso di psicologia);

(e) l'attendibilità di uno strumento è la stessa per tutti i soggetti membri di questa pluralità, e può quindi essere espressa da un solo coefficiente;

(f) il coefficiente è un attributo intrinseco dello strumento; per poterlo calcolare, cioè portare alla luce, si deve far somministrare il test dall'operatore X a un gruppo Y di soggetti nella situazione *al momento nel luogo*. Ma, essendo un attributo intrinseco dello strumento, è naturale che il coefficiente resti lo stesso se il test è somministrato da qualunque altro operatore a qualunque altro gruppo di soggetti in qualunque altra situazione, tempo e luogo (cfr. i rilievi di Tukey 1954).

L'assunto (e) ha natura definitoria e quindi non ha senso sottoporlo a controllo empirico; gli assunti (a) (b) (c) (d) (d') sono controllabili con grave difficoltà. Sarebbe invece estremamente facile controllare empiricamente l'assunto (f). Il fatto che tali controlli praticamente non avvengano conferma appunto la sua natura di assunto. Insieme all'assunto (e), esso ha la funzione di ripristinare una situazione tollerabilmente simile a quella delle scienze fisiche, neutralizzando anche a livello epistemologico — così come il coefficiente di correlazione permette di neutralizzarla a livello tecnico — l'impossibilità di dichiarare esplicitamente irrilevanti le differenze inter-individuali.

In quest'ansia di omologazione, gli psicometrici sono andati oltre i fisici, che non calcolano la precisione di uno strumento indipendentemente dal suo operatore, essendo consapevoli della differenza epistemologica fra i loro oggetti, per i quali vale l'assunto (d), e gli osservatori.

1.4. Pur essendo il più vulnerabile, l'assunto (f) non è stato il più criticato dagli psicologi, piuttosto inclini a generalizzare con leggerezza risultanze relative a specifici campioni, forse in base a un assunto di omogeneità psichica del genere umano.

Sono stati invece diffusamente criticati gli assunti (a) e (b) che, concernendo il mutamento diacronico, sono specifici al procedimento *test-retest*. Si è osservato che, contrariamente all'assunto (a), lo stato di un individuo su una proprietà può cambiare spontaneamente nell'intervallo fra le rilevazioni, e sono stati adottati numerosi risultati di *panels* a prova del fatto che l'entità del cambiamento è associata alla lunghezza dell'intervallo (Rozelle e Campbell 1969; Converse 1970).

D'altra parte, più breve è tale intervallo, più alta è la probabilità che la prima rilevazione alteri lo stato osservato, contrariamente all'assunto (b). Se la proprietà rilevata è o implica qualche forma di abilità o di conoscenza specifica, la prima rilevazione può provocarne la crescita (*learning o practice effects*: Cronbach 1949, 619; Anastasi 1953, 190-1; Webb *et al.* 1966, 19; Siegel e Hodge 1968, 56). La prima rilevazione può aumentare l'interesse e/o mutare l'atteggiamento verso gli argomenti sottoposti (Campbell e Stanley 1963, 20-22; Bohrnstedt 1970, 86; Carmines e Zeller 1979, 39).

In generale, la reazione di un individuo a una seconda rilevazione sarà differente (minor timore e quindi maggiore sincerità; maggior controllo della situazione e quindi maggiore opportunismo; assuefazione e quindi noia, ostilità, etc. Vedi Scott 1968, 239), e questo potrà influenzare il suo stato osservato su una qualsiasi proprietà. Non si possono escludere cambiamenti di qualche rilevanza negli atteggiamenti e comportamenti di chi somministra il test (Webb *et al.* 1966, 22), e in altri aspetti della situazione, che sono innumerevoli (Berka 1983, 198) — né è facile stabilire quali di essi devono essere controllati per evitare variazioni (Kriz 1988, 70).

Tutti questi cambiamenti, naturali e indotti, provocano una riduzione del livello dell'attendibilità *test-retest*. Ma tale livello può anche risultare accresciuto artificialmente se il test riguarda atteggiamenti, e i soggetti ricordano le risposte date in occasione della prima rilevazione, e le ripetono per desiderio di apparire coerenti o per trarsi d'impaccio (Kuder e Richardson 1937, 151; Baughman 1958, 134; Ingram 1977, 17).

La raccomandazione di riportare sempre la durata dell'intervallo fra *test* e *test* (Cronbach 1947, 14-15; Anastasi 1954/1981, 165) accanto al valore del relativo coefficiente potrebbe, se fosse seguita, introdurre un mutamento di rilievo a condizione che il miglioramento spontaneo fosse una funzione lineare del tempo e che fosse il solo fattore perturbatore — due condizioni scarsamente plausibili nella maggior parte dei casi. La stessa Anastasi riconosce che il cambiamento prodotto dalla ripetizione del test è più rilevante del cambiamento spontaneo, salvo che per alcuni tests psicofisici (discriminazione sensoriale, motricità), e che quindi la tecnica *test-retest* “non è idonea per la grande maggioranza dei tests psicologici” (1954/1981, 165).

In effetti, la popolarità del procedimento si è andata riducendo, a causa delle critiche ora viste, e forse anche per banali ragioni pratiche: “La tecnica *test-retest* presenta un grande svantaggio pratico: è spesso praticamente impossibile ottenere il permesso di sottoporre due volte il test a un gruppo di studenti” (Ingram 1977, 17-18). Essa è stata addirittura definita “esageratamente dispendiosa” (Carmines e Zeller 1979, 79) — e dire che a noi sembra invece anche troppo economica a fronte della connessa pretesa di stabilire l'attendibilità di uno strumento una volta per sempre, e per tutte le combinazioni di soggetti e situazioni!

1.5. Riconoscendo il fatto che i loro oggetti di studio non restavano identici a se stessi nel tempo, e reagivano al fatto di essere osservati, gli psicometrici hanno per la seconda volta preso atto della differenza fra tali oggetti e quelli delle scienze fisiche.

Il primo riconoscimento, come si è visto, riguardava la rilevanza delle differenze inter-individuali, che costringeva a lavorare su una pluralità di soggetti, ma aveva anche permesso di operare due sole rilevazioni invece della molteplicità di osservazioni dei fisici. Anche in questo secondo caso, gli psicometrici sono riusciti a dare una qualche soluzione al problema ottenendo nel contempo una riduzione del carico di lavoro.

Invece di correlare i punteggi ottenuti dai vari soggetti sullo stesso test somministrato due volte, negli anni venti si è pensato di correlare i punteggi ottenuti su due tests differenti ma definiti “equivalenti”. E' nata così la tecnica nota come *parallel forms* o *equivalent forms* (Thurstone 1928, 552; Anastasi

1954/1981, 169; Fleishman 1968, 373). Il coefficiente così prodotto è stato battezzato *coefficient of equivalence* (American Psychological Association 1954, 29) e l'attendibilità è stata ridefinita come "il grado in cui strumenti di rilevazione dello stesso tipo producono gli stessi risultati" (D'Andrade 1974, 160; analogamente Ingram 1977, 16).

All'inizio degli anni trenta si è operata un'ulteriore semplificazione, dividendo un test in due parti di uguale lunghezza (spesso semplicemente col sommare separatamente i punteggi ottenuti dal soggetto sugli *items* in posizione pari e in posizione dispari nel test) e correlando i punteggi ottenuti dai soggetti nelle due metà così formate. Il coefficiente così ottenuto è stato denominato *split half* (o anche *odd-even reliability coefficient* (Brownell 1933; Kuder e Richardson 1937; Rulon 1939) oppure *internal consistency coefficient* (Zubin 1934; Sletto 1936).

A questo punto — anni venti-trenta — il concetto di attendibilità ha già subito un'importante modifica: non è più centrale la ripetizione delle osservazioni nel tempo, ma la produzione di due vettori di punteggi da poter correlare. Se con la tecnica *test-retest* i vettori sono prodotti tenendo costante strumento e operatore e mutando il momento della somministrazione, con le tecniche *parallel forms* e *split half* si tiene praticamente costante il momento e (per quanto possibile) l'operatore e si fa variare lo strumento. Data questa struttura logica, basata su due elementi costanti e uno variabile, c'è una terza possibilità: tenere costanti strumento e momento e far variare l'operatore.

E infatti non tardano molto a comparire, non tanto in psicologia quanto nelle discipline in cui si codificano testi scritti, tecniche basate sulla correlazione fra punteggi assegnati da codificatori diversi forniti delle stesse istruzioni (che rivestono il ruolo dello strumento) che operano parallelamente, cioè in costanza, vera o presunta, di tempo. Si forma così un ulteriore concetto di attendibilità come accordo di osservatori diversi (Hempel 1961, 142; Kaplan 1964, 195), che viene designato con una miriade di espressioni: *interindividual reliability* (Zetterberg 1954, 124), *rater reliability* (Bendig 1954), *multi-judge reliability* (Cartwright 1956), *interindividual constancy* (Galtung 1959, 113), *intersubjectivity* (Galtung 1967, 28), *coder o intercoder reliability* (Kalton e Stowell 1979; Singer 1982, 194).

1.6. Un ulteriore slittamento nel concetto psicometrico di attendibilità si produce peraltro verso la fine degli anni trenta, quando due membri del laboratorio di psicomètria di Thurstone propongono una formula per calcolare un coefficiente di attendibilità sulla base di una matrice delle correlazioni fra vettori costituiti dai punteggi dei soggetti su ciascuno degli *items* di un solo test (Kuder e Richardson 1937). Tecnicamente, questo costituisce un deciso progresso rispetto ai coefficienti *test-retest*, *split half* e *parallel forms*, tutti basati su una sola correlazione fra due vettori costituiti dai punteggi individuali su un test (o metà test) globalmente considerato, e rispecchia la crescente familiarità degli psicomètrici con l'algebra matriciale.

In questo modo, il concetto di attendibilità si lega ancor più di prima all'elaborazione delle risposte a una batteria di *items*, e diventa tecnicamente inapplicabile ad ogni altro strumento. Questa restrizione è naturale per gli psicologi di laboratorio, il cui contatto con la realtà empirica avviene appunto quasi esclusivamente attraverso tests in forma di batterie di *items*. Meno naturale è che nelle altre scienze sociali ove si usa una gamma più vasta di strumenti, il concetto/termine 'attendibilità' venga importato dagli psicomètrici senza che la sua stretta limitazione a un particolare strumento sia in alcun modo rilevata.

Dal punto di vista metodologico, inoltre, si può osservare che utilizzare come elementi del calcolo i punteggi su singoli *items* anziché i punteggi su un intero test significa lasciare che le componenti erratiche che pesano su ciascuna singola risposta o prestazione manifestino interamente i loro effetti, mentre tradizionalmente si ritiene, o quanto meno si sostiene (ma qualcuno lo mette in dubbio: ad esempio Dingle 1950, 20), che tali effetti tendano ad elidersi quando i punteggi relativi ai singoli *items* sono sommati, o comunque combinati, in un punteggio globale.

