

*cap.*

2



## **METODI, PROCEDIMENTI E REQUISITI**

Data la vasta utilizzazione degli strumenti nell'attività dell'orientamento è opportuno descrivere brevemente metodi e procedimenti ed esporre i requisiti dei mezzi diagnostici in generale. Ciò al fine di consentire agli operatori l'acquisizione di informazioni adeguate per un corretto uso dei mezzi maggiormente utilizzati - sia in relazione alla scelta degli ambiti di applicazione che per una adeguata interpretazione dei risultati - e, inoltre, la diffusione di elementi concettuali utili in fase di produzione di strumenti di orientamento.

In tal senso uno spazio espositivo particolarmente ampio è dedicato alla descrizione delle caratteristiche e delle tipologie dei mezzi diagnostici standardizzati e alla corretta metodologia per l'utilizzo degli strumenti stessi. Particolare rilevanza riveste, infatti, la possibilità di condividere definizioni, anche relative all'oggetto delle misurazioni (ad es. intelligenza, personalità, ecc.), per costruire, sulla base della letteratura esistente, standard comuni e uniformità interpretative.

### **2.1 METODI ED UTILIZZO DEGLI STRUMENTI**

Di seguito vengono esposti i procedimenti maggiormente impiegati in orientamento e i relativi ambiti di applicazione. In particolare, vengono descritti i metodi (osservazione, colloquio, mezzi standardizzati e non) e successivamente le finalità che tali metodi supportano e i diversi ambiti di applicazione.

#### **2.1.1 Metodi**

*Osservazione* - L'osservazione è un procedimento molto usato per la conoscenza delle persone, ma per essere sufficientemente attendibile deve essere condotta con un minimo di metodo e possibilmente con un mezzo strutturato con cui stabilire l'obiettivo dell'osservazione stessa. In generale viene usato un modulo di annotazione e vengono variati i tempi in cui il soggetto viene osservato. In molti casi il modulo viene riempito dopo l'osservazione del soggetto. I moduli vanno sotto varie denominazioni (scheda personale, protocollo di accoglienza) e sono ampiamente usati. L'osservazione spontanea, episodica e occasionale è di scarsa utilità per la conoscenza del soggetto.

*Colloquio* - Il colloquio consiste in un contatto diretto con il soggetto allo scopo di raccogliere dei dati che si riferiscono ai fatti e agli eventi della sua vita in vista di specifici obiettivi (ad es. per assumere il personale). Conoscere la storia della persona assume notevole importanza per la definizione di un qualsiasi intervento.

A supporto del colloquio può essere usato un modulo (scheda personale, medica, familiare, ecc.) come strumento di annotazione o come traccia per acquisire informazioni (sia personali che sociali), per poi dai fatti e dagli eventi passare all'interpretazione e alla valutazione dei medesimi.

*Mezzi non standardizzati* - Tanto i ricercatori quanto gli operatori sociali elaborano dei mezzi e degli strumenti per conoscere i gruppi sotto le più svariate manifestazioni. Espletata la ricerca termina anche la funzione del mezzo utilizzato. La sua finalità è quella di descrivere il gruppo e non può servire per una diagnosi individuale.

*Mezzi standardizzati* - Si chiamano così tutti gli strumenti che sono stati sottoposti ad una sperimentazione. È l'autore a stabilire come vanno somministrati, come deve essere rilevato il punteggio oggettivamente e, infine, come devono essere interpretati i risultati. Possono essere usati tanto per la diagnosi individuale quanto per la verifica delle ipotesi nella ricerca scientifica. I mezzi standardizzati vanno sotto il termine *test* nel suo significato più ovvio di *prova*. Secondo alcuni autori il test è un procedimento sistematico per misurare un campione di comportamento; secondo altri il test è un procedimento sistematico per osservare il comportamento di un individuo ed esprimere il risultato per mezzo di una scala numerica oppure per mezzo di una categoria. Un ulteriore contributo sulla natura del test viene offerto dalla seguente definizione: il test è uno strumento scientifico di routine, elaborato per esaminare una o più caratteristiche della persona accertate empiricamente; il risultato ottenuto viene espresso con i numeri, secondo una scala che consente di esprimere come tali caratteristiche siano presenti nell'individuo (Lienert, 1967).

### 2.1.2 Applicazione

Dalle definizioni di test riportate sono emerse le principali finalità e quindi anche i possibili ambiti di applicazione.

Conviene però specificare gli usi principali nel modo più dettagliato.

#### *Diagnosi*

In molti casi occorre conoscere il modo di essere di alcuni soggetti per aiutarli a superare i compiti evolutivi ed eventualmente a supportarli nella decostruzione delle difficoltà che ne impediscono lo sviluppo. Un certo numero di test può contribuire efficacemente ad un accertamento delle caratteristiche soggettive. Lo stesso vale anche per un gruppo o per una comunità che intende perseguire delle finalità specifiche. I vari test possono contribuire alla conoscenza tanto dell'individuo quanto del gruppo e in rapporto a tale conoscenza possono essere elaborati e preordinati degli interventi.

#### *Valutazione*

L'educazione si prefigge delle chiare finalità e stabilisce dei precisi obiettivi. Fa parte del pro-

cesso educativo verificare se tali finalità e obiettivi vengono raggiunti dopo un congruo tempo di attività. Questa esigenza è particolarmente sentita per quanto riguarda l'apprendimento e l'acquisizione delle abilità a vari livelli di crescita dei soggetti.

