

CAP. XX

INTERPRETAZIONE DEI RISULTATI

Manoscritto dell'A. con integrazioni 2021. Le parti in carattere più piccolo sono di interesse prevalentemente teorico.

L'interpretazione dei risultati dei test viene considerata da molti un processo più "artistico" che "scientifico"; si presume che gli psicologi apprendano quest'arte nell'ambito di insegnamenti "non tecnici", quali la Psicodiagnostica (insegnamento attualmente non impartito nelle università statali italiane) o gli insegnamenti psicologici riferiti a specifici settori applicativi (psicologia clinica, dell'apprendimento, dell'orientamento, del lavoro, ecc.), inquadrando i problemi diagnostici in stretta connessione con la strutturazione complessiva dell'intervento. La progressiva diffusione delle interpretazioni computerizzate dimostra però non solo la possibilità di un approccio tecnico al problema dell'interpretazione, ma anche la forte richiesta degli operatori per un sostegno tecnico all'interpretazione. Questo capitolo si propone di offrire alcune linee di guida in un'ottica metodologica generale, l'unica possibile quando si deve tener conto di un ampio ventaglio di variabili. Vuole offrire un sussidio ai docenti che si pongono l'obiettivo di far acquisire agli studenti le capacità tecniche che sono premessa e condizione "sine qua non" per un'interpretazione dei risultati che sia profonda e ricca di intuizione, ma non manchi di realismo.

Nell'esposizione verranno privilegiate le problematiche riguardanti l'interpretazione dei risultati di singoli soggetti, riservando solo alcuni cenni alle problematiche specifiche poste dall'interpretazione dei test somministrati a grandi campioni in vista di decisioni amministrative o politiche (cfr. AERA, APA, NCME, 1985, pp.67-69). I paragrafi dedicati al riesame dei concetti di validità e attendibilità in un'ottica pratica sono focalizzati intorno a un concetto di base: nessun test è valido per tutte le situazioni e per tutte le popolazioni (cfr. anche: AERA, APA, NCME, 1985, Standard 1.2) e pochissimi test sono del tutto inutili; quel che importa è scegliere e usare gli strumenti che con una data popolazione e in date circostanze consentono di raccogliere informazioni utili e non distorte.

(A) INTERPRETARE TENENDO CONTO DELLE CARATTERISTICHE DEI TEST

1. Interpretare tenendo conto dei dati sull'attendibilità

L'attendibilità, di cui si è ampiamente trattato nel cap. VI, è una condizione "sine qua non" per la validità: un test attendibile può essere più o meno valido, ma se un test non è attendibile è impossibile accertarne la validità, perché non sempre misura la stessa cosa. Nell'interpretare i risultati di un test somministrato a scopo applicativo le stime dell'attendibilità ottenute con approcci diversi vanno considerate distintamente una dall'altra, anche perché non tutti i dati sull'attendibilità sono ugualmente utili per qualsiasi tipo di applicazione.

Per i test **"a risposta aperta"** (non importa se cognitivi o proiettivi, individuali o di gruppo, ecc.) è fondamentale accertare qual è il grado di accordo fra "correttori" nel siglare le risposte (**"oggettività"**). A questo proposito il manuale dovrebbe riferire dati particolareggiati riguardo ai sussidi usati per la siglatura e al periodo di addestramento a cui erano stati sottoposti i correttori. I dati espressi in "percentuali di accordo" (cfr. cap. VI) sono difficilmente interpretabili, nonostante le apparenze, e di solito "gonfiano" il grado di accordo. L'assenza di dati riguardo all'oggettività di scoring, se il test è "a risposta aperta", non depone a favore della serietà del test.

I dati sull'**omogeneità** del test sono i più frequentemente riportati. Sono indicativi della "purezza" teorica della misura. Ai fini pratici, vanno interpretati congiuntamente ai dati sulla validità riferita a criteri, dato che spesso è stata osservata una diminuzione della validità predittiva corrispondente all'aumento dell'omogeneità. Questo fenomeno non è casuale. Quando si verifica la predittività di un insieme di misure mediante un'equazione di regressione multipla, è noto che i migliori risultati si ottengono utilizzando predittori indipendenti l'uno dall'altro. Questa situazione, con riferimento a un solo test, corrisponde al caso in cui la correlazione media fra gli items è zero, e non certo al caso di buona omogeneità (cfr. Raven, 1989a).

Se il test si presenta in più **forme parallele** va controllato preliminarmente se si tratta di forme parallele "classiche"

(costruite ricercando l'equivalenza fra i contenuti degli stimoli e sottoposte a verifiche empiriche per l'accertamento del parallelismo) o di forme parallele "per casualizzazione" (cfr. cap. VI). Alle forme parallele "per casualizzazione" non si applicano le proprietà metriche attribuite alle forme parallele "classiche". Va tenuto presente che la verifica del parallelismo riguarda soprattutto l'analogia fra deviazioni standard e l'intensità delle correlazioni; le differenze tra punteggi medi sono considerate trascurabili, in quanto possono facilmente esser corrette con un'adeguata standardizzazione. Quando le forme parallele vengono utilizzate sullo stesso campione (p.es.: distribuendo prove a file alterne nella stessa aula d'esame, per controlli "prima" e "dopo" un intervento, ecc.) bisogna fare attenzione e controllare su un campione esterno al proprio se le medie delle due forme si equivalgono; in caso negativo, vanno apportate le correzioni necessarie a non introdurre distorsioni.

I dati sulla **stabilità** dei risultati sono importanti per qualsiasi applicazione implicante predizioni o controlli a distanza di tempo: scuola (AERA, APA, NCME, 1985, Standard 8.8), concorsi (AERA, APA, NCME, 1985, Standard 11.3), orientamento (AERA, APA, NCME, 1985, Standard 9.7), selezione del personale, diagnosi... . Questo tipo di verifica pone tuttavia problemi metodologici di non facile soluzione (cfr. cap. VI). Va anzitutto controllato se test e retest sono stati eseguiti con l'identica forma di test o con due forme parallele. Se è stata impiegata in ambedue i casi la stessa forma, le distorsioni indotte dall'apprendimento e dalla memorizzazione degli stimoli sono molto forti e, peraltro, difficilmente quantificabili e correggibili (cfr. cap. VI e discussioni riguardanti alcuni test proiettivi nel cap. XVIII). Se sono state usate due forme parallele, prima d'interpretare i dati sulla stabilità vanno controllati i dati tecnici sul parallelismo, relativamente a ciascuna delle variabili misurate dal test.

I dati sull'attendibilità danno informazioni utili anche per alcune applicazioni che spesso interessano lo psicologo pratico:

- l'individuazione di differenze tra variabili "significative" all'interno del "**profilo**" di un soggetto (cfr. esempi pratici nel cap. VI);

- l'abbreviazione del test, mediante l'applicazione della formula di Spearman-Brown (in questo caso devono però essere disponibili anche dati analitici item per item, risultanti da item analysis o da analisi fattoriale).

Quando i dati sull'attendibilità sono espressi da coefficienti di correlazione, il livello minimo ritenuto accettabile per gli usi applicativi è fra $r = 0,80$ e $r = 0,90$ (cfr. Ziskin e Faust, 1988, p.547). Questo limite viene solitamente superato dai "buoni" test d'intelligenza (controllare però, per i test adattati da originali stranieri, se l'attendibilità è altrettanto buona per l'adattamento italiano). Quando si danno risultati individuali è comunque buona norma sottoporre a tutti i possibili controlli i punteggi bassi nei test cognitivi (AERA, APA, NCME, 1985, Standard 6.11) e i punteggi che indicano una patologia in atto.