Un secondo rilievo metodologico, più sicuramente fondato, è che la formula di Kuder e Richardson considera solo *items* dicotomici, aventi cioè il formato più sensibile agli errori (in una dicotomia, ogni errore trasforma uno stato nel suo opposto) e alle distorsioni che distribuzioni non equilibrate di risposte introducono nel coefficiente di correlazione (Gangemi 1977).

Questo limite è stato rimosso nel dopoguerra, quando Cronbach ha proposto *alfa*, coefficiente basato sull'elaborazione di matrici di punteggi su *items* di qualsiasi formato. L'evoluzione tecnica è poi proseguita con cadenza rapida, come del resto in tutti i settori dell'analisi dei dati: McKennell (1970) ha proposto una formula semplificata che approssimava *alfa*. Raju (1977) ha proposto *beta*, coefficiente che applica i principi di *alfa* a una situazione in cui un test è diviso in varie parti con un numero diseguale di *items*. Heise e Bohrnstedt (1970) hanno mostrato come attraverso l'analisi fattoriale di una matrice di punteggi su singoli *items* si possa produrre un coefficiente, *omega*, che stima con una sola cifra l'attendibilità globale di tutti i fattori che è possibile estrarre dalla matrice. Armor (1974) ha utilizzato l'analisi delle componenti per produrre un altro coefficiente, *theta*, in due versioni: una per il caso in cui si estragga una sola componente rilevante, un'altra per il caso in cui tali componenti siano più di una.

2. L'attendibilità e la cintura protettiva

2.1. Le varie formule dei numerosi coefficienti di attendibilità introdotti in questo dopoguerra si basano su assunti specifici che sono analizzati criticamente da Giampaglia (1986; più piatta la rassegna di Carmines e Zeller 1979, 43-62), al quale si rinvia. Esse peraltro non introducono alcuna innovazione nel concetto di attendibilità legato in via esclusiva alla manipolazione matematica di una matrice di cifre; l'ultimo passo in questa direzione era già stato compiuto da Kuder e dalla Richardson nel 1937. Già da tempo il problema dell'effettiva corrispondenza delle cifre manipolate agli stati dei soggetti sulle proprietà era stato perso di vista; l'attenzione esclusiva per tecniche via via più sofisticate lo rendeva solo sempre più remoto. Come dichiarò Zetterberg, "l'accumularsi dell'esperienza di questi studi ci ha insegnato ad avere fiducia in certe tecniche di misurazione... Gli studiosi che usano queste tecniche, pertanto, non si sentono più tenuti ad affrontare il lungo e noioso processo di controllare la loro attendibilità" (1954, 124).

Non c'è dubbio che, una volta riempita una matrice con i punteggi di un gruppo di soggetti sugli *items* di un test, far girare il programma che calcola *alfa*, *omega* o *theta*, e magari tutti e tre, è incomparabilmente più rapido e meno faticoso che qualsiasi immaginabile procedura per controllare, in qualche modo e anche in minima parte, la rispondenza di quei punteggi agli stati effettivi dei soggetti sulle proprietà investigate. Offre inoltre il vantaggio di chiudere la questione con una cifra univoca, "scientifica" (prodotta dal calcolatore!) e asettica, risparmiando ogni minuto confronto con la realtà, dai modi sempre opinabili e dagli esiti dubbiosi e generatori di dubbi.

Per l'uomo della strada, il naturale obiettivo di un controllo di attendibilità dovrebbe essere l'individuazione e correzione, per quanto possibile, dei dati che non rispecchiano fedelmente gli stati che rappresentano nella matrice. Per l'uomo in camice bianco, invece, quello che conta non è accrescere il grado di corrispondenza dei propri dati con gli stati effettivi, ma unicamente accrescere la cifra che, per convenzione, tale corrispondenza "misura". In altre parole, anche nel caso dell'attendibilità si produce quella feticizzazione dello strumento che a nostro avviso sempre più caratterizza questo primo secolo di approcci degli scienziati sociali con matematica e derivati.

Il feticismo della cifra si manifesta già negli anni trenta, allorché, malgrado i difetti epistemologici e il maggior impegno richiesto dalla procedura *test-retest*, molti psicometrici esitano ad adottare le procedure *split half* e *parallel forms* perché ciò comporta dimezzare o comunque ridurre il numero di *items* nel test, il che — a parità di altre condizioni — automaticamente abbassa la cifra del coefficiente (Anastasi 1954/1981, 171-2; Hicks 1967; Davis 1977). Per rimediare, viene riesumata una formula elaborata indipendentemente da Spearman e da Brown nel 1910, che — sulla base di un discutibilissimo assunto di uguale effetto di ciascun *items* sull'attendibilità — calcola le variazioni nel coefficiente di attendibilità di un test come funzione (esclusiva) dell'aumento o della riduzione nel numero dei suoi *items*.

Questa formula (battezzata, con ironia non sappiamo quanto consapevole, *prophecy formula* — la sua base logica è infatti nebulosa come quella di una profezia) viene usata come un talismano da tutti gli psicometrici per "correggere" (cioè elevare) il valore dei coefficienti di attendibilità ottenuti con

procedure *split half* o *parallel forms* applicate ai tests da essi proposti, elevando probabilmente di altrettanto la loro auto-stima scientifica e il loro successo (anche commerciale, visto che in quei decenni si va creando in America un florido mercato dei tests, i cui acquirenti sono *colleges* e altre istituzioni educative, *consultants* privati, corsi di addestramento di ogni genere).

Dalla metà degli anni trenta in poi, anche in tempi di relativa desuetudine delle tecniche *split half* di fronte alle nuove tecniche di stima matriciale dell'attendibilità, continuano a comparire nuove formule che "correggono" la riduzione dei coefficienti in modo più semplice, o più efficace, o con assunti più plausibili rispetto alle formule di Spearman e Brown (Horst 1936; Rulon 1939; Callender e Osburn 1977).

Nella psicologia di laboratorio, "i tests oggettivi sono preferiti ai tests non direttivi" cioè ai proiettivi e simili (D. C. A. 1974, 735) perché "i punteggi ottenuti dei tests oggettivi tendono ad essere più attendibili <cioè a produrre coefficienti di attendibilità più alti> dei punteggi ottenuti dai tests non oggettivi" (Ebel 1968, 34).

Come si è anticipato, più numerosi sono gli *items* nel test, più alto sarà, *ceteris paribus*, il coefficiente di attendibilità. Di conseguenza, si tende a produrre tests di lunghezza spropositata. "I soggetti cui questi tests sono somministrati sono di solito tratti da popolazioni 'prigioniere' — scolari, studenti, militari, malati di mente, candidati a un'assunzione — che si sottopongono senza protestare a batterie di tests dell'enorme lunghezza necessaria per raggiungere i coefficienti di attendibilità del livello richiesto" (McKennell 1970, 236).

Per ingrassare il fatal coefficiente, si trascurano gli effetti ben noti (Festinger 1947; Converse 1970, 178-9; Noelle-Neumann 1970, 194; Ferrand e Martel 1986) della fatica e della noia sulle prestazioni dei soggetti e sulla corrispondenza di quello che rispondono alle loro opinioni/atteggiamenti effettivi (da non confondere con l'attendibilità calcolata mediante le formule: lo scarto fra i due concetti non potrebbe essere più evidente); si calpesta anche il diritto morale di ciascun soggetto ad essere "osservato" con uno strumento concepito al fine di valutare le sue capacità, opinioni, atteggiamenti, valori anziché al fine di gonfiare il coefficiente di attendibilità dello strumento stesso.

2.2. Altre ragioni allontanano il concetto di attendibilità formulato dagli psicometrici (e ripreso da sociologi e politologi) dal corrispondente concetto del senso comune. "Il coefficiente di attendibilità di una scala ci dice assai poco delle effettive qualità della scala, perché un alto valore del coefficiente può essere interamente dovuto a fattori di distorsione" (Bartlett *et al.* 1960, 703). Un esempio di questo meccanismo è fornito, senza alcun segno di ironia, da un manualetto assai diffuso e interamente dedicato al problema dell'attendibilità: "Supponiamo che una barra misuri 40 pollici invece degli abituali 36. Di conseguenza, questa barra sistematicamente sottostima la lunghezza dell'oggetto misurato di 4 pollici per ogni iarda... Questo errore non riduce affatto l'attendibilità della barra dato che non produce risultati incongruenti se la misurazione è ripetuta" (Carmines e Zeller 1979, 13). Nelle scienze sociali, un'attendibilità del tutto artificiale può essere prodotta, oltre che allungando i tests, anche riducendo l'estensione del dominio semantico coperto dagli *items* (McKennell 1977, 211) e omogeneizzandone la forma strutturale, i termini impiegati e così via (Hyman *et al.* 1954, 30; McKennell 1970, 234; vedi il concetto di 'rumore linguistico' introdotto da Ricolfi 1987). Inoltre, come deplora Galtung (1967, 28), spesso si riduce il numero di risposte "chiuse" previste dalla struttura della batteria per ciascun *item*, perché questo tende a produrre coefficienti di attendibilità più elevati (specialmente con la procedura *test-retest*), impedendo che le risposte si distribuiscano su categorie più dettagliate — che tuttavia rifletterebero più fedelmente le effettive diversità delle opinioni e degli atteggiamenti. Al limite, una batteria cui tutti i soggetti reagiscono con un *response set* otterrà il massimo coefficiente di attendibilità possibile qualunque si applichi delle formule descritte nei paragrafi precedenti. E si potrebbe allungare a piacere l'elenco degli espedienti e degli artifici tecnici che possono accrescere il valore di un coefficiente di attendibilità.

2.3. Un'altra conseguenza negativa della riduzione del concetto di attendibilità al risultato della manipolazione matematica di due o più vettori è che si è gradatamente perduto il senso della distinzione fra attendibilità e validità.

La validità è universalmente considerata una relazione fra il concetto che ha in mente il ricercatore e il risultato della procedura di rilevazione. Decine e decine di manuali e di articoli riproducono, senza alcuna o con lievissime variazioni, la domanda cui il concetto di validità è inteso rispondere: l'indicatore / la misura / la scala / il test indica / misura / rileva ciò che si crede che indichi / misuri / rilevi? (vedi ad es. Dodd 1942, 488; Zetterberg 1954, 114; Ebel 1968, 39; Fleishman 1968, 373; Frey 1970, 244; Smelser 1976, 185; Nowak 1976, 58; Clark 1977, 109; Ingram 1977, 18; Carmines e Zeller 1979, 12; Singer 1982, 192).

Tuttavia, al fine di evitare interventi "soggettivi" (cioè il giudizio del ricercatore e della comunità scientifica sul grado di corrispondenza semantica fra concetto e indicatore) ogni volta che si è potuto si sono sostituite le forme *soft* di validità (cioè la *face*, la *content*, la *construct*) con quelle *hard* (*criterion*, *predictive*, *concurrent*, *convergent and discriminant*), cioè basate sulla correlazione di due o più vettori di cifre (sulle varie forme di validità, Ammassari 1984; sulla reificazione del concetto di validità, Marradi 1989). La sostanziale identità delle procedure per calcolare i coefficienti di attendibilità e quelli di validità ha inevitabilmente prodotto un grave annebbiamento della radicale differenza fra i due concetti.