### *Decisione*

I soggetti in sviluppo (come anche gli adulti) devono prendere delle decisioni nelle varie fasi della loro esistenza. Ogni decisione deve essere presa dopo un accurato esame delle alternative e delle conseguenze che la decisione comporta. I test possono contribuire al processo decisionale in quanto offrono delle informazioni sul soggetto che deve operare la scelta e spesso anche sulle istituzioni verso le quali il soggetto è orientato.

### *Predizione*

La predizione non è altro che una variazione della decisione già esaminata. La differenza sta nel tempo (prossimo o remoto); mentre la decisione prende in considerazione il futuro immediato, la predizione mira ad un traguardo più lontano. La predizione di un qualsiasi comportamento è una semplice congettura basata sulla probabilità e nello stesso tempo sulla costanza del comportamento umano.

Gli ambiti di applicazione dei metodi esposti possono essere diversi. Di seguito vengono descritti quelli maggiormente rilevanti in relazione alle finalità esposte.

### *Collocamento*

Alcune istituzioni perseguono finalità specifiche e per realizzarle preparano il proprio personale. Data la varietà di compiti, nella preparazione delle persone si propongono di sviluppare delle potenzialità individuali. I test possono contribuire alla conoscenza degli individui e ad un collocamento più rispondente alle loro caratteristiche.

### *Selezione*

Per quanto impopolare, la selezione nei paesi avanzati è inevitabile. In molti casi il numero di concorrenti supera la disponibilità dei posti. Occorre perciò stabilire dei requisiti dei candidati in corrispondenza alle attività che i prescelti devono poi svolgere. Il test può contribuire a questa scelta se rileva almeno parzialmente le caratteristiche desiderate.

### *Ricerca*

Nella ricerca scientifica viene fatto un uso vastissimo di test. Infatti non è possibile condurre nessuna verifica sperimentale senza una rilevazione di dati. I test servono per la descrizione della situazione iniziale e poi per la rilevazione finale per vedere l'effetto del fattore introdotto.

### 2.1.3 Costruzione ed impiego degli strumenti

In una delle definizioni del test è stato detto che esso è un mezzo scientifico. Ciò vuol dire semplicemente che esso viene elaborato con una adeguata metodologia. Al fine di comprendere meglio le diverse tipologie di strumenti e il loro relativo livello di accuratezza nella misurazione delle diverse caratteristiche, conviene tracciare le fasi principali di tale metodologia seguendo Brown (1976).

#### Schema1 - Fasi principali della elaborazione dei mezzi psicodiagnostici



Il lavoro inizia con la fase preparatoria in cui si stabiliscono gli obiettivi che si intendono raggiungere con il test, poi si definisce la dimensione e vengono identificate le sue componenti.

Dopo di ciò si passa dagli obiettivi alla realizzazione, con la raccolta e la formulazione dei quesiti. In tal modo si ottiene la stesura sperimentale del futuro test. Esso viene somministrato ad un ridotto numero di soggetti ai quali lo si intende destinare e raccolti i dati si passa all'analisi dei quesiti per verificare l'*indice di difficoltà* e di *discriminazione*.

La **difficoltà** di un quesito dipende dal numero delle risposte giuste date da tutti i soggetti ai quali la prova è stata somministrata. Se ad un quesito solo pochi hanno dato la risposta giusta, il quesito è difficile; se la metà dei soggetti ha dato la risposta giusta, il quesito è di difficoltà media; se invece la maggior parte ha risposto bene, il quesito è molto facile. La difficoltà del quesito è quindi rappresentata dalla proporzione delle risposte giuste in rapporto al numero totale dei soggetti. Tale proporzione può essere espressa in vari modi, ma quello più abituale e anche più comprensibile è la percentuale delle risposte giuste. Se su cento risposte date le risposte giuste sono cinquanta, la difficoltà del quesito è del 50%.

Il secondo indice esprime la **discriminazione** del quesito. Dalle definizioni del test riportate emerge con chiarezza che la sua finalità è quella di disporre gli individui lungo la dimensione in questione, oppure assegnarli ad una determinata categoria in base alla presenza o assenza della caratteristica che lo strumento intende misurare. La finalità principale dello strumento quindi è quella di discriminare i soggetti in base ad una (o più) caratteristiche. In base ad una informazione esterna al test i soggetti vengono assegnati secondo la loro riuscita in superiori (soggetti in cui si presume presente la caratteristica indagata) e in inferiori (soggetti in cui non si presume presente la caratteristica indagata) per confrontare la loro risposta al test con tale informazione. La capacità discriminante di un quesito (e poi dell'intero strumento) può essere intesa come la precisione con la quale il quesito (o poi lo strumento) riesce a separare, meglio discriminare, il gruppo di soggetti in relazione alla presenza della caratteristica. La capacità discriminante del quesito viene rappresentata dalla proporzione delle risposte giuste date dai due gruppi. La proporzione è espressa in frazioni. Evidentemente non tutti i soggetti del gruppo dei superiori daranno la risposta giusta al quesito, come anche quelli del gruppo degli inferiori daranno la risposta sbagliata. Il numero dei due tipi di risposte viene sottratto da tale unità.

I quesiti con i due indici validi vengono inclusi nel test e in tal modo si ottiene la sua stesura finale. Si passa poi alla standardizzazione del medesimo descrivendo il modo di somministrarlo, di correggerlo e stabilendo la durata dell'esecuzione.