I dati sull'attendibilità, come quelli sulla validità, vanno considerati separatamente, variabile per variabile, anche quando il test sembra unitario: p.es., per test come il Rorschach e lo Holtzman è stato ripetutamente accertato che alcune variabili sono più "oggettive" da misurare e più stabili nel tempo rispetto ad altre: cfr. cap. XVIII.

2. Predizione clinica e statistica

a) Cenni storici

La polemica sull'antitesi "predizione clinica o statistica" maturò nel secondo dopoguerra, stimolata da studi che segnalavano l'inferiorità predittiva dell'"intuizione clinica" rispetto a semplici equazioni di regressione (cfr. Sarbin, 1943; 1944) ed ebbe una notevole risonanza fra il 1950 e il 1970 (cfr. Meehl, 1954; Holt, 1958, 1970; Wiggins, 1973). In una famosa monografia, Meehl confrontò i risultati di predizioni formulate con metodi clinici e statistici, desunti da una ventina di studi (tutti quelli reperiti nella letteratura statunitense dell'epoca), evidenziando una netta superiorità delle predizioni statistiche. La monografia di Meehl ebbe una larga risonanza, suscitò molte polemiche e numerose ricerche sul tema. La tab. XX.1 riassume i risultati raccolti da Meehl in più riprese (Meehl, 1954; 1957; 1965).

TAB. XX.1

Sintesi di studi in cui si confrontano predizioni cliniche e statistiche (da Wiggins, 1973)

Fonte	Numero degli studi	Ambito predittivo	Le predizioni più accurate sono:		
			Statistiche	=	Cliniche
Meehl (1954)	16-20	Successo nell'istruzione accademica o militare; recidive e violazione della libertà vigilata; remissioni di psicosi	11	8	1
Meehl (1957)	27	Successo nell'istruzione accademica o militare; recidive e violazioni della libertà vigilata; remissioni di psicosi; descrizioni di personalità; risultati di psicoterapie	17	10	0
Meehl (1965)	51	Successo nell'istruzione accademica o militare; recidive e violazioni della libertà vigilata; remissione di psicosi; descrizioni di personalità; risultati di psicoterapia; reazione a un trattamento di shock; nosologia psichiatrica formale; successo e soddisfazione nel lavoro; diagnosi medica (non psichiatrica)	33	17	1

Le cause della scarsa efficienza delle predizioni cliniche sono riconducibili all'imperfetto controllo delle distorsioni che operano nella conoscenza interpersonale "ingenua" (cfr. cap. I): convinzioni infondate riguardo alla correlazione fra certi tipi di prestazione e certe caratteristiche psicologiche (p.es.: tra occhi grandi nel disegno della figura umana e paranoia: "correlazioni illusorie", cap. XVIII), indicatori di anormalità individuati solo in base alla frequenza in soggetti anormali, senza controllare se sono altrettanto frequenti nei normali, ricorso a descrizioni ovviamente vere perché generiche e riferibili a qualsiasi persona, fiducia in indicatori non controvalidati, generalizzazioni indebite da una popolazione all'altra (queste e altre pecche dei "rapporti psicologici" sono elencate e discusse in Ziskin e Faust, 1988, p. 547 ss.).

La controversia - posta in termini di cruda contrapposizione - contribuì a provocare conseguenze gravi per la professione clinica: escludere il clinico dalla psicodiagnostica e la psicodiagnostica dalla formazione clinica, anziché cercar di trovare l'impiego ottimale per i clinici, dopo una preparazione adeguata (cfr. Holt, 1986). Cessata la polemica, attualmente l'opinione prevalente è che il clinico sia una buona fonte di informazioni aggiuntive (cfr. già Sawyer, 1966) ma che un programma computerizzato sia capace di interpretazioni e predizioni più valide di quelle espresse dalla media dei clinici. Viene peraltro riconosciuta e privilegiata la distinzione qualitativa tra approccio clinico e approccio statistico, ed evidenziata la necessità di ambedue (cfr. Sarbin, 1986). Anche la computerizzazione viene progettata non in contrapposizione al processo clinico, ma come simulazione e razionalizzazione di esso (cfr. cap. XIV, in riferimento ai programmi computerizzati per l'interpretazione del MMPI e, più oltre, il paragrafo sui rapporti psicologici generati dal computer; v. anche: Holt, 1986; Einhorn, 1986). Viene auspicata da alcuni l'adozione di modelli epistemologici come la teoria dei sistemi e l'analisi della decisione, che giustifichino filosoficamente l'approccio clinico (Holt, 1986).

b) Esigenze di base nell'esame critico dei dati sulla predittività riferiti nel Manuale

Se ci si orienta verso una predizione-interpretazione in termini statistici, è necessario rispettare alcune condizioni preliminari:

- per interpretare i risultati di qualsiasi test bisogna riferirsi necessariamente a un **manuale** in cui siano riportate le prove empiriche della validità e dell'attendibilità delle interpretazioni consentite; questa norma vale anche se si dispone di un'interpretazione computerizzata dei risultati (AERA, APA, NCME, 1985, Standard 6.13); bisogna

diffidare dei test che non sono accompagnati da un manuale accurato e aggiornato sotto questo aspetto;

- se l'interpretazione comporta una predizione (p.es.: del successo accademico, della riuscita della terapia, ecc.) bisogna disporre di prove adeguate della **validità predittiva** di quel particolare test in quella particolare modalità d'uso: p.es., per la predizione dei risultati accademici non sono sufficienti i coefficienti di correlazione: servono anche equazioni di regressione ed errori standard della stima, oppure "tavole di attesa" (AERA, APA, NCME, 1985, Standard 8.9);

- **se il test è tradotto** da un'altra lingua, la validità e l'attendibilità dell'adattamento devono essere nuovamente verificate (AERA, APA, NCME, 1985, Standard 13.4);

- i dati sulla validità vanno esaminati analiticamente, per capire se la predittività varia in funzione di **caratteristiche anagrafiche dei soggetti** (p.es.: sesso, appartenenza etnica, età: cfr. AERA, APA, NCME, 1985, Standard 7.6 e, per eventuali distorsioni a danno di una minoranza: AERA, APA, NCME, 1985, Standard 9.5);

- la predizione ottenuta usando il test dovrebbe essere più accurata di quelle basate su informazioni attuariali elementari, quali la frequenza del fenomeno nella popolazione; ancor più utile sarebbe quantificare l' **incremento alla validità predittiva dato dall'uso di un particolare test** (Ziskin e Faust, 1988, pp. 549-50).

La maggior parte dei dati riportati nei manuali riguardano la validità del test rispetto a criteri "concorrenti" o "predittivi" (cfr. cap. V). Nonostante molti dati riportati siano spesso inutili, vale la pena esaminarli criticamente per vedere quel che di utile può essere nascosto nella selva di numeri scarsamente commentati e spesso riferiti a campioni inappropriati.

Il primo dato da controllare riguarda proprio i **campioni su cui la validazione è stata compiuta**:

- se si tratta di campioni stranieri o esaminati oltre dieci anni fa, qualunque conclusione interpretativa è solo ipotetica;

- nel caso dei test di profitto, va controllato anche il mese di somministrazione: i livelli di profitto variano nettamente lungo l'anno scolastico (p.es.: secondo la distanza dalle vacanze estive);

- se un test è stato validato su soggetti normali, i dati sulla validità non possono essere generalizzati a soggetti portatori di handicap o appartenenti a minoranze.