Likert propose infatti (1932, 93), sotto il nome di *internal consistency*, una tecnica di misurazione dell'attendibilità che riproduce esattamente la tecnica detta *known groups*, abitualmente considerata una forma di accertamento della validità (Thurstone e Chave 1929; Kendall e Lazarsfeld 1950; Scott 1929, 253; Frey 1970, 248-9). Sin dagli anni trenta, le oscillazioni terminologiche sono frequenti (Horst 1934; Horst 1936), e alcuni usano tranquillamente un termine al posto dell'altro (Gulliksen 1936; Parry e Crossly 1950; Suchman e Phillips 1958; Schuman 1966; Traugott e Katosh 1979). Siegel e Hodge sono quasi eufemistici quando osservano (1968, 55) che "questioni di validità e questioni di attendibilità sono confuse in molta letteratura delle scienze sociali".

In effetti, come hanno sottolineato Campbell e Fiske (1959, 83), se ci si limita a correlare in entrambi i casi due o più vettori di cifre, attendibilità e validità sono i due estremi di un continuum:

"I concetti di attendibilità e validità richiedono entrambi che si mostri un accordo fra misure. Le varie forme di validità condividono — a differenza dell'attendibilità — la caratteristica che questo accordo consiste nella convergenza di approcci indipendenti fra loro. Naturalmente, l'indipendenza è una questione di gradi, e in questo senso attendibilità e validità possono essere viste come regioni di un continuum. L'attendibilità è l'accordo fra due tentativi di misurare lo stesso tratto con tecniche quanto più simili fra loro. La validità consiste nell'accordo fra due tentativi di misurare lo stesso tratto con tecniche quanto più differenti fra loro. Uno *split-half reliability coefficient* è leggermente più simile a un coefficiente di validità di quanto lo sia un *test-retest reliability coefficient*, perché gli *items* non sono identici. Una correlazione fra sub-tests diversi è probabilmente ancora una misura di attendibilità, ma è ancora più vicina alla regione della validità" (1959, 83).

L'idea che l'attendibilità consista nell'accordo di vettori che rappresentano "tecniche quanto più simili", mentre la validità consiste nell'accordo di vettori che rappresentano "tecniche quanto più differenti" è stata spesso ripresa (Siegel e Hodge 1968; Bentler 1972, 343). Una variante, che accenna a fuoriuscire dal quadro della matrice dei dati per porsi sul piano epistemologico, è l'idea che "ciò che meglio contraddistingue le questioni di attendibilità è l'assunto che i punteggi siano determinati solo dai concetti che si vuole rilevare, e da errori accidentali. Per contro, questioni di validità sorgono quando si ritiene che altri fattori, oltre a quel concetto e agli errori accidentali, influenzino i punteggi" (Althausen e Heberlein 1970. Vedi anche Werts e Linn 1970). Berka sostiene qualcosa di non dissimile quando riduce la differenza fra attendibilità e validità a quella fra errori accidentali ed errori sistematici (1983, 203).

Altri autori si limitano a constatare che "attendibilità e validità sono inestricabilmente intrecciate" (Davies 1977, 51) o "strettamente collegate" (Carmines e Zeller 1979, 34). Si stabiliscono semplici relazioni matematiche tra le formule dei coefficienti di validità e di quelli di attendibilità (si veda per esempio Cronbach 1949, 171); per Lord e Novick (1968) la relazione è semplicissima: il primo è la radice quadrata del secondo. Si concepiscono dei modelli *path* "per trattare simultaneamente questioni

di attendibilità e di validità” oltre che le relazioni causali fra le variabili del modello (Siegel e Hodge 1968, 56; vedi anche Bohrnstedt 1969; Werts e Linn 1970; Allen 1974, 52).

2.4. Le uniche proteste contro questa commistione fra i concetti di attendibilità e validità sono per ora venute dagli interazionisti simbolici (Blumer 1954, 7; Deutscher 1966a, 34), coerentemente con le loro critiche a una forma di ricerca empirica che si esaurisce nell’elaborazione matematica della matrice dei dati. Ma la differenza fra il concetto di attendibilità degli psicometrici e il corrispondente concetto dell’uomo della strada non è sfuggita anche a qualche altro studioso pur tutt’altro che ostile alle elaborazioni matematiche:

“L’uso del termine ‘attendibilità’ nel suo significato tecnico di congruenza fra due o più vettori di risposte è inopportuno. Il significato non tecnico di ‘attendibile’ è ‘degnò di fiducia’. Gli scienziati sociali possono essere in grado di ripetere con esattezza la definizione tecnica del termine, ma temo che tuttavia avranno più fiducia di quanto dovrebbero in un test di attendibilità semplicemente per l’effetto di alone del significato ordinario del termine. Sarebbe assai meglio chiamarli ‘tests di congruenza’, cosa che sono effettivamente, piuttosto che tests di attendibilità, cosa che nel senso ordinario del termine non sono affatto” (Naroll 1968, 265-6).

Entrato ormai in maniera stabile nelle abitudini di psicologi, sociologi e politologi, il concetto di attendibilità e le tecniche relative sbarrano la strada allo sviluppo di concetti e tecniche più adeguati alla situazione epistemologica delle scienze sociali:

“Dato che non si disponeva di altre informazioni sul passato dei soggetti e sul loro sviluppo, l’attendibilità dovette essere stabilita mediante manipolazione statistica dei dati degli stessi tests. Sembra che questi tests, immaginati per ovviare all’assenza di altre forme di controllo dell’attendibilità, ora funzionino come barriere contro la diffusione di ogni altra procedura... Le tecniche universalmente in uso... danno informazioni solo sul complesso degli individui studiati, non su ogni singolo soggetto” (Frank 1939, 399-400).

Già prima della guerra, alla fine di un decennio che aveva visto un vigoroso sviluppo delle tecniche basate sulla correlazione fra vettori, uno psicologo clinico come Frank, abituato ad usare tests proiettivi e altri strumenti di analisi del profondo, si rendeva conto dell’inammissibilità di un concetto vettoriale di attendibilità, e della necessità di valutare separatamente la fedeltà di ciascun atto di rilevazione.

Solleva peraltro qualche riserva la sua opinione che gli psicometrici abbiano fatto ricorso alla manipolazione statistica “per ovviare all’assenza di altre forme di controllo dell’attendibilità”. Come vedremo nella sezione 3, tali forme esistono, e non mancano i ricercatori che ad esse fanno ricorso, occasionalmente o anche abitualmente. Ma farvi ricorso esige un grande investimento di tempo, di pazienza, e talvolta anche di creatività da parte del ricercatore, e — soprattutto — non è compatibile con l’immagine dello scienziato che, in camice bianco nella sua torre d’avorio, manipola ipotesi e teorie con l’ausilio della logica formale, e/o matrici di dati con l’ausilio del calcolatore.

2.5. Secondo il modello oggi dominante nelle scienze sociali, la ricerca conserva uno e un solo punto di contatto con il suo oggetto, al momento della raccolta dei dati. Ma anche tale raccolta è effettuata con operazioni quanto più standardizzate e meccaniche possibile, al fine di evitare la “soggettività” — e con essa ogni occasione di contatto diretto del ricercatore con gli individui studiati. Infatti, queste operazioni sono affidate del ricercatore a una vera catena di collaboratori: esperti di campionamento, agenzie di sondaggi, enti pubblici di raccolta di informazioni, somministratori di tests, intervistatori, operatori censuali, contatori di parole e altri analisti del contenuto, e così via. Liberatosi così di tutti i compiti ancillari, e con essi — a quanto pare senza rimpianti — della sostanza della ricerca, il sociologo attende alla teorizzazione — o, più banalmente, alla pubblicazione. Per quanto le negligenze, gli errori, le superficialità, le piccole e grandi astuzie di ciascuno degli anelli di questa catena di montaggio siano state denunciate più volte (sulla totale inaffidabilità delle

“statistiche ufficiali” vedi Merritt e Rokkan 1966; Denzin 1969; La Palombara 1970; Taylor e Hudson 1972; sui trucchi di intervistatori e codificatori il classico è Roth 1965; sul modo in cui sono effettivamente estratti campioni “casuali” di comuni e di individui e dai loro collaboratori, vedi Marradi 1988, cap. 4), gran parte dei direttori di ricerca non esercitano alcun controllo sulle operazioni di raccolta, ufficialmente per scrupolo di neutralità, in realtà per non restare coinvolti in faccende che richiedono tempo, attenzione e impegno. Così le innumerevoli decisioni, minute e non, che sono necessarie ad ogni passo della ricerca, sono prese da esecutori che nella maggior parte dei casi non hanno alcun interesse per essa, e quindi seguono il criterio del minimo sforzo. Ognuna di queste decisioni, che collettivamente determinano il risultato della ricerca, viene presa di volta in volta da un esecutore, senza alcun coordinamento con gli altri che eseguono lo stesso compito, e senza canali di comunicazione da e per il direttore della ricerca (vedi la denuncia di Boccuzzi 1985).

2.6. Naturalmente, è impensabile che tutte le decisioni minute possano essere registrate nei rapporti di ricerca; ma anche delle più generali e importanti vi si trova traccia assai di rado. E' ancor più di rado, per non dire mai, si trova traccia nei manuali di metodologia del fatto che di questa fitta rete di decisioni c'è assoluto bisogno per far procedere giorno per giorno la ricerca. Non si può permettere che questo fatto affiori alla coscienza collettiva perché turberebbe la visione epistemologica, ancora dominante quanto meno nel subconscio dei ricercatori, del processo di osservazione e di rilevazione come una fotografia della realtà, che non ha bisogno di alcuna decisione salvo quella di fotografare — il che, detto fra parentesi, è un'immagine assai riduttiva anche dell'attività del fotografo. L'osservazione dev'esser fotografica perché il suo frutto — la matrice dei dati — dev'essere una fotografia della realtà, anzi la realtà stessa. Quando il ricercatore (o piuttosto, ancora una volta, i suoi collaboratori: statistici e *computer men*) tratta una matrice di dati con delle tecniche di analisi, non sembra affatto consapevole del fatto che la matrice rappresenta (qualche aspetto della) realtà solo grazie a una vera impalcatura di convenzioni e a una miriade di decisioni “soggettive” e scoordinate prese dai vari esecutori.

Queste convenzioni e decisioni (così come le convenzioni e decisioni necessarie a qualsiasi trattamento statistico, e ad ogni interpretazione dei risultati di tale trattamento) devono essere rimosse dalla consapevolezza individuale e collettiva affinché si possa pensare, e proclamare, che i risultati di questo articolato e complicato processo si riferiscono direttamente alla realtà.