Successivamente si passerà alla verifica empirica del medesimo, esaminando la sua fedeltà, validità ed elaborando delle norme basate su un notevole numero dei soggetti. Ottenuti risultati positivi si passa alla diffusione del test.

## 2.2 CARATTERISTICHE FONDAMENTALI

Ogni mezzo diagnostico deve possedere delle caratteristiche per poter essere utilizzato. Alcune caratteristiche sono puramente desiderabili come l'economicità, la facilità di somministrazione, la qualità grafica. Tre caratteristiche, invece, sono fondamentali: oggettività, fedeltà e validità.

### 2.2.1 Oggettività

L'oggettività è il primo requisito essenziale del test e nello stesso tempo è anche il presupposto degli altri due requisiti, e cioè la fedeltà e la validità. Nessun rilevamento potrebbe essere fedele e valido se non fosse innanzitutto oggettivo. L'oggettività però è solo uno dei presupposti del test, che non assicura necessariamente né la fedeltà né la validità.

La maggior parte degli autori considera l'oggettività del test nel suo risultato finale, definendola come accordo di giudizio emesso da varie persone. Qualche autore la considera però nel suo significato etimologico, notando che essa consiste nell'indipendenza dei risultati da colui che utilizza il test. In ogni caso il test deve essere elaborato in modo da escludere al massimo ogni giudizio arbitrario nella valutazione del comportamento. In altre parole, il test deve permettere a colui che lo usa di controllare tutti i fattori contrari alla finalità del test stesso. In questa prospettiva l'oggettività è nettamente contrapposta alla soggettività, ossia alla valutazione del comportamento del soggetto basata su criteri personali di colui che fa la rilevazione (Lienert, 1967).

È del tutto evidente che si riscontra una notevole variazione nell'oggettività degli strumenti. Questo dipende dalle aree che essi rilevano, dai metodi sui quali si basa la rilevazione e dalla capacità o meno dei soggetti stessi di dare risposte conformi alle finalità del test.

L'oggettività si ottiene per mezzo di **tre fasi**: a) **somministrazione**, b) **correzione** o rilevazione del punteggio, c) **interpretazione**. Il modo di procedere nelle tre fasi deve essere definito con la massima precisione possibile e deve essere basato su verifiche sperimentali. Per ogni fase l'autore deve stabilire delle norme alle quali l'utente del test deve attenersi con precisione.

#### a) Somministrazione

La prima fase consiste nella somministrazione individuale oppure collettiva del mezzo diagnostico. Essa deve avvenire in un ambiente tranquillo, ben illuminato e comodo per tutti i

soggetti. Occorre introdurre il test per motivare i soggetti e dare delle precise indicazioni sul suo svolgimento.

### **b) Correzione (scoring)**

L'autore del mezzo diagnostico deve fornire delle informazioni sul come correggere le risposte oppure sul come rilevare il punteggio (*scoring*). Per molti mezzi la correzione viene fatta con le griglie, assicurando in tal modo il massimo dell'oggettività. In tanti altri casi le risposte possibili sono molteplici e l'autore può solo proporre qualche risposta "modello".

### **c) Interpretazione**

In questa fase viene dato un significato ai dati rilevati. Colui che interpreta i dati del test mette a confronto le varie informazioni ottenute, le inserisce in un quadro più ampio del comportamento del soggetto, cerca i legami tra le varie manifestazioni e rileva possibilmente le cause del comportamento. A tale scopo si serve delle informazioni dell'autore del test, particolarmente degli studi in cui è stata esaminata la sua validità. Questa fase è la più creativa ma anche la più delicata. I risultati del test vengono integrati in una teoria e viene elaborato un intervento a favore del soggetto.

#### *Disposizioni permanenti*

Sin dall'inizio di questo secolo è stato constatato che i soggetti, rispondendo ai questionari, danno a volte delle risposte non strettamente attinenti al contenuto dei quesiti. Il modo di rispondere assume varie strutture, note sotto il termine *stili di risposta* che consistono in risposte evasive, estreme e incoerenti. I più noti stili sono: risposte in serie (*response set*), acquiescenza e desiderabilità sociale. Quando sono stati scoperti tali stili essi sono stati considerati modalità indesiderabili di rispondere al test, ma successivamente sono stati considerati aspetti diagnostici della personalità stessa. Conviene descriverli almeno brevemente.

*Risposte in serie (response set)* consistono nella predisposizione del soggetto a preferire un tipo di alternativa, del tutto o parzialmente indipendentemente dal contenuto del quesito. Alcuni soggetti tendono a marcare molti Sì, altri molti No; altri scelgono le alternative tenendo presenti le risposte già date in una o in un'altra colonna.

*Acquiescenza* è una specificazione del *response set*, in quanto il soggetto tende a preferire, particolarmente in caso di dubbio, le risposte positive (Sì, Vero, D'accordo) anziché quelle negative.

Alcuni autori hanno cercato di elaborare delle scale per scoprire l'acquiescenza adottata dai soggetti nel rispondere ai questionari. Qualche autore però ha notato che le risposte in serie correlavano con alcuni costrutti psicologici, tra cui in modo particolare il grado di stima di sé.

Qualche altro ricercatore ha constatato che i soggetti che segnavano le risposte positive risultavano essere più ottimisti e meglio adattati socialmente.