In secondo luogo bisogna **riflettere sull'appropriatezza e la validità dei criteri**:

- in linea di massima, i confronti con altri test sono utili solo se riferiti a test che hanno una qualche connessione per contenuto con quello esaminato;

- nell'interpretare un risultato, si tenga conto dell'analogia fra il criterio usato e le situazioni della vita reale: p.es. un test che misura la memoria facendo ripetere sillabe senza senso può essere usato come stima dell'efficienza di memoria nella vita reale - supponiamo nell'ambito di una perizia giudiziaria - solo se corredato da prove empiriche della correlazione tra prestazioni al test ed efficienza di memoria nella vita quotidiana (cfr. Ziskin e Faust, 1988, p. 553);

- **i criteri "scolastici"** vanno considerati in riferimento alle strutture educative e alla prassi docimologica (p.es.: sono poco generalizzabili alla nostra situazione test validati rispetto a esami costituiti da prove scritte a scelta multipla, secondo la prassi anglosassone); bisogna accertare che gli obiettivi perseguiti dagli insegnanti degli alunni esaminati coincidano con quelli a cui si riferiscono i criteri su cui è stato validato il test (cfr. AERA, APA, NCME, 1985);

- **i criteri clinici** vanno valutati in base almeno alla validità di contenuto (p.es.: chi e quanti erano i clinici che hanno dato e unificato le diagnosi-criterio); le categorie nosografiche usate per la diagnosi - criterio devono essere esplicitate e definite, e dev'essere specificato il metodo usato per arrivare alla diagnosi (AERA, APA, NCME, 1985, Standard 7.2);

- **per le applicazioni al lavoro**, il criterio dovrebbe essere formulato in termini concreti; si dovrebbe dare una descrizione particolareggiata del lavoro usato come criterio e degli elementi in base ai quali il criterio è stato considerato pertinente per quel dato test (AERA, APA, NCME, 1985, Standard 10.3). L'accuratezza della descrizione è fondamentale per valutare se i dati sulla validità predittiva possono essere generalizzati a professioni affini, dato che spesso si opera in riferimento a professioni molto specifiche ed è impossibile trovare dati su un lavoro esattamente identico (AERA, APA, NCME, 1985, p.59). I dati sulla validità predittiva che assumono come criteri la riuscita in professioni di livello elevato sono spesso basati su campioni poco numerosi e molto selezionati

(AERA, APA, NCME, 1985, p. 63) e quindi risultano poco attendibili.

c) Interpretazione in base ai dati sulla predittività del test riferiti nel Manuale

L'esame critico dei criteri solitamente induce a scartare come inutili o non interpretabili molti dati numerici riferiti nei manuali.

I dati riferiti a criteri significativi possono essere interpretati in prima istanza individuando o calcolando: (a) **l'errore tipico** delle misure, se si tratta di differenze tra campioni contrastanti o fra "prima" e "dopo" un trattamento (dell'errore tipico si parla anche più oltre, in riferimento all'interpretazione delle scale, nel ≈ 4) (b) **la percentuale di varianza comune** se sono espressi mediante coefficienti di correlazione (basta calcolare il quadrato di r). Per i coefficienti di correlazione va compiuto un ulteriore controllo: se si dice che sono stati corretti "per l'attenuazione" (cfr. cap. VI) il loro valore è stato amplificato (indebitamente, a parere di molti studiosi).

Nelle applicazioni diagnostiche e forensi si deve accertare che il test produca misure sufficientemente precise per la diagnosi individuale (cfr. cap. VI; Ziskin e Faust, 1988).

Se i dati sulla validità del test vengono letti in funzione di problematiche particolari, vanno prese in considerazione anche le condizioni che fanno variare la validità del test "differenzialmente" ("**predicibilità differenziale**"): p.es., se il test viene utilizzato per fare predizioni su caratteristiche personali "non cognitive", vanno tenuti presenti anche i dati sulla stabilità al retest (cfr. cap. VI) e le caratteristiche personali che fanno variare "differenzialmente" la predicibilità (cfr. cap. V); se si lavora con test cognitivi (non importa se "verbali" o "non verbali") e con campioni appartenenti a gruppi socio-culturali minoritari, bisogna controllare se e quanto i test siano sensibili a distorsioni socioculturali ed emotive (cfr. cap. V e IX, oltre a quanto detto altrove in questo capitolo. Le norme professionali statunitensi consigliano alle scuole di coordinarsi tra loro per mettere insieme, nel giro di qualche anno, dati sufficienti a stimare la validità predittiva dei test per alunni appartenenti a minoranze: AERA, APA, NCME, 1985, Standard 8.10).

I dati sulla validità rispetto a criteri sono prioritari in psicologia del lavoro quando interessa predire in quale misura un candidato farà bene il lavoro a cui aspira. In questo caso va indicata la relazione matematica fra i punteggi nel test e un qualche indice numerico di successo nel lavoro (cfr. Lawshe, 1985).

Se si utilizzano i test per assunzioni in posti di lavoro e una parte degli aspiranti appartiene a minoranze linguistiche, bisogna prendere i dovuti accorgimenti perché la componente linguistica influisca solo relativamente a criteri di lavoro per cui la padronanza linguistica è oggettivamente importante (cfr. AERA, APA, NCME, 1985, Standard 13.5 e 13.7).

Nelle **applicazioni diagnostiche** va fatta una netta differenza fra test di "screening" (da usare solo come pre-test: cfr. AERA, APA, NCME, 1985, Standard 6.12), per i quali è sufficiente sia documentata la validità nel distinguere soggetti "normali" da soggetti "anormali" o "a rischio", e test impiegati per la diagnosi differenziale, per cui deve essere documentata la capacità del test di distinguere fra prestazioni di due o tre gruppi nosografici ben identificati (AERA, APA, NCME, 1985, Standard 7.3). Analogò è il problema che si pone nelle applicazioni all'**orientamento** e alla **psicologia del lavoro**: quando si deve decidere fra due o tre diverse opzioni, ci si dovrebbe basare su dati specifici che dicono la differenza di predittività del test fra quelle opzioni (AERA, APA, NCME, 1985, Standard 10.2). Se poi si usa stabilmente un insieme di test per un dato scopo (l'esempio tipico è dato dalle "batterie" per l'orientamento), bisogna accertare che esistano prove sulla validità predittiva di quel particolare insieme (AERA, APA, NCME, 1985, Standard 8.11).

3. Interpretare in funzione delle caratteristiche tecniche del test: comprendere qual è la variabile misurata

I termini psicologici che descrivono caratteristiche della personalità umana (nei suoi aspetti cognitivi e non cognitivi, "normali" e "patologici") e, conseguentemente, le variabili misurate dai test, sono imprecisi. Nei capitoli precedenti le ragioni dell'imprecisione sono state illustrate con riferimenti allo sviluppo storico delle teorie. Informazioni specifiche dovrebbero esser date dal manuale del test, in forma più o meno documentata. La completezza e la chiarezza dei Manuali è soggetta a eccezioni:

- i dati attinenti alle verifiche di "validità del costrutto" (cfr. cap. VII), che si accumulano attraverso lunghi periodi di tempo, vanno spesso reperiti altrove, soprattutto per i test in circolazione da decenni (p.es. Wechsler, Rorschach,

MMPI, 16 PF, ecc.);

- se il Manuale è stato scritto molti anni fa, le categorie psicologiche o nosologiche non possono essere quelle elaborate negli ultimi vent'anni e attualmente in uso;

- i termini usati da Autori che si riferiscono a teorie diverse hanno significati diversi anche quando usano parole simili: ad esempio, il termine "estroversione" e le sue varianti ("estraversione", "extroversione", "estratensione", ecc.) possono essere correttamente compresi solo se si conosce il nome dell'Autore alla cui teoria è riferito.