La rimozione è favorita da una vera e propria cintura protettiva di concetti e termini scelti e usati in modo da obnubilare il fatto che la matrice dei dati non è la stessa cosa della realtà, allo scopo di permettere al sociologo di minimizzare i contatti con il suo oggetto pur conservando la pretesa di parlarne in modo non speculativo ma “scientifico”.

Più specificamente, alcuni concetti/termini della cintura hanno la funzione di celare la differenza fra un aspetto della realtà e il suo corrispettivo convenzionale nella matrice dei dati. Importante a questo proposito è il ruolo del concetto/termine ‘variabile’, i cui referenti sono simultaneamente un vettore di cifre nella matrice, l'idea che un ricercatore si fa di una certa proprietà dei soggetti che studia, e un insieme di stati e attributi individuali che sono — in modo più o meno convenzionale — considerati relativi alla stessa proprietà (per l'appunto quella che ha in mente il ricercatore)

Il concetto/termine ‘caso’ designa simultaneamente un individuo (o comunque un oggetto) reale e un vettore-riga nella matrice. Per ‘dato’ si intende simultaneamente lo stato individuale su una proprietà, l'informazione non trattata a proposito di questo stato (come nell'espressione ‘raccolta dei dati’), e la stessa informazione passata attraverso tutte le manipolazioni previste dalla definizione operativa, e collocata sotto forma di simbolo numerico in una posizione univocamente definita della matrice.

Altri concetti/termini della cintura assolvono la funzione di coprire e legittimare il fatto che vengono ridotte a delle manipolazioni matematiche (quindi “oggettive”, quindi “scientifiche”) di vettori quelle che sono — e anche più dovrebbero essere — delle operazioni intellettuali che richiedono giudizi e decisioni “soggettive” del ricercatore, come possessore di conoscenza tacita sia sull'oggetto e il contesto della sua ricerca, sia sui vari aspetti del suo mestiere (come si deve interpretare quel certo diagramma, che grado di fiducia si può avere nella fedeltà dei dati raccolti con quella certa definizione operativa, etc.).

In questa sezione si è visto come tale funzione sia svolta dal concetto/termine ‘attendibilità’ elaborato dagli psicometrici e diffusosi fra gli scienziati sociali che fanno uso di matrici di dati. Nel par. 2.3 si è accennato a come la stessa funzione sia svolta dal concetto/termine ‘validità’ (per una trattazione più diffusa, vedi Marradi 1989, sez. 2). Un altro caso importante è il concetto/termine ‘spiegazione’, che nel linguaggio della statistica per le scienze sociali è ridotto a designare una forma di correlazione fra il vettore di cifre che rappresenta una variabile dipendente e uno o più vettori che rappresentano una o più variabili indipendenti.

Come si accennava, una funzione della cintura protettiva è di minimizzare il contatto del sociologo con il suo oggetto, e quindi il tempo, le conoscenze, la creatività che deve impegnare nelle sue ricerche. Tuttavia, pensiamo che questa funzione di tipo *zweckrational* sia secondaria rispetto a una funzione *wertrational*, che la cintura svolge consentendo di preservare l’immagine della scienza come un’impresa “oggettiva”. Si preserva tale immagine standardizzando e spersonalizzando ogni procedura di raccolta e di analisi dell’informazione, e sviando l’attenzione, appunto con gli artifici terminologici che abbiamo esemplificato, dagli ampi spazi che queste procedure standardizzate, scontrandosi con la realtà, lasciano da coprire mediante interventi decisionali *ad hoc* del ricercatore e dei suoi collaboratori.

Ma il risultato veramente cruciale che si ottiene interpretando ogni operazione cognitiva come un’operazione sulla matrice dei dati ed entro i suoi limiti è stabilire una situazione di monismo gnoseologico. Tale monismo è la condizione di ogni pretesa all’oggettività: se si ammette una pluralità di fonti di verità, fra queste fonti si può produrre un conflitto, da cui la necessità di decidere fra due o più affermazioni concorrenti di pari status — il che è evidentemente incompatibile con l’idea di un’oggettività impersonale e fotografica.

Nel nostro caso, non si può avere monismo gnoseologico se si ammette che il sociologo abbia fonti di informazione sulla realtà al di fuori della matrice dei dati; in particolare se si riconosce che un patrimonio di esperienze e di conoscenze non formalizzate è necessario in qualsiasi fase della ricerca, dalla scelta degli indicatori e delle definizioni operative alla valutazione della fedeltà dei dati, dalla scelta dei modelli e delle tecniche di analisi all’interpretazione dei risultati.

3. Fedeltà, affidabilità a priori, affidabilità a posteriori

3.1. E’ da ritenere che oltre cinquant’anni di pratica abbiano legato il termine ‘attendibilità’ alle tecniche di elaborazione bi-vettoriale o matriciale che abbiamo visto, e che sia del tutto illusorio proporre un’accezione alternativa. Peraltro, analizzando sia gli aspetti epistemologici del problema sia la pratica della ricerca, e in particolare dei sondaggi, si evidenziano aree concettuali che non possono essere designate dal termine ‘attendibilità’ nell’accezione corrente e che tuttavia hanno bisogno di una delimitazione e di una designazione.

Iniziamo da una revisione degli assunti individuati nei parr. 1.1 e 1.3. Conviene articolare e rendere più cauto l’assunto ontologico circa gli stati effettivi:

(c’) anche supponendo che, in ogni dato istante, ci sia uno stato effettivo

di ciascun oggetto sulla proprietà che ci interessa, nella maggior parte dei casi non si può esser sicuri che lo si conosce, e non si può postularlo uguale alla media o al risultato di qualunque altra elaborazione logico-matematica degli stati osservati (questa postulazione, abituale nelle scienze fisiche, è criticata anche da Berka 1983, 195-6). Tuttavia, più si approfondisce la ricerca, più possibilità si hanno di giungere alla conoscenza di tale stato effettivo, o di approssimarlo, o di constatare la sua inesistenza (il soggetto S non può avere alcuno stato su una proprietà che ci interessa: ad esempio, nessuna opinione sul problema X). Viceversa, più superficiale, standardizzata, meccanica è la ricerca, più è probabile che si registri un dato che non rispecchia affatto lo stato effettivo, o che non corrisponde ad alcuno stato effettivo.

Anche i due assunti relativi al mutamento dello stato devono essere rivisti perché, quanto meno per molte delle proprietà che interessano le scienze sociali,

(a') lo stato di un soggetto può cambiare spontaneamente fra una rilevazione e l'altra;

(b') la rilevazione può alterare tale stato, o registrarne uno anche laddove non ne esiste alcuno.

Gli assunti (d') (e) (f) devono essere sostituiti dal:

(d'') sono rilevanti non solo le differenze fra individui, ma anche ogni differenza fra le situazioni di rilevazione, inclusa naturalmente la persona del rilevatore.

La letteratura abbonda di testimonianze e considerazioni sulla rilevanza di questo o quell'aspetto della situazione per l'esito della rilevazione. Le condizioni fisiche e mentali del soggetto al momento di un test di abilità influenzano le sue prestazioni (Thorndike 1949). Queste, insieme alle fluttuazioni di umore, influenzano le sue opinioni e i suoi atteggiamenti dichiarati (Kendall 1954; Converse 1970; Noelle-Neumann 1970). Moltissime ricerche (Schank 1932; Gorden 1952; Rose 1961; Slomczynski 1969; Lutynska 1978; Przybylowska e Kistelski 1981; Pawson 1982) hanno mostrato che le risposte della stessa persona alla stessa domanda possono mutare radicalmente a seconda della sua percezione di chi somministra il test o il questionario e del proprio ruolo, sia nella specifica situazione sia nel contesto più generale (gli si chiedono le opinioni dominanti nella sua cultura, le sue opinioni come cittadino, come operaio, come padre, o le sue opinioni più intime e private?). "Un individuo può avere una varietà di atteggiamenti verso lo stesso oggetto, che sarebbero incompatibili tra loro se manifestati nello stesso contesto" (Rose 1961, 266; sulla differenza fra opinioni pubbliche e opinioni private vedi anche Cicourel 1964, 56-7; Deutscher 1972; Gostkowski 1966 e 1974).

Dato che il linguaggio dei tests e dei questionari è poco familiare alla maggioranza dei soggetti (Pinto 1964, 700; Bourdieu *et al.* 1968, 63), alcuni soggetti interpretano ogni domanda e ogni compito non elementare sulla base delle indicazioni fornite dalle circostanze, cioè dallo sviluppo dell'intervista fino a quel momento (Thurstone 1922; Sargent 1945; Cicourel 1964, 80; Belson 1981). Se il questionario verte su temi estranei al loro mondo vitale — il che succede spesso con soggetti marginali per istruzione, occupazione, età — l'intervistato risponderà a caso (Converse 1964 e 1970; Gergen e Back 1966; Lutynski 1988) e potrà essere influenzato da un qualsiasi accidente della formulazione verbale di quelle o di domande precedenti, o da un qualsiasi dettaglio della situazione (Haynes 1964; Bishop *et al.* 1982). Se l'intervistato ha fretta o è mal disposto darà la risposta più immediata e facile: "il sociologo... prende per espressione di atteggiamenti profondi delle risposte superficiali date in fretta a domande che appaiono inutili" (Bourdieu *et al.* 1968, 57; analogamente Pinto 1964, 674; si vedano anche le osservazioni di intervistatori riportate da Hyman *et al.* 1954).

Se a questo quadro si aggiungono gli errori accidentali di intervistatori e codificatori, ce n'è abbastanza per concludere che ciascun dato rispecchia lo stato corrispondente (ammesso che esista) in maniera più o meno fedele per ragioni largamente indipendenti dalle analoghe ragioni che valgono per ciascun altro dato. Purtroppo, malgrado questa imponente letteratura specifica — che si è citata solo in minima parte — sono in pochissimi (ad esempio Kriz, che parla di reificazione: 1988, 68) a tirare la conclusione che attribuire un grado di attendibilità a uno strumento in quanto tale è ingiustificabile. Oltre a Kriz, anche Frank (nel brano citato al par. 2.4) e Cicourel (1964, 80 e 109-111) accennano al fatto che non ci interessa una proprietà dello strumento, ma una proprietà del singolo atto di rilevazione, e quindi del singolo dato che ne è il risultato. Si tratta di giudicare quanto fedelmente tale dato, tenuto conto delle convenzioni introdotte dalla definizione operativa, corrisponda al (supposto) stato effettivo di un soggetto su una data proprietà. Un controllo del genere permette ovviamente di correggere i dati non corrispondenti, migliorando la qualità della rilevazione.

Per distinguere questo concetto, situato al livello del singolo atto e fondato su un confronto fra le cifre nella matrice e la realtà fuori, dal concetto degli psicometrici, relativo a un intero strumento, fondato su elaborazioni matematiche e privo di alcun riscontro con ciò che è fuori dalla matrice, ci sembra che il termine 'fedeltà' sia il più indicato. Esso ha il vantaggio di essere più specifico rispetto all'espressione 'qualità dei dati' che è stata talvolta usata (vedi ad es. Naroll 1962) in quelle ricerche che operavano un confronto fra le cifre codificate e gli stati effettivi (vedi il par. successivo).