*Desiderabilità sociale* è la tendenza dei soggetti ad attribuirsi delle caratteristiche che socialmente sono più accettate e apprezzate e a rifiutare quelle che sono socialmente considerate come indesiderabili. Anche la desiderabilità sociale, come a sua volta l'acquiescenza, è stata rivalutata nel suo significato. Questa modalità di rispondere ai quesiti è diventata pure una dimensione della personalità e sotto diversi punti di vista socialmente accettabile. Infatti esistono delle scale che accertano il grado di presenza di questa dimensione negli individui per capire e predire il loro comportamento costruttivo.

Da questa descrizione è evidente che i tre tipi di predisposizione sono in una certa misura contrari all'oggettività. Occorre tenerne conto nella interpretazione dei risultati individuali. Tali dati devono essere tenuti presenti anche nelle interpretazioni dei dati delle ricerche. I tre tipi di predisposizione svolgeranno il ruolo di variabili intervenienti, come fattori che interagiscono tra le variabili in esame.

#### *Contraffazione*

Il fatto che si possano contraffare le risposte in tutti i tipi di mezzi diagnostici è noto da molto tempo. Le prime ricerche in questo senso sono state condotte sugli inventari degli interessi professionali e sui questionari di personalità. Quando è stato introdotto l'uso dei test proiettivi nella diagnosi della personalità, vi era la convinzione universale che essi non sarebbero stati soggetti alla contraffazione. Attualmente esiste una ricca documentazione sulla possibilità di contraffare tutti i mezzi diagnostici, anche se, naturalmente non nella stessa misura. È del tutto evidente che la contraffazione conscia o/e inconscia compromette l'oggettività della diagnosi. Gordon e Gross (1978) definiscono la contraffazione come "... la vulnerabilità d'un mezzo diagnostico in base alla quale il soggetto risponde ai quesiti sistematicamente in modo distorto per dare un'impressione favorevole e in alcuni casi persino sfavorevole di se stesso". La contraffazione dipende da alcune dimensioni personali, particolarmente dall'atteggiamento difensivo. La contraffazione però, diversamente da quello che si pensa, non è molto frequente.

#### *Apprendimento*

Somministrando un mezzo diagnostico ad un gruppo di soggetti, si suppone che tutti siano nelle stesse condizioni in quanto hanno uguale familiarità con tali mezzi. L'uso esteso dei mezzi diagnostici fa sì che questa supposizione non sia più fondata. Inoltre nei paesi nei quali i test sono stati usati per l'ammissione alla scuola secondaria molti genitori hanno sottoposto i loro figli ad un metodico esercizio sui test attitudinali e sulle prove oggettive di profitto per

superare tali prove. Attualmente chi deve affrontare i test per concorrere ad un posto di lavoro oppure sottoporsi alla selezione nell'esercito cerca delle informazioni per trovarsi avvantaggiato nelle prove. Al giorno d'oggi esistono numerose pubblicazioni per prepararsi ad una prova basata sui test.

In generale si ottiene solo un modesto miglioramento nel risultato ai test attitudinali; invece, spesso notevole risulta il miglioramento ai questionari di personalità, di interessi professionali e degli atteggiamenti.

*Test wiseness* - Sotto il termine *test wiseness* si intende l'abilità del soggetto ad utilizzare gli aspetti puramente formali d'un mezzo diagnostico e della situazione in cui si svolge la prova, in modo tale da aumentare il punteggio indipendentemente dalle sue capacità oppure dal suo modo di essere. In sostanza, la *test wiseness* sta nell'abilità del soggetto di padroneggiare gli aspetti gestionali del mezzo diagnostico e ottenere in conseguenza un punteggio migliore, indipendentemente dal contenuto del mezzo stesso. Per questa ragione i ricercatori considerano la *test wiseness* una fonte di errore.

Nell'interpretazione dei risultati occorre tenere conto delle modalità esposte e interpretare i risultati con cautela. Naturalmente alcuni metodi di rilevazione sono maggiormente condizionati da tali modalità.

L'esaminatore stesso deve chiedersi se non assume verso il soggetto un atteggiamento fondato su una sua struttura personale stabile, introducendo nel processo dell'interpretazione una dose di errore. Schafer (1971) ha esaminato questo rischio riferendosi all'interpretazione dei dati ottenuto da uno strumento proiettivo. L'autore ha distinto quattro possibili atteggiamenti dell'esaminatore verso il soggetto:

- l'aspetto voyeuristico (egli esplora nell'intimo di molti individui senza impegnarsi in un rapporto sociale);
- l'aspetto autocratico (controlla e domina il soggetto);
- l'aspetto oracolare (trae delle prodigiose deduzioni dai segni e dai simboli, "penetra" nei significati nascosti, predice l'evolversi degli eventi);
- l'aspetto da santo (si considera utile all'umanità).

In questa galleria di tipi non poteva mancare l'esaminatore con un certo sadismo, colui che per mezzo del test scova aspetti "degradanti" e "umilianti" nel soggetto; l'esaminatore masochista che con il suo atteggiamento provoca sin dall'inizio l'ansia nel soggetto fino a dimostrare che egli è "realmente cattivo".

Questa tipologia, per quanto approssimativa, rivela quanto lontani si sia da una rilevazione oggettiva della situazione, alla quale l'esaminatore partecipa con i suoi problemi personali.