Nelle trattazioni riguardanti i singoli test si è cercato di dare una sintesi della letteratura e indicazioni anche bibliografiche per chi desidera approfondire la conoscenza dello strumento. Qui vengono presentate alcune problematiche generali.

a) Inquadrare correttamente il tipo di variabile misurato

Nel cap. I si è accennato all'esigenza di conoscere qual è il tipo di variabile oggetto della misurazione: se i risultati del test esprimono una classificazione in "tipi" o valutano l'intensità di presenza di un "tratto", se misurano uno stato momentaneo o una caratteristica permanente, e così via. Tenendo conto delle caratteristiche della maggior parte dei test attualmente commercializzati in Italia, i problemi più rilevanti ci sembrano i seguenti.

Per i **test cognitivi** bisogna fare attenzione anzitutto al livello di generalizzazione della misura ottenuta (cfr. parte III): al livello più alto di generalizzazione si collocano i test di "fattore g" (descritti nel cap. X), che sostanzialmente misurano la capacità dei soggetti di affrontare con successo problemi cognitivi espressi in modo insolito, seguiti da vicino dai test di "abilità generale" (descritti nel cap. IX), che includono anche problemi già sperimentati dai soggetti nella vita quotidiana e nella scuola. I test di abilità e di attitudine (per la distinzione fra questi termini, cfr. paragrafo seguente) misurano invece la capacità di risolvere problemi caratterizzati da tipi specifici di contenuto e, in alcuni casi, dall'attivazione di processi che sono implicati solo in quei tipi di problema. I **test di profitto** (cfr. cap. XII: in particolare, la fig. XII.1 e il relativo commento) sono ancor più specifici, e vanno interpretati tenendo presenti gli obiettivi e i contenuti dell'insegnamento impartito ai soggetti, anche quando la loro apparenza è simile a quella di test di abilità. Ad esempio, un test X di abilità verbale può essere interamente composto da items che chiedono di trovare sinonimi e antonimi di parole, scelte in modo che i significati si distribuiscano fra un'ampia gamma di contenuti e lungo un'ampia gamma di difficoltà: dà una misura di abilità verbale, generalizzabile a molti compiti in cui si chiede di comprendere concetti espressi con parole. Un test Y può presentarsi in forma identica al test X, ma essere composto di items scelti all'interno di un'area di contenuti ristretta (p.es.: termini attinenti ai colori, alla natura in autunno, alla navigazione...), su cui sono stati compiuti appositi interventi didattici, e con un gamma di difficoltà adeguata al livello degli interventi compiuti: dà una misura di profitto, che al massimo della generalizzazione può essere interpretata come stima della padronanza lessicale in un settore. Anche i **test "neuropsicologici"** (cfr. cap. XIII), qualunque sia la loro apparenza (p.es.: anche se somigliano a test di fattore g o a test di profitto) vanno interpretati solo relativamente al deficit specifico per la cui diagnosi sono stati validati.

Quanto ai test "di personalità" vanno chiaramente distinti, ai fini dell'interpretazione, alcuni tipi di misure (cfr. cap. I):

(a) classificazioni riferite a una **tipologia** (p.es.: il questionario junghiano di Myers-Briggs, test caratterologici ispirati a Heymans-Le Senne o a Sheldon, classificazione secondo il TRI nel Rorschach): in questi casi è curata al massimo la distinzione fra un "tipo" e l'altro, e sono considerate irrilevanti eventuali differenze tra individui all'interno dello stesso tipo: interpretazioni più sfumate (p.es.: individuazione di "tipi misti") vanno compiute riferendosi esclusivamente alle indicazioni, documentate da studi, fornite dal Manuale; se la classificazione non è solo dicotomica, ma espressa da un punteggio (p.es.: Sheldon), si tenga presente che il punteggio esprime l'intensità di presenza di una caratteristica in riferimento al numero dei sintomi presenti in assoluto, senza riferimento all'intensità di presenza delle caratteristiche in altri soggetti: non è quindi confrontabile con misure "di tratto" (anche se la denominazione fosse identica), che invece esprimono l'intensità in riferimento alla media;

(b) misure di **stati** relativamente instabili; la variabile psicologica su cui sono stati compiuti più studi al riguardo è l'ansia (cfr. cap. XIV), ma è presumibile che anche per altre variabili (p.es.: depressione) la distinzione stato - tratto sia importante; vanno letti accuratamente i dati sulla stabilità della misura al retest ed eventuali altre

informazioni riportate nel Manuale;

(c) misure di **tratti** relativamente stabili della personalità o di **caratteristiche psicopatologiche** che non hanno remissione spontanea nell'arco di pochi giorni (la maggior parte dei test misura variabili di questo tipo); vanno interpretate tenendo conto delle indicazioni generali date in questo capitolo.

b) Individuare termini a cui vengono attribuiti significati diversi

Le variabili misurate dai test di largo uso applicativo sono in genere costrutti complessi, in ciascuno dei quali confluiscono più variabili "elementari": ad esempio, nell'Estraversione di Eysenck sono state individuate le componenti "ricerca di sensazioni" e "impulsività". Alcune variabili, inoltre, sono state studiate da decenni e in prospettive teoriche diverse: fattori questi che hanno indotto un ulteriore incremento nella varietà dei significati. Esempi tipici sono l'assai confusa distinzione fra i termini "abilità", "attitudine" e "intelligenza" (cfr. capp. IX e X) e la scarsa chiarezza con cui vengono designate alcune patologie (v. appresso). La denominazione della variabile è un'interpretazione riduttiva e spesso ambigua: per chiarirne il significato bisogna esaminare i contenuti degli items, i dati sulla validità rispetto a criteri noti e attendibili, i dati sulla validità del costrutto.

Quando la variabile designa una sindrome patologica, va controllato qual è il sistema nosografico a cui l'A. del test si riferisce. Attualmente i sistemi più usati in campo internazionale sono l'ICD e il DSM (cfr. cap. XIV), fra i quali esistono connessioni evidenziate dagli stessi curatori. Sono anche reperibili lavori che consentono di stabilire rapporti tra questi sistemi e altri di recente pubblicazione (cfr. Robins e Helzer, 1986; Mezzich e von Cranach, 1988). Il significato dei termini riferiti a nosografie diverse da queste va chiarito consultando le opere citate nel Manuale del test, senza mai dare per scontato che una denominazione nota abbia significato conosciuto. Le etichette diagnostiche che hanno una storia più lunga (p.es.: schizofrenia, ansia, depressione, psicosi) sono le più ambigue nell'uso comune e nei test. Ad esempio, sottoponendo un elenco di 94 caratteristiche della schizofrenia a un campione di 183 professionisti (psicologi clinici, psichiatri e assistenti sociali psichiatrici), sono state trovate otto distinte componenti fattoriali (Fitzgibbons e Shearn, 1972), ossia concezioni chiaramente differenziabili per attribuzione di etiologia, dipendenza più o meno forte dalla fenomenologia descritta da Bleuler, severità della prognosi, convinzione che la schizofrenia sia una malattia, che sia un disordine mentale, ecc. Si è anche visto che le discordanze interpretative riguardo a questa sindrome possono avere un base culturale (cfr. Frighi, 1985), il che induce molta cautela soprattutto nei confronti di studi non europei e non "occidentali". Per l'ansia, la gamma delle descrizioni sintomatiche è tale da assimilarla da un lato alla "emotività" come tratto temperamentale contrapposto alla riservatezza schizoide, dall'altro a una patologia nevrotica. Si potrebbe continuare con molte altre esemplificazioni, che non arriverebbero però mai a coprire la totalità delle denominazioni ambigue. Si può quindi soltanto ribadire la necessità di raccogliere dati di fatto sul significato che un particolare termine ha in riferimento a un dato test. E' ovvio che gli stessi problemi si pongono anche quando la diagnosi non si basa sui test.