3.2. Tre questioni si pongono: a che serve un concetto non-vettoriale, squisitamente idiografico, di fedeltà? In che modo è possibile valutare, o addirittura misurare, la fedeltà di un atto di rilevazione? Sono stati fatti tentativi in tal senso?

A nostro avviso, il concetto di fedeltà ha la fondamentale funzione epistemologica di ricordare agli scienziati sociali che i loro strumenti di raccolta di informazioni non hanno la virtù di fotografare automaticamente la realtà, e che l'obiettivo di una registrazione fedele deve essere perseguito controllando accuratamente il funzionamento effettivo degli strumenti nei singoli atti di rilevazione. Se si adotta l'idea che il problema di una registrazione fedele è stato o può essere risolto una volta per tutte con questa o quella formula di manipolazione delle cifre, o in qualsiasi altro modo che non sia un accurato controllo sul campo durante la rilevazione, ci si risparmia senza dubbio una quantità di lavoro ma si producono degli artefatti che hanno, nella migliore delle ipotesi, un rapporto tenue indiretto e incerto con l'obiettivo dichiarato di una registrazione fedele.

Pur costituendo un'esigua minoranza rispetto alla massa imponente delle ricerche sociali eseguite secondo il modello tayloristico descritto al par. 2.5, le ricerche che includono dei controlli idiografici di fedeltà non mancano. Naturalmente, il termine di confronto cui si ricorre di preferenza sono i documenti ufficiali, che per alcune proprietà (dalla residenza al possesso di patente) sono costitutivi dello stato effettivo, e quindi permettono di confrontarlo al corrispondente dato senza alcun intervento valutativo.

Già nel 1944 un controllo di questo genere è stato operato, sotto l'etichetta "studio di validità", dal National Opinion Research Center, che allora aveva sede a Denver. Si controllò sui registri pubblici la fedeltà delle dichiarazioni di un migliaio di intervistati circa la loro età, possesso di beni immobili, automobili, telefono, patente, etc. (Parry e Crossly 1950). Katona riferisce di analoghi controlli operati sulla situazione economica dei suoi soggetti (1951, 319-22). Con un disegno di ricerca più sofisticato, si è controllata sui registri pubblici la fedeltà delle ammissioni di colpevolezza per una serie di reati (bancarotta, guida in stato di ebbrezza, etc.) ottenute con quattro diverse tecniche di intervista (Locander *et al.* 1976).

I controlli di fedeltà pongono ovviamente maggiori problemi se non possono appoggiarsi su alcuna documentazione ufficiale circa gli stati sulla proprietà che interessa. Tuttavia, controlli del genere sono eseguiti in varie forme da vari istituti di ricerca. Il Centro di Metodologia presso l'Università polacca di Lodz è il più attivo nei controlli sul campo. In decine di ricerche empiriche cui il Centro presta consulenza la fedeltà dei dati viene controllata ricorrendo, a seconda delle proprietà in questione, a documenti ufficiali, all'osservazione diretta da parte di ricercatori, a informazioni fornite da vicini e conoscenti, e a "interviste di controllo in profondità... cioè interviste libere, ma guidate da uno schema di raccolta di informazioni sui processi cognitivi ed emotivi dell'intervistato nell'atto di rispondere a una domanda chiusa. Tutta l'intervista in profondità era registrata e l'intervistatore aveva l'incarico di aggiungere le sue impressioni e qualunque altra informazione rilevante poteva raccogliere... Questa 'intervista sulla intervista' si è dimostrata assai efficace per acquisire informazioni sul livello di comprensione delle domande e sui processi di formazione delle risposte" (Lutynski 1988, 182-3). Questo complesso di controlli ha permesso di stabilire che le due principali cause di infedeltà dei dati sono la difficoltà delle domande e gli errori commessi dagli intervistatori (Kistelski 1984). Quanto alle domande sulle opinioni, "si può dire con ragionevole sicurezza che solo una minima parte delle risposte... danno informazioni esatte... I risultati dei nostri studi legittimano conclusioni piuttosto pessimistiche sulla credibilità dei risultati di molti studi sociologici, e dovrebbero indurre i ricercatori a manifestare una cautela molto maggiore" (Lutynski 1988, 183). Controlli mediante interviste in profondità sono eseguiti anche dal Survey Research Center presso la London School of Economics e da altri centri di ricerca inglesi (Belson 1981). Una combinazione di controlli su registri pubblici e mediante interviste è impiegata a partire dal 1950 dalla Response Research Unit, incaricata di valutare la fedeltà dei dati censuali negli Stati Uniti.

Una tecnica che affida allo stesso intervistato il controllo della fedeltà di alcuni dati è stata proposta da Schuman (1966) con un certo successo. Alla fine dell'intervista, l'intervistatore sceglie alcune delle domande del questionario secondo una tabella di numeri casuali e chiede al soggetto di spiegare perché ha scelto l'una o l'altra risposta chiusa, in modo che si possa confrontare la sua interpretazione della risposta con quella che ne dà il ricercatore.

Il Centro di Metodologia di Lodz ha anche elaborato un modello ideale del processo di comunicazione fra ricercatore, intervistatore, intervistato e codificatore, con cui confronta l'andamento dei processi effettivi. Si è manifestata, com'era da attendersi, una relazione positiva fra grado di correttezza del processo e grado di fedeltà dei dati (Koniarek 1984). Un modello simile, limitatamente al processo di intervista, è stato approntato da Cannell e Robinson (1971) per il Survey Research Center dell'Università del Michigan: tutta l'intervista è registrata, e ogni comportamento verbale che si discosta dal modello dà luogo a una codifica in vettori a parte, ma paralleli a quelli che riportano le risposte. Questa procedura, detta Verbal Interaction Coding, è applicata anche in centri di ricerca inglesi (Morton-Williams 1979).

Infine, può essere senz'altro considerato un controllo idiografico della fedeltà dei dati il confronto, che viene effettuato abbastanza spesso analizzando i dati di un *panel*, degli stati che il soggetto ricorda nell'ondata *n* (ad esempio per quale partito ha votato nell'elezione *X*) con le sue dichiarazioni sullo stesso argomento in un'ondata precedente (vedi ad es. Powers *et al.* 1978).

3.3. Una delle tre domande poste al paragrafo precedente (è possibile misurare, o quanto meno valutare, la fedeltà di un dato?) non è stata fin qui affrontata perché conviene prima porsi un'altra: a che serve misurare, o valutare, la fedeltà di un dato, visto che la principale preoccupazione di chi individua un dato infedele sarà ovviamente sostituirlo con un dato (ritenuto) più fedele?

Una volta operata questa correzione, conservare traccia di come e/o quanto il dato precedente era infedele può servire solo se si vogliono raccogliere informazioni sull'affidabilità della definizione operativa (vedi par. 3.4). In tal caso, sarà opportuno creare dei vettori paralleli, che registrino questa informazione per ogni dato che è stato controllato, esattamente come si fa nel Verbal Interaction Coding (par. 3.2). Così codificata, questa informazione potrà essere utilizzata per valutare l'affidabilità della definizione operativa, come vedremo nel paragrafo seguente.

Delimitatine così gli scopi, consideriamo ora la possibilità di misurare, o soltanto valutare, la fedeltà idiografica. Nell'accezione rigorosa del termine, misurare la fedeltà è possibile a due condizioni:

- (i) che lo stato effettivo sia conoscibile, come sicuramente accade per almeno alcune proprietà: ad esempio le proprietà "demografiche" dell'individuo (nazionalità, luogo di nascita e di residenza, titolo di studio), e in generale tutte quelle in cui la documentazione ufficiale non si limita a registrare ma *costituisce* lo stato (possessione di patente, fatto di aver votato, situazione della fedina penale, etc.).
- (ii) che la proprietà in questione sia misurabile, cioè che sia definibile un'unità di misura dello scarto fra lo stato effettivo e quello registrato.

Resta da vedere se esistono proprietà che soddisfano entrambe queste condizioni, il che a prima vista sembra da escludere, dato che non ci sembrano conoscibili gli stati effettivi su una qualunque proprietà continua, e le proprietà discrete non sono misurabili nel senso stretto del termine.

Se tuttavia liberalizziamo la condizione (i), cioè siamo disposti a confrontare lo stato registrato da normali collaboratori in normali operazioni di *routine* non con lo stato effettivo ma con lo stato registrato con operazioni molto più accurate e direttamente controllate dal ricercatore, allora il grado di infedeltà di molti tipi di dati è quantificabile — mediante misurazioni o conteggi a seconda della natura della proprietà. Con questo adattamento la fedeltà di quasi tutti i dati censuali — tutti quelli cioè che sono frutto di conteggi e misurazioni — potrebbe essere controllata e quantificata ripetendo con cura molto maggiore le relative operazioni. Lo stesso vale naturalmente anche per i dati relativi a proprietà individuali misurabili o enumerabili.

Per i dati relativi alle altre proprietà, la fedeltà non è quantificabile, ma è comunque valutabile con sicurezza maggiore o minore a seconda della natura della proprietà in questione. Abbiamo visto i molti modi in cui alcuni centri di ricerca hanno controllato la veridicità delle dichiarazioni dei soggetti, il comportamento sul campo dei collaboratori e in genere il funzionamento delle definizioni operative. Anche queste informazioni sulla fedeltà di ogni dato possono essere codificati in una serie di vettori paralleli, in forma dicotomica (dato fedele / infedele) o con una classificazione *ad hoc* dei vari tipi rilevanti di infedeltà, oppure con una semplice scala di gravità (fedele / non gravemente infedele /

gravemente infedele) o di probabilità (sicuramente fedele / forse infedele / probabilmente infedele / sicuramente infedele).

Naturalmente i giudizi su tipo, gravità e probabilità non si escludono l'un l'altro: possono essere registrati in vettori separati, o combinati in una tipologia. Saranno naturalmente emessi dal ricercatore o da un suo collaboratore competente che si è occupato dei controlli.

Una volta che si sono registrate queste informazioni sulla fedeltà dei dati di un vettore della matrice, nulla vieta di sintetizzarle nella forma adeguata, che può essere: nel caso il giudizio sia dicotomico (fedele/infedele), la proporzione dei dati fedeli sul totale; nel caso il giudizio assuma la forma di una classificazione, ordinata o non, la percentuale dei dati in ciascuna classe o tipo; nel caso il giudizio sia quantificato, una media. Beninteso, qualunque sia il modo in cui si perviene a questa sintesi, essa riguarda esclusivamente quello specifico vettore di dati, registrati con una data definizione operativa, in un dato ambito spazio-temporale, su determinati soggetti in specifiche situazioni. Ogni estensione ad altre situazioni è illegittima. Pur entro tali stretti limiti, questa informazione di sintesi sulla fedeltà dei dati di un vettore può rendere dei servizi in una ricerca, come vedremo nel paragrafo che segue.