### 2.2.2 Fedeltà

Lo scopo del test è quello di aiutare gli esperti nella valutazione del comportamento, in modo da ottenere risultati oggettivi e nello stesso tempo anche *fedeli*. È universalmente noto che il comportamento umano dipende da tanti fattori di cui alcuni interni al soggetto (stati d'animo) ed altri esterni (le situazioni in cui si trova). Lo scopo di ogni diagnosi è quello di accertare la reale situazione del soggetto, le sue concrete abilità, i suoi *veri* interessi e i suoi *genuini* atteggiamenti. Solo in questo modo sarà possibile accertare la costanza del comportamento.

Con il termine di fedeltà si intende il grado di precisione con cui un test misura una dimensione della personalità, oppure qualche aspetto del comportamento, senza indagare se li misura effettivamente (Lienert, 1967). Qualche autore la intende nel suo aspetto statistico: la fedeltà può essere definita come il rapporto tra la vera varianza di una serie di punteggi di un test e la varianza totale o osservata (rilevata) del test (Brown, 1976). Oppure in un modo più ampio: la fedeltà si esprime per mezzo della consistenza del punteggio ottenuto dallo stesso soggetto, esaminato con lo stesso test in condizioni differenti, oppure con una sua forma parallela. Alla base di questo procedimento sta la convinzione che sia possibile identificare l'errore di misurazione e predire l'estensione della fluttuazione, dovuta a fattori casuali del punteggio del singolo soggetto (Anastasi, 1976).

La fedeltà viene accertata sostanzialmente con due procedimenti: il primo consiste nella riapplicazione del test ad una ragionevole distanza di tempo, confrontando i risultati della prima applicazione con quella della seconda. Il risultato viene espresso con il coefficiente di correlazione. Più alto è il coefficiente e più fedele è il test. Il secondo sta nella verifica della consistenza dei quesiti che compongono il test. La consistenza viene data dall'indice alfa. Più alto è tale coefficiente e maggiore è anche la consistenza del test. Spesso la consistenza corrisponde anche all'omogeneità del test. Più omogeneo è un test e maggiormente univoca sarà la reazione del soggetto ai quesiti di cui il test è composto.

#### *Errore di misurazione*

Dalla definizione di fedeltà nel senso statistico risulta che essa è impostata sull'accertamento della vera varianza del test e nello stesso tempo sulla identificazione dell'estensione dell'errore nella misurazione. Con i metodi prima esposti è possibile accertare l'entità dell'errore. Le fonti dell'errore sono numerose e una volta conosciute possono essere controllate o almeno tenute presenti durante l'interpretazione dei dati ottenuti.

#### *Tipi di errore*

Sostanzialmente si possono distinguere **due tipi** di errore: *sistematico* e *asistematico*. L'errore sistematico si manifesta in modo ordinato, costante da una situazione all'altra della rilevazio-

ne; così si può notare un progressivo aumento del punteggio oppure una regolare diminuzione. Nella riapplicazione del test l'errore si manifesta talvolta in modo ciclico. Le cause che fanno aumentare il punteggio in modo sistematico sono esercizio, apprendimento, sviluppo, ecc. Una progressiva diminuzione del punteggio è da attribuirsi alla stanchezza, al calo di memoria, all'invecchiamento e simili. L'oscillazione ciclica del punteggio può essere dovuta all'umore e ad alcune abitudini. L'errore asistemático si manifesta del tutto casualmente da una situazione all'altra; ciò non significa che l'errore casuale sia dovuto al caso (esso è dovuto a cause specifiche), ma tali cause non agiscono in modo costante e ordinato. La fluttuazione dell'attenzione è un buon esempio dell'errore asistemático.

I due tipi di errore possono essere presenti nella stessa rilevazione simultaneamente.

#### *Fonti di variazione*

Brown (1976) ha identificato varie fonti di variazione tra le quali il test stesso, la sua somministrazione e il soggetto a cui il test è somministrato. La prima fonte di variazione è il test. Per quanto l'autore durante l'elaborazione si sia sforzato di preparare dei quesiti formulati in modo chiaro, una certa ambiguità del significato rimane sempre e nel caso di ambiguità, il soggetto non può rispondere in un modo coerente.

Vari errori vengono commessi anche durante la somministrazione. Alcuni non afferrano subito o completamente le istruzioni, altri commettono degli sbagli nel segnare le risposte, altri ancora non rispettano il tempo stabilito per la prova, accorciandolo o allungandolo. Durante il lavoro si verificano talvolta delle interruzioni o distrazioni. Tutto questo contribuisce, in generale, all'abbassamento del punteggio. Tali influssi si fanno sentire maggiormente se le classi sono numerose oppure se i soggetti sono molto giovani.

#### **2.2.3 Validità**

L'esposizione sulla validità comprende la definizione, l'esposizione dei tipi principali e l'accento ai tipi complementari.

La definizione più ricorrente è quella secondo la quale il test è valido se misura effettivamente quello che intende misurare (Anastasi, 1976). Questa definizione ha scarsa utilità ai fini della scelta dei metodi di verifica della validità stessa. Più precisa, ma limitativa, è la definizione secondo la quale un test è valido se è correlato con qualche informazione esterna e indipendente dal test. Tale informazione viene chiamata "criterio". Questa definizione, se da un lato ha il pregio di offrire un facile metodo per verificare la validità del test, dall'altro però si limita a prendere in considerazione un solo tipo di validità (si veda in seguito).