c) Caratteristiche degli stimoli e validità del contenuto

I manuali di molti test commercializzati in Italia non contengono una documentazione ampia e aggiornata sulla validità del costrutto misurato. Soprattutto in questi casi, è fondamentale sfruttare al massimo le informazioni sulla validità del contenuto, che l'utente può ottenere anche con un'analisi personale. E' evidente che nessun test può contenere stimoli che sondano tutti gli aspetti di un costrutto psicologico o tutti i punti di cui si compone l'apprendimento di una disciplina scolastica. Per una corretta interpretazione è però importante sapere quali punti sono tenuti presenti.

(1) Caratteristiche del contenuto rilevanti per l'area applicativa

Nell'ambito dei **test di profitto** la validità di contenuto è di primaria importanza (una discussione della problematica in riferimento al sistema scolastico statunitense si trova in: Tittle, 1979). In particolare, se un test è utilizzato come parte sostanziale di un *esame avente valore legale*:

- "la gamma dei contenuti, sia del test sia del programma previsto per quel particolare livello scolastico/accademico, dovrebbe esser descritta in modo abbastanza particolareggiato da consentire, senza compromettere la riservatezza degli stimoli, di valutare la corrispondenza fra contenuti del test e del programma" (AERA, APA, NCME, 1985, Standard 8.4);

- "dovrebbe essere documentato che il test comprende quesiti solo sulle nozioni (specifiche e generali), le capacità e le abilità che gli studenti hanno avuto l'opportunità di apprendere" (AERA, APA, NCME, 1985, Standard 8.7);
- se il test da utilizzare come parte sostanziale di un esame avente valore legale è stato elaborato da un ente pubblico (p. es. un distretto scolastico, la regione, lo stato) "le guide per l'utente e il rapporto tecnico sul test dovrebbero essere messi a punto e diffusi" (AERA, APA, NCME, 1985, Standard 8.5);
- per gli esami di abilitazione professionale i contenuti più importanti sono quelli che si riferiscono al costrutto di cui è documentata l'appropriatezza alla professione in questione (AERA, APA, NCME, 1985, p. 63); in ogni caso, è essenziale che siano date prove per documentare la connessione fra i contenuti oggetto del test e le conoscenze o le capacità necessarie per svolgere bene la professione a cui l'esame si riferisce (AERA, APA, NCME, 1985, Standard 10.5, 10.6, 11.1).

Nelle applicazioni di **psicologia del lavoro** la validità del contenuto è fondamentale quando si vuol valutare se il candidato a un lavoro possiede conoscenze o competenze essenziali per svolgere bene un dato lavoro; con metodi razionali (non necessariamente quantitativi) va evidenziata l'analogia tra il comportamento stimolato dal test e quello richiesto dal lavoro (Lawshe, 1985).

La validità del contenuto, per quel che riguarda la selezione, classificazione e promozione in ambiente di lavoro, si basa sull'"**analisi del lavoro**" (*job analysis*). Il lavoro nel quale si vuol predire la riuscita dovrebbe esser descritto in riferimento a caratteristiche del candidato che possano esser tenute presenti fra i contenuti dei test e che non cambino sostanzialmente nell'arco di un periodo di tempo precisato (AERA, APA, NCME, 1985, Standard 10.4).

Le prove di validità di contenuto di un test usato in applicazioni di psicologia del lavoro devono essere verificate dall'operatore per accertare che:

- c'è uno stretto legame fra contenuti del test e contenuti del lavoro (AERA, APA, NCME, 1985, Standard 10.5);
- c'è una spiegazione razionale che illustra in modo convincente che le conoscenze, capacità e abilità assunte come definizione dell'ambito professionale sono le maggiori determinanti della buona riuscita in quell'ambito (AERA, APA, NCME, 1985, Standard 10.6);
- se un test è stato validato in riferimento a una professione X, ma si dice che può essere utilizzato in riferimento anche alle professioni Y e Z, per ognuna delle professioni aggiuntive devono essere dati elementi concreti che dicano la somiglianza tra i contenuti di questi lavori e di quello per cui il test è stato validato (AERA, APA, NCME, 1985, Standard 10.7).

Nella **psicodiagnostica** la conoscenza dei contenuti è importante per non ottenere misure distorte. Ad esempio, l'uso del MMPI per la valutazione di caratteristiche della personalità normale è stato criticato proprio perché il metodo di costruzione del test ha comportato la sovracampionatura dei sintomi patologici (cfr. Costa et alii, 1985). Le distorsioni possono riguardare anche singoli costrutti: oltre agli esempi di analisi del contenuto di test di abilità, di profitto e di personalità (tematici) esposti nel cap. IV, si vedano nel cap. XIV le differenze tra scale della stessa denominazione ("Ansia") ricavate dal MMPI includendo o escludendo certe componenti dell'ansia e nel cap. XVIII gli studi che documentano il ruolo determinante delle caratteristiche degli stimoli del TAT nell'induzione di storie tristi o bizzarre, che facilmente spingono lo psicologo ad attribuire al soggetto patologie che non ha (cfr. Sharkey e Ritzler, 1985).

(2) Caratteristiche del contenuto eque per le minoranze.

Nell'interpretare i risultati di test somministrati a **portatori di handicap** è bene sapere se i contenuti sono stati modificati adeguatamente e se, sul testo modificato, è stata compiuta una nuova validazione.

Nella valutazione delle capacità cognitive è ormai noto che vada controllata, all'interno dei test "di livello", qual è la proporzione di stimoli che potrebbero porre in situazione di svantaggio gli **appartenenti a minoranze etniche o culturali o gli appartenenti all'uno o all'altro sesso** (si veda, più oltre, il paragrafo sull'interpretazione in funzione di caratteristiche personali dei soggetti).

(3) Caratteristiche dei contenuti nelle "forme parallele" o "equivalenti" di un test.

L'esame dei contenuti ha anche un ruolo importante nella valutazione critica dei risultati ottenuti dagli stessi soggetti rispondendo a due forme parallele del test oppure, a distanza di tempo, a forme equivalenti per fasce d'età diverse.

Le **forme parallele** di un test dovrebbero essere costruite "a priori" in modo da includere gli stessi contenuti.

Questo parallelismo "di contenuto" è molto importante in alcuni casi: p.es. se si tratta di test di profitto utilizzati all'interno di una stessa sessione d'esame distribuendoli a file alterne, se le forme parallele devono essere usate per confronti tra "prima" e "dopo" un trattamento, ecc. Se il manuale non dice nulla riguardo al parallelismo di contenuto, oppure se si tratta di forme parallele "per casualizzazione" (cioè estratte a caso da una banca di items, senza verifiche empiriche sul parallelismo: cfr. cap. VI), conviene operare un controllo empirico del parallelismo di fatto, in modo da interpretare poi correttamente i risultati.