3.4. Si è detto di aree concettuali lasciate del tutto scoperte dal termine 'attendibilità' nell'accezione ormai consacrata dagli psicometrici. Una è la corrispondenza del singolo dato allo stato effettivo che esso rappresenta nella matrice, concetto per il quale si è proposto il termine 'fedeltà'. Un'altra area concettuale per ora non denominata è l'idea che il ricercatore si fa di come funzionerà la sua definizione operativa sul campo, cioè della probabilità che essa produca, in quella data ricerca (cioè in quell'ambito spazio-temporale, con quei soggetti studiati e quei collaboratori), dati più o meno fedeli. Per esempio, si sa che in certe culture delle domande dirette sul reddito sono destinate a raccogliere risposte non veritiere, e quindi a produrre dati infedeli; si sa che, specialmente se un questionario è somministrato da intervistatori poco interessati alla ricerca, delle domande organizzate in batteria produrranno numerosi *response-sets*; si sa che, se non si esercita alcun controllo, la codifica di articoli di giornale ai fini di un'analisi del contenuto produrrà un certo tasso di affermazioni errate; si è a conoscenza del fatto che, in precedenti ricerche nostre o altrui su soggetti e in ambiti spazio-temporali non molto dissimili, quella certa definizione operativa ha prodotto una certa proporzione di dati infedeli e/o una certa distribuzione dei vari tipi e gradi di infedeltà (e in questo consiste il principale servizio che può rendere alla comunità scientifica chi, oltre a controllare la fedeltà dei propri dati, registra questa informazione nelle forme viste al par. 3.3 e la rende accessibile). L'ampiezza e la profondità di queste conoscenze (inclusa la conoscenza del dove e del come procurarsi informazioni sugli esiti di una definizione operativa nelle ricerche precedenti) contraddistinguono il ricercatore esperto. Si tratta di una conoscenza essenziale ai fini della scelta di una definizione operativa; è in base ad essa che il ricercatore ha fiducia nella sua scelta, oppure nutre perplessità ma si adatta perché non riesce ad immaginare niente di meglio.

Questo grado di fiducia nel funzionamento di una definizione operativa in base a conoscenze ed esperienze precedenti a quella specifica raccolta di dati è una proprietà della definizione operativa stessa; è un concetto trascurato dalla letteratura metodologica, per il duplice motivo che si basa su conoscenze tacite e che è legato a un problema di scelte, cioè di decisioni "soggettive". Penso si possa opportunamente denominare 'affidabilità a priori' di una definizione operativa. Affidabilità perché la sua radice richiama l'elemento essenziale del fidarsi di uno strumento, dell'affidarsi ad esso; 'a priori' per distinguerla da un'altra forma, che vedremo subito.

Il grado di fiducia di un ricercatore nella sua definizione operativa può essere modificato da ogni sorta di informazione sugli esiti della raccolta dei dati, acquisita durante la raccolta stessa oppure in seguito, durante un'apposita fase di controllo o anche in sede di analisi della matrice dei dati. Gli intervistatori possono riferire che una domanda che *a priori* si giudicava scabrosa sulle abitudini sessuali ha provocato imbarazzo e reazioni negative in misura molto minore del previsto; i collaboratori che somministrano un test possono riferire che certi *items* sono stati mal compresi o mal interpretati da numerosi soggetti; controllando sui documenti ufficiali cui possiamo accedere le dichiarazioni dei soggetti circa il titolo di studio, ricaviamo l'impressione che la temuta tendenza a dichiarare un titolo di studio più alto del vero si sia manifestata solo di rado; passando in rassegna i questionari compilati possiamo constatare un'altissima incidenza di *response-sets*; controllando, con interviste in profondità

a un subcampione dei nostri soggetti, la fedeltà di alcuni dati sulle opinioni, registrando in matrice le relative informazioni e analizzandole, giungiamo alla conclusione che alcune domande sono state eluse o mal interpretate; calcolando la media dei redditi dichiarati dagli occupati a vario titolo, si può riscontrare che il reddito medio dei lavoratori indipendenti, anziché risultare più basso (come nelle dichiarazioni IRPEF) di quello dei lavoratori dipendenti, risulta molto più alto, come è opinione comune che sia; applicando le relative formule a una batteria di *items* si può ottenere un *alfa*, o un *theta*, o un *omega*, di livello troppo basso; e così via.

Attraverso processi come quelli descritti e innumerevoli altri, il grado di fiducia che un ricercatore nutre nella sua definizione operativa può modificarsi più volte, e anche non raggiungere uno stato definitivo, almeno fintantoché si analizzano quei dati e quindi se ne possono scoprire sempre nuovi aspetti. Conviene tuttavia distinguere l'affidabilità a priori, che una definizione operativa ha sulla base di informazioni note prima di quella specifica raccolta di dati, e che sono servite al ricercatore per sceglierla, dall'affidabilità a posteriori, che una definizione operativa ha sulla base di informazioni di ogni tipo relative all'andamento di quella raccolta e ai suoi esiti.

La fondamentale ragione per distinguere tra affidabilità a priori e a posteriori sta nel fatto che la prima è connessa alla scelta di quella data definizione operativa, e la seconda non lo è. Auspicabilmente, una parte delle informazioni su cui si basa l'affidabilità a posteriori della definizione operativa X eserciterà una certa influenza sull'affidabilità della definizione operativa Y in una ricerca successiva. Esemplificando sopra alcuni tipi di informazione rilevante, si è anticipata la tesi che l'attendibilità calcolata nel modo tradizionale, cioè mediante formule applicate ai coefficienti di correlazione fra vettori, è solo una componente dell'affidabilità a posteriori, e limitatamente ad alcune variabili — gli *items* organizzati in batteria. Il contributo informativo dei coefficienti di attendibilità, comunque, è ben lungi dall'essere conclusivo: si è visto (par. 2.2) che un alto valore del coefficiente può essere prodotto da un *response set* o da vari artifici linguistici nella formulazione degli *items*. Simmetricamente, un valore basso può dipendere dal fatto che gli *items* sono scarsamente correlati perché non riguardano la stessa dimensione semantica — il che è spiacevole, ma non ha nulla a che vedere con la fedeltà dei singoli dati.

In una comunità scientifica in cui fosse abituale un approfondito controllo dei membri sui risultati empirici raggiunti dai colleghi dello stesso settore, potrebbe servire un terzo concetto di affidabilità, relativo alla fiducia che gli altri esperti hanno in una definizione operativa data dal ricercatore X. Visto che assai raramente gli altri esperti possono e vogliono interferire nella scelta di una definizione operativa, la distinzione a priori / a posteriori sarebbe in questo caso irrilevante. Ad ogni modo, individuare e designare un concetto del genere ci sembra di scarsa utilità, visto che il dibattito metodologico sulle ricerche altrui è rarissimo, e ancor più raramente investe gli aspetti di cui ci siamo occupati in questo saggio.

Note

1 Si dice *response set* una serie di risposte uguali che un intervistato dà meccanicamente a una batteria di domande, senza riflettere sul loro significato. E' quindi un fatto paradossale, e rivelatore, che questo fenomeno di assenza di significato possa produrre il massimo di attendibilità secondo i canoni e le formule correnti.

Nella sterminata letteratura sui *response sets* e le loro cause, si segnalano Cronbach (1946), Frederiksen e Messick (1959), Berg (1967), Messick (1968), Hui e Triandis (1985).

2 Nella rapida rassegna dei significati di questi termini della cintura protettiva ci limitiamo rigorosamente a quelli abituali fra gli psicologi, i sociologi e i politologi — con i relativi metodologi — che fanno ricerca con la matrice dei dati. Ad esempio, il termine 'variabile' ha ancora altri significati, più o meno diversi fra loro, per i logici, per i matematici e per gli statistici; da questi ultimi lo ha mutuato, con qualche ritocco, la comunità degli scienziati sociali. Da questa poi il termine, come sostantivo e con i significati più bislacchi, sta ora rimbalzando nel linguaggio comune, dove prima era solo un aggettivo.

Praticamente ogni termine della cintura protettiva ha vari altri significati nel linguaggio comune e/o in altri linguaggi specializzati.

Bibliografia

- Allen, M. P.
1974 £Construction of Composite Measures by Canonical-Factor-Re-£
 £gression Method£, in H. L. Costner (a cura di), £Sociological£
 £Methodology 1973-74£, San Francisco, Jossey Bass, pp. 51-78.
- Althausen, R. P. e Heberlein, T. A.
1970 £Validity and the Multitrait-Multimethod Matrix£, in E. F.
 Borgatta e G. W. Bohrnstedt (a cura di), £Sociological£
 £Methodology 1970£, San Francisco, Jossey-Bass, pp. 151-169.
- American Psychological Association
1954 £Technical Recommendations for Psychological Tests and£
 £Diagnostic Techniques£, in "Psychological Bulletin", LI,
 Supplement, part 2, pp. 1-38.
- Ammassari, P.
1984 £Validità e costruzione delle variabili: elementi per una rifles-£
 £sione£, in "Sociologia e Ricerca Sociale", V, n. 13, pp. 141-156.
- Anastasi, A.
1953 £Differential Psychology£, London, Macmillan.
1955 £Psychological Testing£, London, Macmillan. Citazioni dalla
 trad. it. £I tests psicologici£, Milano, Angeli, 1981.
- Armor, D. J.
1974 £Theta Reliability and Factor Scaling£, in H. L. Costner (a
 cura di), £Sociological Methodology 1973-1974£, San Francisco,
 Jossey-Bass, pp. 17-50.
- Bartlett, C. J., Quay, L. C. e Wrightsman, L. S.
1960 £A Comparison of Two Methods of Attitude Measurement:£
 £Likert Type and Forced Choice£, in "Educational and
 Psychological Measurement", XX, n. 4, pp. 699-704.
- Belson, W. A.
1981 £The Design and Understanding of Questions in the Survey£
 £Interview£, London, Gower.
- Bendig, A. W.
1954 £Reliability and the Number of Rating Scale Categories£, in
 "Journal of Applied Psychology", XXXVIII, n. 1, pp. 38-40.
- Berg, I. A.
1967 £Response Set in Personality Assessment£, Chicago, Aldine.
- Berka, K.
1983 £Measurement. Its Concepts, Theories and Problems£, Dordrecht, Reidel.
- Bertrand, J. L. F.
1889 £Calcul des probabilités£, Paris, Gauthiers-Villars.