Il termine validità è un termine che abbraccia vari concetti, reciprocamente correlati, in dipendenza dai metodi con i quali vengono elaborati e rapportati alla situazione specifica. La

validità di un test non è un concetto universale ma legato al campione, al metodo e alla finalità, ossia alla situazione in cui viene utilizzato. Non è corretto perciò porre la domanda se un test è valido o no, ma conviene invece chiedersi se un test è valido per una determinata categoria di soggetti e per una specifica finalità.

Come la fedeltà anche la validità può essere chiarita ulteriormente in termini statistici. In tal senso la validità è intesa come proporzione della "vera" varianza rispetto alla finalità per cui il test viene usato. Se il test viene confrontato con un criterio allora la validità viene valutata in base alla varianza comune del test e del criterio; se invece il test viene valutato in base ad un costrutto psicologico, la validità viene giudicata secondo la varianza comune del costrutto (Brown, 1976).

Parlando della fedeltà abbiamo detto che accanto alla vera varianza esiste sempre anche la varianza dell'errore. *La varianza dell'errore* è presente naturalmente anche nel concetto della validità; oltre alla vera varianza e alla varianza dell'errore, esiste però un terzo tipo di varianza chiamata *irrilevante*. Questo tipo di varianza, pur essendo costante, non corrisponde alla varianza del criterio o del costrutto. In tal modo la vera varianza del test è composta dalla vera varianza e da quella *irrilevante*.

Da questa equazione si può capire facilmente che minore è la varianza dell'errore e maggiore è la fedeltà; ma la validità è alta solo se la varianza irrilevante è bassa e di conseguenza sarà alta la varianza valida. Da ciò si può comprendere anche meglio che la validità è sempre in stretto rapporto con la finalità per cui il test viene usato. Teoricamente un test può possedere un'elevata varianza "vera", ma rispetto ad un criterio oppure ad un costrutto tale varianza "vera" può essere del tutto irrilevante e quindi il test risulta inadatto a tale uso.

Gli autori distinguono **tre tipi** fondamentali di **validità**: validità basata sul **criterio**, che si divide ulteriormente in concorrente e predittiva, validità di **contenuto** e validità di **costrutto**. I tre tipi di validità sono in un rapporto operativo e logico; in altre parole essi non sono altro che tre aspetti della validità che devono essere distinti metodologicamente ma, eseguita l'analisi, devono essere ricomposti in un quadro unitario.

#### *Validità di criterio*

Una delle verifiche più semplici da condurre è quella di confrontare il test con un'informazione indipendente, come può essere per es. la diagnosi di un esperto, oppure con un fatto sociale, come per es. un successo scolastico, il rendimento professionale, ecc. L'informazione indipendente dal test oppure una situazione che ha significato sociale si chiama criterio di validità. In altre parole il test è subordinato al criterio ed è giudicato, a seconda del grado in cui corrisponde al criterio, come valido o meno. La validità di criterio consiste semplicemente nella verifica dell'accordo tra i risultati del test e il criterio.

Il criterio però può essere disponibile nello stesso tempo in cui viene somministrato il test oppure può realizzarsi nel futuro. Per questa ragione la validità di criterio si distingue ulteriormente in validità concorrente e validità predittiva. Evidentemente si tratta di validità concorrente se il criterio è disponibile nello stesso periodo in cui viene somministrato il test e di validità predittiva se l'informazione non è disponibile al momento della somministrazione oppure il fatto sociale non è ancora avvenuto (la promozione, la guarigione, ecc.). Il risultato del test viene messo a confronto con il criterio per verificare la corrispondenza tra i due dati. I due tipi di dati vengono rilevati e poi confrontati. Tale confronto viene eseguito in generale con il calcolo del coefficiente di correlazione.

La validità predittiva è definita semplicemente come capacità del test di predire il criterio. L'utilità del test che "predice" un fatto socialmente importante è notevole. In base ai dati del test si possono evitare eventi indesiderabili, come per es. l'insuccesso scolastico, l'aggravarsi di una malattia psichica, ecc. Naturalmente anche in questi casi viene operata la diagnosi oppure esplicitata la natura del criterio, con la differenza che i dati vengono riferiti all'evento futuro che non si è ancora verificato al momento della somministrazione del test. Il rapporto tra il test e il criterio viene espresso anche qui con il coefficiente di correlazione. Eseguita la somministrazione del test, occorre tenere segreti i dati fino al momento della rilevazione del criterio. A parità di condizioni i coefficienti di validità predittiva sono più bassi rispetto a quei di validità concorrente in quanto la caratteristica in esame può subire delle variazioni positive (maturazione) o negative (involuzione o regressione).

Parlando del criterio si può avere l'impressione che esso sia un'informazione "sicura" o un fatto "vero", in altre parole un'informazione valida. Da qualche decennio è stata sottolineata la necessità di dedicare attenzione all'accertamento del criterio per non vanificare l'accurato lavoro di elaborazione del test e confrontarlo con un criterio poco affidabile. Il criterio deve essere, infatti, inteso come un costrutto che richiede un'elaborazione altrettanto accurata quanto il mezzo diagnostico (Cronbach, 1971).