Per quel che riguarda le differenze tra **forme equivalenti per fasce d'età diverse**, in linea di massima non va dato per scontato che i problemi sottoposti a soggetti di età diverse stimolino gli stessi processi; il Manuale dovrebbe dare elementi di fatto per concludere che le differenze di contenuto fra stimoli per livelli diversi hanno un serio fondamento evolutivo e che non introducono distorsioni tali da cambiare il significato dei punteggi. Si rammenti che le grandi polemiche sulle distorsioni a favore della borghesia operate dalla scala Binet (si veda l'analisi di alcuni subtest della Stanford-Binet presentata nel cap. IV) sono connesse con una prevalenza di contenuti verbali nelle prove per ragazzi di 12 anni, prevalenza motivata da Binet con un dato evolutivo oggettivo (a quell'età le prestazioni verbali-culturali sono il mezzo migliore per evidenziare differenze evolutive tra un individuo e l'altro), ma ignorando l'altro dato, pure oggettivo, riguardante la distorsione culturale operata dai quesiti di competenza lessicale.

(4) Contenuti attinenti all'esame, ma estranei al test.

L'esame della validità dei contenuti di un esame psicologico dovrebbe estendersi all'insieme di tutte le informazioni utili per il trattamento di un caso, incluse le *informazioni estranee al test*. Questo problema è particolarmente importante quando i risultati del test vengono "autogestiti" dal soggetto o affidati ad altri operatori (p.es.: medici, insegnanti) che più dello psicologo possono esser tentati di "decontestualizzare" il test e di attribuirgli superpoteri. Per ora, gli strumenti di cui è più diffusa l'autogestione in Italia sono le guide per l'orientamento che includono test d'interessi autosomministrati e autointerpretati (cfr. cap. XVI). Se è lo psicologo a distribuire questi strumenti, dovrebbe almeno raccomandare agli utenti di utilizzare insieme ai risultati del test tutte le altre informazioni a loro disposizione (p.es.: dati sulla riuscita scolastica, sul modo di utilizzare il tempo libero, ecc.). Queste avvertenze, secondo la normativa professionale statunitense, dovrebbero essere stampate sul pacchetto dei test autogestiti (AERA, APA, NCME, 1985, Standard 9.6).

(5) Utilizzare i dati sui contenuti per interpretare coefficienti di correlazione.

Le caratteristiche di contenuto vanno tenute presenti anche per interpretare correttamente i dati sulla validità espressi in forma di correlazione. E' stato osservato che alcuni indici di correlazione sono "gonfiati" perché i contenuti dei questionari usati per misurare costrutti diversi sono talvolta molto simili tra loro. Ad esempio, se per misurare la Mascolinità si ricorre prevalentemente ad items che inducono descrizioni di autostima, l'elevata correlazione tra scala di Mascolinità e misure di stima di sé è ovvia ("tautologica") e va interpretata solo come sovrapposizione fra contenuti, non come correlazione fra costrutti indipendenti ("Mascolinità" e "Stima di sé", nell'esempio).

d) Unitarietà o specificità delle variabili misurate

Parlando della scelta dei test, si è detto che in alcuni casi è conveniente articolare l'esame in più fasi, misurando prima variabili di tipo generale e poi variabili specifiche.

Il livello di generalità delle misure dev'essere tenuto ben presente al momento dell'interpretazione, così da non mettere sullo stesso piano, come fossero indipendenti e paritetiche, la misura di una caratteristica generale (fattore "di ordine superiore") e quella di una sua componente (fattore "di primo ordine").

Si è visto, ad esempio, che l'"**intelligenza**", in base all'insieme delle ricerche compiute durante questo secolo (cfr. cap. X) non si può ragionevolmente considerare una variabile unitaria. In fase interpretativa, conseguentemente, andrà evidenziato per ciascuna delle misure cognitive usate qual è il suo grado di specificità; va detto esplicitamente se mancano misure di livello cognitivo generale e se mancano misure di abilità specifiche rilevanti alla finalità dell'esame (si veda anche quel che è stato detto nella prima sezione di questo Manuale riguardo alla scelta dei test). Se sono state usate sia misure di abilità sia misure di profitto, vanno commentati gli eventuali dislivelli di rendimento, in modo da consentire l'identificazione dei soggetti il cui rendimento scolastico

(dipendente sistematicamente dalla stimolazione istituzionalmente offerta dalla scuola) è al di sotto o al di sopra del rendimento nelle prove di abilità, che invece dipende in gran parte dalla reazione alla stimolazione ambientale generale (cfr. Cronbach, 1984, p. 254 segg.).

Anche le variabili esaminate da test "di personalità" possono essere più o meno specifiche, com'è stato evidenziato sia da studi teorici sia da studi empirici.

Raramente i Manuali dei test riportano dati di tipo tassonomico che consentano di individuare il livello di specificità di una variabile, ma spesso riferiscono i risultati di una o più **analisi fattoriali** (da qui in poi: a.f.) che consentono di ottenere informazioni in proposito.

Altrove sono state illustrate le caratteristiche tecniche dell'a.f. (cfr. cap. VII), da cui si desumono alcune raccomandazioni fondamentali per l'interpretazione di questo tipo di dati.

- I risultati dell'a.f. dipendono pesantemente dal **numero e dalla qualità delle variabili** che il ricercatore ha deciso di inserire nella matrice di analisi: p.es., se oltre ai punteggi dei subtest viene inserito anche il loro totale, si trova un buon fattore generale, anche se i subtest sono pochissimo correlati fra loro. Comunque, i fattori sono interpretabili solo come rapporti all'interno della struttura costituita dalle variabili analizzate e non dicono nulla sui rapporti fra queste ed altre variabili ad esse esterne. Ad esempio, se l'insieme dei test sottoposti ad a.f. include solo un test attinente alla "estroversione", è probabile che questo test abbia saturazioni basse in tutti i fattori, in quanto composto di items estranei all'insieme per contenuto e costruito; ciò non significa peraltro che quel test misuri male l'estroversione, né che i soggetti inclusi in quel campione siano poco estroversi, né che i test in genere siano inadatti a misurare l'estroversione.

- I risultati dell'a.f., come tutti i dati empirici, dipendono molto **dalle caratteristiche dei soggetti che compongono il campione** (cfr. esempi nel cap. X): p.es., se un campione è molto eterogeneo si ottengono coefficienti di correlazione elevati e conseguentemente un buon fattore generale, se il campione è piuttosto omogeneo (p.es. studenti al termine della secondaria superiore o all'interno di una sola istituzione; pazienti selezionati da uno screening preliminare) i coefficienti di correlazione sono bassi e producono una pluralità di piccoli fattori.

- Le decisioni sul **numero di fattori "significativi"** e sulle **rotazioni** ottimali vengono prese arbitrariamente, in base per lo più alla congruenza fra modello teorico ipotizzato e quadro fattoriale risultante;

- **L'interpretazione psicologica del significato di ciascun fattore** è arbitraria e in genere si basa sulla lettura dei contenuti degli items più saturi in quel fattore: vanno quindi lette criticamente le ragioni addotte dal ricercatore per motivare l'identificazione fra certi fattori e certi costrutti psicologici.