- Bishop, George *et al.*
 1982 £Effects of Presenting One versus Two Sides of an Issue in Survey£
 £Questions£, in “Public Opinion Quarterly”, XLI, n. 1, pp. 69-85.
- Blumer, H.
 1954 £What Is Wrong With Social Theory?£, in “American Sociological
 Review”, XIX, n. 1, pp. 3-10.
- Bocuzzi, E.
 1985 £Parcellizzazioni e reificazioni nelle ricerche sociologiche:£
 £il punto di vista di un'intervistatrice£, in “Rassegna
 Italiana di Sociologia”, XXVI, n. 2, pp. 239-260.
- Bohrnstedt, G. W.
 1969 £A Quick Method for Determining the Reliability and Validity£
 £of Multiple-Item Scales£, in “American Sociological Review”,
 XXXIV, n. 4, pp. 542-48.
 1970 £Reliability and Validity Assessment in Attitude Measurement£,
 in G. F. Summers (a cura di), £Attitude Measurement£,
 Chicago, Rand-McNally, pp. 80-99.
- Booth, C.
 1891-1903 £Life and Labours of the People in London£, London.
- Bourdieu, P., Chamboredon, J.-C. e Passeron, J.-C.
 1968 £Le métier de sociologue. Problèmes épistémologiques£,
 Paris, Mouton.
- Brown, W.
 1910 £Some Experimental Results in the Correlation of Mental Abilities,£
 in “British Journal of Psychology”, III, pp. 296-322.
- Brownell, W. A.
 1933 £On the Accuracy with which Reliability May Be Measured by£
 £Correlating Test Halves£, in “Journal of Experimental
 Education”, I, pp. 204-15.
- Callender, J. C. e Osburn, H. G.
 1977 £A Method for Maximizing Split-Half Reliability Coefficients£, in
 “Educational and Psychological Measurement”, XXXVII, n. 4, pp. 819-25.
- Campbell, D. T. e Fiske, D. W.
 1959 £Convergent and Discrimination Validation by the Multitrait-Multimethod£
 £Matrix£, in “Psychological Bulletin”, LVI, n. 2, pp. 81-105
- Campbell, D. T. e Stanley, J. C.
 1963 £Experimental and Quasi-Experimental Designs for Research£,
 Chicago, Rand-McNally.
- Cannell, C. F. e Robinson, J. P.
 1971 £Analysis of Individual Questions£, Ann Arbor, Survey Research Center.

- Carmines, E. G. e Zeller, R. A.
1980 £Reliability and Validity Assessment£, London, Sage.
- Cartwright, D. S.
1956 £A Rapid Non-Parametric Estimate of Multi-judge Reliability£,
in “Psychometrika”, XXI, n. 1, pp. 17-29.
- Cicourel, A. V.
1964 £Method and Measurement in Sociology£, New York, Free Press.
- Clark, R.
1977 £The Design and Interpretation of Experiments£, in J. P. B.
Allen e A. Davies (a cura di), £Testing and Experimental£
£Methods£, London, Oxford University Press.
- Converse, P. E.
1964 £The Nature of Belief Systems in Mass Publics£, in David E. Apter (a
cura di), £Ideology and Discontent£, Glencoe, Free Press, pp. 202-61.
1970 £Attitudes and Non Attitudes: Continuation of a Dialogue£, in
Edward R. Tuft (a cura di), £The Quantitative Analysis of£
£Social Problems£, Reading, Addison-Wesley.
- Cronbach, L. J.
1946 £Response Sets and Test Validity£, in “Educational and
Psychological Measurement”, VI, pp. 475-94.
1947 £Test Reliability: Its Meaning and Determination£, in
“Psychometrika”, XII, n. 1, pp. 1-16.
1949 £Essentials of Psychological Testing£, New York, Harper & Row.
1951 £Coefficient Alpha and the Internal Structure of Tests£, in
“Psychometrika”, XVI, pp. 297-334.
- D. C. A.
1974 £Measurement: Psychological£, in £The New Encyclopaedia£
£Britannica£, 15th edition, Vol. XI, pp. 734-9.
- D’Andrade, R. G.
1974 £Memory and the Assessment of Behaviour£, in H. M. Blalock
(a cura di), £Measurement in the Social Sciences. Theories£
£and Strategies£, Chicago, Aldine.
- Davies, A.
1977 £The Construction of Language Tests£, in J. P. B. Allen e A.
Davis (a cura di), £Testing and Experimental Methods£, London,
Oxford University Press.
- Denzin, N.
1969 £Symbolic Interactionism and Ethnomethodology: A Proposed Synthesis£,
in “American Sociological Review”, XXXIV, n. 6, pp. 922-934.
- Deutscher, I.
1966a £Looking Backward: Case Studies in the Progress of Methodology£
£in Sociological Research£, in “American Sociologist”, IV,
n. 1, pp. 34-42.

- 1966b £Words and Deeds: Social Science and Social Policy£, in “Social Problems”, XIII, pp. 233-54.
- 1972 £Public and Private Opinions: Social Situations and Multiple£
£Realities£, in S. Z. Nagi e R. G. Corwin (a cura di), £The£
£Social Contexts of Research£, New York, Wiley.
- Dingle, H.
1950) £A Theory of Measurement£, in “British Journal of the
Philosophy of Science”, I, n. 1, pp. 5-26.
- Dodd, S. C.
1942 £Operational Definitions Operationally Defined£, in “American
Journal of Sociology”, XLVIII, n. 4, pp. 482-9.
- Ebel, R. I.
1968 £Achievement Testing£, in £International Encyclopedia of the£
£Social Sciences£, vol. I, pp. 33-39.
- Ferrand, D. J. e Martel, J. M.
1986 £Le choix multicritère des *items* d’une échelle de mesure£, in
“Mathématique et Sciences Humaines”, n. 89, pp. 35-59.
- Festinger, L.
1947 £The Treatment of Qualitative Items by Scale Analysis£, in
“Psychological Bulletin”, XLIV, pp. 149-61.
- Fleishman, E. A.
1968 £Aptitude Testing£, in £International Encyclopedia of the£
£Social Sciences£, vol. I, pp. 369-74.
- Fleiss, J. L.
1965 £Estimating the Accuracy of Dichotomous Judgments£,
in “Psychometrika”, XXX, pp. 469-79.
- Frank, L. R.
1939 £Projective Methods for the Study of Personality£, in “Journal
of Psychology”, VIII, n. 2, pp. 389-413.
- Frederiksen, N. e Messick, S. J.
1959 £Response Set as a Measure of Personality£, in “Educational
and Psychological Measurement”, XIX, pp. 137-57.
- Frey, F.
1970 £Cross-Cultural Survey Research in Political Science£, in R. T.
Holt e J. E. Turner (a cura di), £The Methodology of£
£Comparative Research£, New York, Free Press, pp. 173-294.
- Galton, F.
1869 £Hereditary Genius£, London, MacMillan.
1909 £Essays in Eugenics£, London, Eugenics Society (postumo).
- Galtung, J.
1959 £An Inquiry into the Concepts of ‘Reliability’, ‘Intersubjec-£

- activity' and 'Constancy', in "Inquiry", II, n. 2, pp. 107-25.
- 1967 Theory and Methods of Social Research, London, Allen & Unwin.
- Gangemi, G.
 1977 Le misure di associazione tra due dicotomie nella ricerca sociale, in "Rassegna Italiana di Sociologia", XVIII, n. 4, pp. 599-628.
- Gergen, K. J. e Back, K. W.
 1966 Communication in the Interview and the Disengaged Respondent, in "Public Opinion Quarterly", XXX, n. 3, pp. 385-98.
- Ghiselli, E. E.
 1960 Differentiation of Tests in Terms of the Accuracy with which they Predict for a Given Individual, in "Educational and Psychological Measurement", XX, pp. 675-84.
- Giampaglia, G.
 1986 Alfa, *omega* e teta: è attendibile la misura dell'attendibilità?, in "Sociologia e Ricerca Sociale", VII, n. 21, pp. 75-99.
- Gorden, R. L.
 1952 Interaction between Attitude and the Definition of the Situation in the Expression of Opinion, in "American Sociological Review", XVII, n. 1, pp. 50-58.
- Gostkowski, Z.
 1966 Some Assumptions and Postulates Concerning the Empirical Research on Research Techniques in Sociology, in "Polish Sociological Bulletin", n. 1.
 1974 Toward Empirical Humanization of Mass Surveys, in "Quality and Quantity", VIII, n. 1, pp. 11-26.
- Grubbs, F. E.
 1948 On Estimating Precision of Measuring Instruments and Product Variability, in "Journal of the American Statistical Association", XLIII, pp. 243-64.
- Gulliksen, H. O.
 1936 The Content Reliability of a Test, in "Psychometrika", I, n. 3, pp. 189-94.
- Guttman, L. A.
 1946 The Test-Retest Reliability of Qualitative Data, in "Psychometrika", XI, n. 1, pp. 81-95.
- Hake, H. W. e Garner, W. R.
 1951 The Effect of Presenting Various Numbers of Discrete Steps on Scale Reading Accuracy, in "Journal of Experimental Psychology", XLII, pp. 358-66.
- Halbwachs, M.
 1912 La théorie de l'homme moyen, Paris, Alcan.

Haynes, D. P.

1964 £Item Order and Guttman Scales£, in “American Journal of Sociology”, LXIX, n. 1, pp. 51-8.

Heise, D. R. e Bohrnstedt, G. W.

1970 £Validity, Invalidity, and Reliability£, in E. F. Borgatta e G. W. Bohrnstedt (a cura di), £Sociological Methodology 1970£, S. Francisco, Jossey-Brass, pp. 104-129.

Hempel, C. G.

1961 £Fundamentals of Taxonomy£, in J. Zubin (a cura di), £Field£ £Studies in Mental Disorders£, New York, Grune & Stratton.

Hicks, J. M.

1967 £Comparative Validation of Measures by the£ £Multitrait-Multimethod Matrix£, in “Educational and Psychological Measurement”, XXVII, pp. 985-95.

Hill, K. Q.

1980 £Measurement Problems in Cross-National Analysis: Persisting£ £Dilemmas and Alternative Strategies£, in “Quality & Quantity”, XIV, n. 3, pp. 397-413.

Horst, A. P.

1934 £Item Analysis by the Method of Successive Residuals£, in “Journal of Experimental Education”, II, pp. 254-63.

1936 £Item Selection by Means of a Maximizing Function£, in “Psychometrika”, I, n. 4, pp. 229-44.

Hui, H. C. e Triandis, H. C.

1985 £The Instability of Response Sets£, in “Public Opinion Quarterly”, XLIX, n. 2, pp. 253-60.

Hyman, H. H. £et al.£

1954 £Interviewing in Social Research£, Chicago University Press.

Ingram, E.

1977 £Basic Concepts in Testing£, in J. P. B. Allen e A. Davies (a cura di), £Testing and Experimental Methods£, London, Oxford University Press.

Jahoda, M., Deutsch, M. e Cook, S. W.

1951 £Research Methods in Social Relations£, New York, Dryden.

Jones, C.

1921 £A First Course in Statistics£, London, Bell.

Kalton, G. e Stowell, R.

1979 £A Study of Coder Variability£, in “Applied Statistics”, XXVIII, n. 3, pp. 276-89.

Kaplan, A.

1964 £The Conduct of Inquiry£, San Francisco, Chandler.