Vari autori hanno indicato i requisiti del criterio o hanno proposto delle classificazioni o, ancora, hanno delineato il procedimento per rilevarlo. A titolo puramente esemplificativo viene riportata la seguente classificazione:

1. Criteri semplici: ad esempio, onorario di un professionista
2. Criteri combinati: ad esempio, tasse pagate, articoli di lusso acquistati, vacanze, ecc.
3. Criteri oggettivi: ad esempio, tempo impiegato per la traduzione di una pagina
4. Criteri valutativi: ad esempio, giudizio globale sulla produzione di un operaio;
5. Criteri quantitativi: ad esempio, volume di vendita di un rappresentante
6. Criteri qualitativi: ad esempio, giudizio sulla maturità affettiva di un adolescente.

Un'altra importante distinzione è basata sulla dimensione temporale. Certi criteri non sono che una tappa per il raggiungimento del criterio ultimo. Il successo in un corso di formazione professionale è solo un mezzo per l'esercizio di un'attività professionale e viene considerato come criterio intermedio rispetto alla soddisfazione professionale che è il criterio terminale. Spesso tra i due tipi di criteri esiste solo una moderata correlazione.

#### *Validità di contenuto*

Esiste un notevole accordo tra gli autori nel definire la validità di contenuto come un esame sistematico del contenuto del test per determinare se esso rileva il campione di comportamento definito dall'autore. In altre parole, se il comportamento rilevato dal test possa essere considerato come campione rappresentativo delle situazioni reali o ipotetiche che caratterizzano l'area che il test intende misurare.

Per poter verificare questo tipo di validità occorre definire bene l'area del comportamento, confrontare i quesiti con l'area prescelta e avere la possibilità di rilevare il comportamento. La validità di contenuto è strettamente legata alla denominazione del test. La verifica della validità riguarda naturalmente la corrispondenza del titolo al contenuto (Cronbach, 1971).

#### *Validità di costrutto*

Questo tipo di validità è stato proposto negli anni '50 da Cronbach e Meehl (1955). Il costrutto è il risultato di una organizzazione delle esperienze in un quadro unitario, coerente, in cui sono note le manifestazioni, le relazioni tra le diverse componenti ed è possibile prevedere le conseguenze. Secondo un'altra versione il costrutto è un'elaborazione con metodi scientifici delle conoscenze acquisite, per definire e spiegare dei comportamenti unitari e ben differenziati; ne sono esempi abilità numerica e maturità professionale.

In modo approssimativo la validità di costrutto può essere intesa come il grado di accordo tra il test e il costrutto psicologico che il test intende misurare (Anastasi, 1976).

In realtà il processo di validazione è molto più complesso. Il costrutto rappresenta già un sistema teorico consistente e coerente. Tale sistema è formato da norme, supposizioni, ipotesi e leggi, noto sotto il termine *nomological network*, elaborato, in generale, prima della verifica della validità del test. Dalla struttura formata da una serie di proposizioni teoriche intercorrelate, il ricercatore può dedurre alcune ipotesi verificabili sperimentalmente. Se le ipotesi sono confermate è dimostrata la validità di costrutto del test. È da precisare che nel caso contrario il test non necessariamente deve ritenersi invalido (Cronbach, 1970; Fiske, 1971).

Da quanto esposto si può capire facilmente che la verifica di validità di costrutto è la più impegnativa di tutte le verifiche presentate prima. Per condurre questo tipo di verifica occorrono vaste conoscenze nel campo della metodologia della ricerca e una comprensione del pro-

cesso di elaborazione delle teorie. Inoltre i risultati ottenuti dalla verifica non trovano un'immediata applicazione pratica (come per esempio, quelli della validità di criterio), ma rappresentano un contributo al progresso della scienza.

I metodi per la verifica della validità di costrutto sono numerosi: differenze tra gruppi ben caratterizzati, calcolo dei coefficienti di correlazione e analisi fattoriale, esame della struttura interna del test, efficacia di un fattore sperimentale, ecc. Questi metodi sono adatti, naturalmente, anche alla verifica degli altri tipi di validità prima esaminati, però qui sono in funzione di un processo sperimentale più vasto (Cronbach e Meehl, 1955).

### TIPI COMPLEMENTARI

I tipi complementari o secondari di validità (rispetto ai tre esposti) sono numerosi. Qui ci si limita solo a **due: validità apparente e validità convergente e discriminante.**

#### *Validità apparente.*

Da un rapido esame del contenuto di un test, anche una persona poco esperta in materia è indotta a credere che esso misuri la dimensione per cui il test è stato elaborato. Così un test formato da figure geometriche può sembrare a chiunque un mezzo valido per la selezione dei disegnatori. La validità apparente ha la sua importanza nella prassi, in quanto fa accettare il test in relazione alla finalità per cui è usato. La validità apparente è importante anche per motivare i soggetti a impegnarsi nel test. Essa però non è mai sufficiente perché il test possa essere utilizzato.

#### *Validità convergente e discriminante.*

La verifica di tale validità consiste nel fatto che due test che rilevano vari tratti di personalità devono dare risultati simili quando rilevano lo stesso tratto e differenti invece quando rilevano tratti differenti. In altre parole i due (o più test) devono convergere sui tratti identici e divergere su quelli differenti, per discriminare in tal modo i differenti tratti. Questo tipo di validità è usato in modo particolare per la verifica della validità di costrutto (Campbell e Fiske, 1959).

### 2.2.4 Norme

Le norme sono delle guide per l'interpretazione del punteggio grezzo ottenuto al test. Esse sono prevalentemente quantitative ed elaborate in base alle medie ottenute dai soggetti ai quali è destinato il test. Dovrebbero essere basate sulle risposte al test date da un campione rappresentativo di soggetti. In generale, le norme sono basate su soggetti occasionali, talvolta residenti in varie regioni.