Dati i margini di soggettività appena segnalati, prima di accettare una data interpretazione "fattoriale" di un test, bisognerebbe **avere risultati convergenti di almeno tre ricerche compiute su campioni diversi**.

I risultati sulle analisi fattoriali sono utili per comprendere quali e quante sono le variabili misurate dal test e qual è l'ampiezza di ciascuna variabile. Tuttavia questi dati, da soli, non possono essere interpretati come prova della "validità di costrutto" del test, come invece alcuni Manuali indurrebbero a credere.

e) Utilizzare i dati sulla validità del costrutto

I risultati delle analisi fattoriali vengono di solito inquadrati nell'ambito della validità del costrutto: si vedano nel cap. VII i dati sul CPI, nel cap. X i dati sul contributo dell'analisi fattoriale all'elaborazione del costrutto "intelligenza", nel paragrafo sui fattori biologici e culturali la storia della validazione del costrutto G_f come indicatore del deterioramento mentale. La validità del costrutto dovrebbe però essere provata in base anche ad altre informazioni, esterne al test.

Molte volte i dati sulla validità del costrutto sono voluminosi e vanno quindi reperiti al di fuori del Manuale, soprattutto per i test più usati (p.es. Wechsler, Rorschach, MMPI, 16 PF), che sono in circolazione da decenni. In questo volume vengono dati ampi riferimenti bibliografici, con particolare attenzione alla letteratura specialistica italiana, proprio per sopperire alle carenze dei Manuali dei test; è auspicabile che si diffondano sintesi rigorose (p. es. "meta-analisi") dei dati sulla validità, che consentano agli operatori di mantenersi aggiornati senza affrontare compiti di spoglio bibliografico ai limiti del possibile.

Le informazioni di base da ricercare nei dati sulla validazione del costrutto rispondono al quesito: ci sono prove sperimentali che il test misuri proprio la variabile che interessa misurare (cfr. AERA, APA, NCME, 1985, Standard

7.4), nella popolazione su cui si opera? in quale misura incidono sul risultato altri fattori estranei alla variabile che interessa?

In particolare, i dati sulla validità del costrutto possono dare informazioni su:

- **significato della denominazione della variabile**, in aggiunta a quanto evidenziato dall'analisi dei contenuti: p. es. la distinzione fra misure di "attitudine", "abilità" e "profitto", che possono avere una notevole sovrapposizione fra loro (cfr.: AERA, APA, NCME, 1985, Standard 8.3; Ziskin e Faust, 1988, p. 540), viene evidenziata bene dai dati sperimentali che contrappongono le diverse condizioni in cui sono ottenuti certi risultati (p.es.: differenze tra risultati ottenuti prima e dopo un ciclo d'istruzione, differenze di predittività, differenze tra risultati ottenuti da soggetti di livelli scolastici diversi, ecc.);

- **fondatezza dell'interpretazione che s'intende dare** (p.es., in campo clinico: se ci sono elementi di fatto a sostegno di una particolare interpretazione o previsione: AERA, APA, NCME, 1985, Standard 7.1);

- nel caso di test usati come parte sostanziale di esami di abilitazione e di licenza o in altre applicazioni al lavoro, prove del fatto che il costrutto misurato dal test ha un'attinenza reale con l'attività professionale in cui si vuol predire la riuscita (cfr. AERA, APA, NCME, 1985, Standard 11.2; Lawshe, 1985);

- nelle **applicazioni alla psicologia del lavoro e alla programmazione sociale** bisogna anche evitare di considerare i risultati dei soggetti come unicamente dovuti all'intervento degli operatori. Per esempio, un test può avere comprovata validità come misura del profitto ed evidenziare che gli alunni della classe X hanno conseguito, in media, profitto più elevato degli alunni della classe Y; non è giustificato però inferire che l'insegnante della classe X ha lavorato meglio di quello della classe Y, perché mancano informazioni importanti: qual era il livello iniziale del profitto degli alunni, qual era il loro livello di abilità, quali sono state le condizioni di lavoro, ecc.

I dati sulla validità del costrutto possono variare a seconda delle popolazioni su cui le ricerche sono state condotte. Vanno quindi compiuti alcuni controlli, soprattutto se si lavora con soggetti dalle caratteristiche "non standard".

Ad esempio:

- per i test usati nell'orientamento, se si lavora con adulti, portatori di handicap, o altri soggetti "atipici", non si può presumere di applicare i dati sulla validità ottenuti somministrando il test a studenti giovani inseriti in corsi di studio regolari;

- **se si usano test modificati nei contenuti o nelle modalità di somministrazione** per adattarli alle esigenze di portatori di handicap, bisogna cercare prove esplicite del fatto che i test misurino ancora lo stesso costrutto (cfr. AERA, APA, NCME, 1985, pp. 78-79);

- se si modificano i **tempi di somministrazione** dei test non è giustificato presumere che i costrutti misurati rimangano gli stessi (cfr. paragrafo. sull'influsso della rapidità nel cap. XIX; v. anche Nunnally, 1978, pp. 628-641). La norma generale è che i dati sulla validità del costrutto vanno interpretati con logica, come i risultati di qualsiasi ricerca, evitando le generalizzazioni infondate di qualsiasi tipo.

4. Interpretare in funzione delle caratteristiche tecniche del test: comprendere cosa significano i livelli di misura adottati

Le modalità adottate per la quantificazione delle risposte pongono problemi interpretativi su cui l'operatore può intervenire più attivamente, rispetto a quelli riguardanti la validità e l'attendibilità. Soprattutto nell'attuale situazione italiana, data la carenza generalizzata di norme statistiche adeguate, l'operatore è molto libero di intervenire modificando almeno in parte il processo di quantificazione.

a) Modalità della quantificazione: risposte "aperte" e "chiuse"

Qualsiasi test, attualmente, si conclude con un qualche tipo di quantificazione, sia pure a livello elementare (p.es.: scala nominale). E' importante vedere nei dettagli qual è il processo seguito per trasformare risposte qualitative in dati quantitativi.

Ai fini dell'interpretazione, va fatta una prima, ovvia distinzione tra test a risposta "aperta" e "chiusa": ciascuno dei due tipi comporta distorsioni nelle risposte, ma di tipo diverso. Le **risposte aperte** comportano molti svantaggi pratici: esigono lungo addestramento per la siglatura e, comunque, molto tempo per l'assegnazione del punteggio; per di più l'accordo tra siglatori è spesso insoddisfacente, il che compromette la validità del punteggio. Si

preferisce quindi utilizzare questo tipo di quesito solo quando fornisce informazioni che altrimenti andrebbero perdute: p.es. quando servono informazioni su aspetti della capacità produttiva grafica, linguistica, ecc. Negli altri casi, si valutano i pro e i contro che le diverse distorsioni indotte dai due tipi di quesito comportano. Ad esempio, nei test di personalità le risposte ad alternative aperte si prestano di più a distorsioni indotte dalla desiderabilità sociale e dallo stile di risposta personale (p.es. la tendenza a rispondere "Sì" a qualsiasi domanda). Le risposte ad **alternative chiuse** possono però forzare i soggetti a compiere scelte casuali o quasi, e comunque lasciarsi sfuggire informazioni utili a causa di un'inadeguata rappresentatività dei contenuti delle alternative.

Qui appresso vengono date alcune raccomandazioni utili nell'interpretazione di ambedue i tipi di item (v. anche quanto detto in proposito nei capitoli II e IV).