- Katona, G.
1951 £Psychological Analysis of Economic Behaviour£, New York, McGraw-Hill.
- Kendall, P.
1954 £Conflict and Mood£, New York, Free Press.
- Kendall, P. e Lazarsfeld, P. F.
1950 £Problems of Survey Analysis£, in R. K. Merton e P. F. Lazarsfeld (a cura di), £Continuities in Social Research£, Glencoe, Free Press.
- Keynes, J. M.
1921 £A Treatise on Probability£, London, Macmillan. Ristampato come vol. VIII dei £Collected Writings£, London, Macmillan, 1973.
- Kistelski, K.
1984 £Bledy ankieterskie w wywiadzie kwestionariuszowym i ich skutki£ (Gli errori degli intervistatori nell'intervista su questionario e le loro conseguenze), in J. Lutynski (a cura di), £Acta£ £Universitatis Lodziensis. Folia Sociologica£, Lodz, vol. IX.
- Koniarek, J.
1984 £Porównanie różnych sposobów weryfikacji informacji uzyskanych w£ £badaniach z wykorzystaniem kwestionariusza£ (Confronto fra vari modi di controllare l'informazione raccolta con i sondaggi), in J. Lutynski (a cura di), £Acta Universitatis Lodziensis. Folia£ £Sociologica£, Lodz, vol. IX.
- Kriz, J.
1988 £Facts and Artefacts in Social Science. An Epistemological and£ £Methodological Analysis of Empirical Social Science Research£ £Techniques£, New York, MacGraw-Hill.
- Kuder, G. F. e Richardson, M. W.
1937 £The Theory of the Estimation of Test Reliability£, in "Psychometrika", II, n. 3, pp. 151-60.
- La Palombara, J.
1970 £Parsimony and Empiricism in Comparative Politics: an£ £Anti-Scholastic View£, in R. T. Holt e J. E. Turner (a cura di), £The Methodology of Comparative Research£, New York: Free Press.
- Likert, R.
1932 £The Method of Constructing an Attitude Scale£, in R. Likert, £A Technique for the Measurement of Attitudes£, in "Archives of Psychology", monogr. n. 140, pp. 44-53.
- Locander, W., Sudman, S. e Bradburn, N.
1976 £An Investigation of Interview Method, Threat and Response£ £Distorsion£, in "Journal of the American Statistical Association", LXXI, n. 354, pp. 269-75.
- Lord, F. M. e Novick, M. R.

- 1968 *Statistical Theory of Mental Test Scores*, Reading, Addison-Wesley.
- Lutynska, K.
 1978 *Ankieterzy i badacze. Z badan nad wplywem ankieterskim*, in "Przeglad Socjologiczny", XXX, pp. 143-73.
- Lutynski, J.
 1988 *Un centro di ricerca sulle tecniche di raccolta dei dati*, in A. Marradi (a cura di), *Costruire il dato*, Milano, Angeli.
- Macrae, D.
 1970 *Issues and Parties in Legislative Voting: Methods of Statistical Analysis*, New York, Harper & Row.
- Marradi, A.
 1988 *Casualità e rappresentatività di un campione nelle scienze sociali. Contributo a una sociologia del linguaggio scientifico*, in R. Mannheim (a cura di), *I sondaggi elettorali e le scienze politiche: Problemi metodologici*, Milano, Angeli, pp. 51-134.
 1989 *La validité des indicateurs et la fidélité des données*, in J. Duchêne e G. Wunsch (a cura di), *L'explication en sciences sociales*, Louvain-la-Neuve, CIACO.
- McKennell, A. C.
 1970 *Attitude Measurement: Use of Coefficient Alpha with Cluster or Factor Analysis*, in "Sociology", IV, n. 2, pp. 227-45.
 1977 *Attitude Scale Construction*, in C. O' Muirheartaigh e C. Payne (a cura di), *The Analysis of Survey Data*, New York, Wiley, vol. I.
- McKenzie, D.
 1981 *Statistics in Britain, 1865-1930. The Social Construction of Scientific Knowledge*, Edinburgh University Press.
- Merritt, R. L. e Rokkan, S. (a cura di)
 1966 *Comparing Nations. The Use of Quantitative Data in Cross-National Research*, New Haven, Yale University Press.
- Messick, S. J.
 1968 *Response Sets*, in *International Encyclopedia of the Social Sciences*, vol. XIII, pp. 492-6.
- Morgenstern, O.
 1963 *On the Accuracy of Economic Observations*, Princeton, Princeton University Press.
- Morton Williams, J.
 1979 *The Use of "Verbal Interaction Coding" for Evaluating a Questionnaire*, in "Quality & Quantity", III, n. 1, pp. 59-75.
- Naroll, R.
 1962 *Data Quality Control*, New York, Free Press.
 1968 *Some Thoughts on Comparative Method in Cultural Anthropology*,

in H. M. Blalock e A. B. Blalock (a cura di), *Methodology in Social Research*, New York, McGraw-Hill, pp. 236-277.

Noelle-Neumann, E.

1970 *Wanted: Rules for Wording Structured Questionnaires*, in "Public Opinion Quarterly", XXXIV, n. 2, pp. 191-201.

Nowak, S.

1976 *Understanding and Prediction. Essays in the Methodology of Social and Behavioral Theories*, Dordrecht, Reidel.

Parry, H. e Crossly, H. M.

1950 *Validity of Responses to Survey Questions*, in "Public Opinion Quarterly", XIV, n. 1, pp. 61-80.

Pawson, R.

1982 *Desperate Measures*, in "British Journal of Sociology", XXXIII, n. 1, pp. 35-63.

Phillips, B.

1966 *Social Research. Strategy and Tactics*, London, Macmillan.

Pinto, R.

1964 *Méthodes des sciences sociales*, Paris, Dalloz.

Powers, E., Goudy, W. J. e Keith, P.

1978 *Congruence Between Panel and Recall Data in Longitudinal Research*, in "Public Opinion Quarterly", XLII, n. 3, pp. 380-9.

Przybyłowska, I. e Kistelski, K.

1981 *The Social Context of Questionnaire Interview*, Lodz, Instytut Sojologii Univ. Łódzkiej.

Quetelet, L. A. J.

1869 *Physique sociale, ou essai sur le développement des facultés de l'homme*, Bruxelles, C. Murquaedt.

Raju, N.

1977 *A Generalization of Coefficient Alpha*, in "Psychometrika", XLII, n. 4, pp. 549-65.

Ricolfi, L.

1987 *Sull'ambiguità dei risultati delle analisi fattoriali*, in "Quaderni di Sociologia", XXXIII, n. 8, pp. 95-129.

Rose, A. M.

1961 *Inconsistencies in Attitudes Toward Negro Housing*, in "Social Problems", VIII, pp. 266-82.

Roth, J.

1965 *Hired Hand Research*, in "American Sociologist", I, n. 1, pp. 190-6.

Rozelle, R. e Campbell, D. T.

- 1969 £More Plausible Rival Hypotheses in the Cross-Lagged Panel£
£Correlation Technique£, in “Psychological Bulletin”, LXXI,
pp. 74-80.
- Rulon, P. J.
1939 £A Simplified Procedure for Determining the Reliability of a Test£
£by Split-halves£, in “Harvard Educational Review”, IX, n.1 pp. 99-103.
- Sargent, H. D.
1945 £Projective Methods: Their Origins, Theory and Application in£
£Personality Research£, in “Psychological Bulletin”, XLII, n. 5,
pp. 257-93.
- Schanck, R. L.
1932 £A Study of a Community and Its Groups and Institutions Conceived£
£as a Behaviour of Individuals£, in “Psychological Monographs”.
- Schuman, H.
1966 £The Random Probe: A Technique for Evaluating the Validity of Closed£
£Questions£, in “American Sociological Review”, XXV, n. 1, pp. 3-25.
- Scott, W. A.
1968 £Attitude Measurement£, in G. Lindzey e E. Aronson (a cura di),
£Handbook of Social Psychology£, Reading, Addison-Wesley,
2nd edition, vol. II.
- Shaw, M. E. e Wright, J. M.
1967 £Scales for the Measurement of Attitudes£, New York, McGraw-Hill.
- Siegel, P. M. e Hodge, R. W.
1968 £A Causal Approach to the Study of Measurement Error£, in H. M.
Blalock e A. B. Blalock (a cura di), £Methodology in Social£
£Research£, New York, McGraw-Hill.
- Singer, J. D.
1982 £Variables, Indicators, and Data£, in “Social Science History”,
VI, n. 2, pp. 181-217.
- Sletto, R. F.
1936 £Critical Study of the Criterion of Internal Consistency in£
£Personality Scale Construction£, in “American Sociological
Review”, I, n. 1, pp. 61-8.
- Slomczynski, K. M.
1969 £Conditions of Interview: Their Impact Upon Statements of£
£Respondents£, in “Polish Sociological Bulletin”, n. 2.
- Smelser, N. J.
1976 £Comparative Methods in the Social Sciences£, Englewood Cliffs,
Prentice-Hall.
- Spearman, C.
1910 £Correlation Calculated from Faulty Data£, in “British Journal of

Psychology”, III, pp. 271-295.

Spector, P. E.

1981 £Research Design£, London, Sage.

Taylor, C. L. e Hudson, M. C.

1972 £World Handbook of Political and Social Indicators, II£,
New Haven, Yale University Press.

Thorndike, R. L.

1949 £Personnel Selection. Test Measurement Techniques£, New York, Wiley.

Thurstone, L. L.

1922 £The Stimulus-Response Fallacy£, in “Psychological Review”, XXX, n. 5.

1928 £Attitudes Can Be Measured£, in “American Journal of Sociology”,
XXXIII, n. 4, pp. 529-54.

Thurstone, L. L. e Chave, E. J.

1929 £The Measurement of Attitude£, University of Chicago Press.

Traugott, M. W. e Katosh, J. P.

1979 £Response Validity in Surveys of Voting Behaviour£, in
“Public Opinion Quarterly”, XLIII, pp. 395-77.

Tukey, J. W.

1954 £Causation, Regression and Path Analysis£, Ames, Iowa State College Press.

Upshaw, H. S.

1968 £Attitude Measurement£, in H. M. Blalock e A. B. Blalock (a cura di),
£Methodology in Social Research£, New York, McGraw-Hill, p. 60-111.

Webb, E. J., Campbell, D. T., Schwartz, R. D. e Sechrest, L.

1966 £Unobtrusive Measures: Nonreactive Research in the Social£
£Sciences£, Chicago, Rand McNally.

Werts, C. E. e Linn, R. L.

1970 £Cautions in Applying Various Procedures for Determining the£
£Reliability and Validity of Multiple Item Scales£, in “American
Sociological Review”, XXXV, n. 4, pp. 757-9.

Zetterberg, H. L.

1954 £On Theory and Verification in Sociology£, Totowa, Bedminster Press.

Zubin, J.

1934 £The Method of Internal Consistency for Selecting Test Items£,
in “Journal of Educational Psychology”, XXV, pp. 345-56.