Ci sono parecchi tipi di norme. Qui descriviamo solo le **norme di posizione** e **norme di sviluppo**. Le prime collocano il soggetto nel gruppo al quale appartiene e le seconde stabiliscono il grado evolutivo individuale della caratteristica rilevata dal test.

### **Norme basate sul gruppo**

Il modo più semplice per interpretare il punteggio grezzo di un test è quello di confrontarlo con i punteggi ottenuti da un gruppo di soggetti ben scelti. I risultati di tutti i soggetti del gruppo vengono sommati e viene calcolata, in generale, la media aritmetica con lo scarto quadratico medio. Il singolo risultato viene poi confrontato con la media del gruppo, determinando di quanto si discosta da essa. Il punteggio del singolo assume quindi significato in relazione alla "qualità" del gruppo. Si può allora comprendere come la composizione del gruppo sia di fondamentale importanza.

Nell'elaborazione delle norme non si possono ignorare alcune variabili che influiscono sui risultati di molti test, come: sesso, età, grado di istruzione, tipo di occupazione, livello socio-economico, luogo di residenza, razza, ed altre ancora. L'autore del test dovrebbe aver verificato, almeno per alcune variabili, se l'influsso sui risultati del test è significativo. Nel caso positivo dovrebbero essere disponibili norme differenziate (ciò viene fatto frequentemente per età e per sesso).

### **Norme di sviluppo**

Molte caratteristiche psicologiche hanno uno sviluppo sistematico e progressivo nell'arco evolutivo dei soggetti. Uno degli obiettivi dell'educazione è quello di favorire lo sviluppo di tali caratteristiche. Per poter raggiungere questa finalità è necessario conoscere come avviene tale processo nei singoli. Le norme di sviluppo possono contribuire notevolmente a questa conoscenza. Lo sviluppo può essere considerato in rapporto al tempo per le caratteristiche che si sviluppano spontaneamente (intelligenza, interessi) oppure in rapporto alle istituzioni scolastiche (apprendimento delle varie materie). In tal modo le norme si articolano in *Norme per età* e in *Norme per classe*.

*Norme per età* - Un modo molto semplice per valutare lo sviluppo psichico, particolarmente quello intellettuale, è basato sull'Età Mentale (EM). Denominata così da Binet durante l'elaborazione della sua scala di intelligenza, l'Età Mentale è in stretto rapporto con l'Età Cronologica. Infatti l'Età Mentale corrisponde alla capacità dei soggetti di una determinata età di dare risposte giuste, le quali sono poi considerate tipiche di tale età.

Una variante dell'Età Mentale è il Quoziente di Intelligenza (QI). Il quoziente esprime numericamente il rapporto tra Età Mentale ed Età Cronologica (EC). L'Età Mentale viene divisa per l'Età Cronologica e il prodotto viene poi moltiplicato per 100. Nel caso in cui l'EM è perfettamente uguale all'EC il QI è uguale a 100. Si tratta di un risultato prettamente medio.

*Norme per classe* - Un secondo tipo di norme, particolarmente adatte alle prove che intendono rilevare l'apprendimento nelle varie materie, è quello in cui i soggetti sono raggruppati in

base al tipo di scuola e alla classe che frequentano al momento della verifica. Il modo di procedere è identico a quello per l'elaborazione delle norme per età. Viene estratto il campione rappresentativo dalle classi appartenenti ai vari livelli scolastici (per esempio, dalla prima alla terza media), si somministra la prova, si calcolano le medie aritmetiche, si riportano sul grafico e per interpolazione si può valutare se l'apprendimento del singolo è in anticipo o in ritardo (per esempio, un allievo del primo anno di scuola media può possedere già i contenuti del secondo anno).

Riassumendo possiamo affermare che questi due tipi di norme hanno il vantaggio di essere facilmente comprensibili e permettono di stabilire il grado di sviluppo della dimensione misurata oppure il livello di apprendimento del soggetto. Presentano però l'inconveniente di non possedere l'unità di misura a tutti i livelli di età e di istruzione e quindi di non essere direttamente confrontabili. Tale inconveniente viene superato con una trasformazione lineare del punteggio grezzo, adattandolo alla curva della distribuzione normale ed ottenendo in tal modo il punteggio normalizzato.

### ***Rilievi conclusivi sulle norme***

Esaminati i vari tipi di norme, conviene fare qualche rilievo che può essere utile particolarmente per la prassi. *Prima di usare le norme riportate nel manuale di un test occorre stabilire se il soggetto o il gruppo che si intende sottoporre al test rassomiglia nelle principali caratteristiche al campione scelto dall'autore.* Si tengano presenti le caratteristiche più note: età, sesso, razza, livello di istruzione, livello socio-economico e residenza (urbana o rurale). In questo modo saranno controllate le variabili più importanti. Per molte altre variabili ci saranno certamente tra il campione e il gruppo delle differenze indesiderate, come per es. motivazione, esercizio, ecc., difficilmente verificabili.

Con la diffusione della cultura e i mutamenti sociali le norme si modificano. Per le attitudini si è constatata un'accelerazione nello sviluppo mentale delle nuove generazioni, per i test di personalità un cambiamento di mentalità (si pensi alle varie "rivoluzioni" sociali) Si sostiene la necessità di revisionare le norme a distanza di una generazione, cioè ogni 20 o 25 anni.