Se il test è **a risposta aperta** stimola una maggior varietà di risposte. Quando, p.es., si somministrano test di profitto finalizzati alla diagnosi o al ricupero, le informazioni date dalle risposte aperte possono essere sfruttate per costruire una tipologia d'errori più chiaramente interpretabile e più utile agli operatori.

Quando si lavora con test a risposta aperta bisogna prendere alcune precauzioni prima di procedere nell'interpretazione:

- individuare quali sono le modalità specifiche seguite per la "codifica" delle risposte (per la maggior parte dei test a risposta aperta esistono più varianti del sistema di codifica: cfr. in particolare il cap. XVIII) e cercare dati empirici sull'"oggettività" (cfr. cap. VI) del metodo di codifica specificamente adottato, tenendo presente che i dati sull'oggettività sono di solito calcolati sulle prestazioni di persone che hanno lunga esperienza di un metodo, non di principianti. I dati sull'accordo tra siglatori vanno combinati con i dati sulla validità predittiva per avere una stima dell'errore standard delle misure riguardanti singoli soggetti (si veda un esempio riguardante il Rorschach nel cap. XVIII).

- Se chi interpreta il test deve anche codificare le risposte, è preferibile che scelga e apprenda il metodo che dà maggiori garanzie di oggettività e che faccia siglare il test, indipendentemente, anche da un'altra persona. I casi di siglatura discordante vanno discussi e unificati, formulando quando è possibile norme aggiuntive che evitino il ripetersi della discordanza.

I test basati su **alternative di risposta chiuse** danno informazioni più scarse ma, se il test è valido, più precise. In particolare, è molto buona l'oggettività (accordo tra correttori), a condizione che vengano compiuti alcuni controlli (necessari quando si danno risultati individuali):

- (1) controllo "per correttore": tenendo distinti i plichi corretti da ciascun correttore, si ricorregge un protocollo ogni cinque-dieci; se si trova più del 10% di errori o se si trovano anche pochi errori ma rilevanti (p.es.: uguali a una deviazione standard), si ricorreggono tutti i protocolli di quel correttore;

- (2) controllo "per soggetto": si esaminano tutti i protocolli con punteggio molto basso (p.es.: 10% inferiore) e si controlla che non ci siano stati errori sistematici nell'uso del foglio di risposta: p.es. risposte segnate in modo da non essere visibili attraverso la griglia o da non essere leggibili mediante il lettore ottico, tutta una sequenza di risposte sistematicamente scritta sulle caselle sbagliate perché il soggetto ha "saltato" un item senza accorgersene, ecc.; si esaminano tutti i soggetti con punteggio molto alto e si controlla se hanno eseguito tutti gli esempi (in caso negativo, potrebbero aver cominciato ad affrontare gli items della prova prima degli altri), se stavano seduti vicini fra loro (potrebbero aver collaborato), se dalle anamnesi risultano già sottoposti a quel test. Se ci sono elementi per sospettare scarsa accuratezza nella somministrazione (p.es.: esempi non completati, copie) il controllo di queste caratteristiche va esteso a tutti i soggetti e i fatti accertati ("esempi non completati", "risposte identiche al vicino nell'X% degli items") vanno scritti all'inizio del "rapporto psicologico" in cui sono commentati i risultati di quei soggetti.

Nell'interpretare i risultati bisogna poi sapere se sono state apportate correzioni al punteggio per tener conto di risposte "indovinate a caso" (cfr. cap. II): è noto che questa è spesso un'ipercorrezione, cioè che abbassa il punteggio più di quel che è ragionevole.

b) Caratteristiche della distribuzione standardizzata: le "norme"

Nei Manuali dei test sono di solito riportate "norme statistiche" che facilitano la standardizzazione dei punteggi.

Queste norme non vanno mai utilizzate senza una riflessione previa. Anzitutto, bisogna ricordare che l'accettazione del riferimento al livello medio comporta implicazioni di valore, di tipo socio-politico (cfr. cap. III): se non si accettano i presupposti e, soprattutto, le conseguenze interpretative e pratiche del riferimento al livello medio, non ha senso utilizzare le norme statistiche offerte nei manuali.

In secondo luogo, quando si decide per l'accettazione del riferimento al livello medio, bisogna controllare che il gruppo di riferimento assunto come "norma" appartenga alla stessa popolazione di cui fanno parte i soggetti che si vogliono valutare. In concreto, bisogna verificare se il campione normativo descritto dal manuale è recente (p.es.: esaminato non oltre dieci anni fa), se è composto di soggetti italiani (e a seconda della variabile misurata e delle caratteristiche del test andrebbe controllata anche l'omogeneità con il proprio campione rispetto alla zona - Nord, Sud, Centro, Isole - e al tipo di centro abitato - città o paese - in cui i soggetti vivono), simili per età (AERA, APA, NCME, 1985, 9.8), sesso (AERA, APA, NCME, 1985, 9.3) e scolarità a quelli da noi esaminati.

Quando si opera con **portatori di handicap** bisogna decidere di volta in volta a quale popolazione riferirsi: in alcuni casi è corretto confrontare il risultato del soggetto con quelli della popolazione "normale" (p.es. se si vuol misurare l'entità del disturbo), in altri è più corretto riferirsi alle prestazioni di soggetti portatori dello stesso handicap (AERA, APA, NCME, 1985, 14.8).

Le caratteristiche del campione normativo sono di fondamentale importanza quando i punteggi standard sono ancorati al livello medio (come avviene in quasi tutti i test pubblicati): questo vale sia per i test cognitivi sia per i "non cognitivi", data la provata relatività culturale di molte variabili psicologiche (cfr. paragrafo successivo e cap. III). Questa esigenza attualmente è così sentita che si tende a considerare preferibile il riferimento a norme statistiche locali recenti (cfr. cap. III), come quelle facilmente calcolabili sui dati raccolti da alcuni operatori locali nel giro di uno-due anni.

L'uso di norme inappropriate può comportare sostanziali variazioni nell'interpretazione. Ad esempio, in molti test d'intelligenza si nota un incremento nel rendimento medio della popolazione da un'edizione all'altra delle norme (cfr. Flieller, Saintigny, Schaeffer, 1986; Raven, Court, Raven, 1986b; Ziskin e Faust, 1988, p. 535): chi viene valutato sulle norme "vecchie" in molti casi può essere considerato più intelligente di quel che è. Distorsioni ancor più gravi derivano dall'uso di norme inappropriate perché calcolate su altre popolazioni (si veda, più oltre, il paragrafo sulle differenze culturali).

Nei rari casi in cui i punti standard sono ancorati a livelli massimi (p.es.: mastery learning) o minimi (p.es.: test per la diagnosi neurologica) bisogna cercare nel Manuale o nella letteratura specializzata indicazioni che provino la validità di fissare a un dato livello anziché ad un altro la soglia che delimita la "padronanza" dalla "non padronanza", la "patologia" dalla "normalità" (cfr. cap. VI; AERA, APA, NCME, 1985, 6.9) o che giustifica la non assunzione di un soggetto in un posto di lavoro (AERA, APA, NCME, 1985, 10.9).

c) Interpretazione dei punti standard tenendo conto dell'errore standard e della distribuzione dei punteggi

Alcune scale standard si articolano in numerosi livelli (p.es.: QI, punti T). In questi casi è bene controllare **l'ampiezza dell'errore standard** della misura (cfr. cap. VI): se l'errore standard ha un valore più grande di 1, vuol dire che una differenza di un punto è trascurabile. Conviene allora usare una scala standard a livelli meno numerosi e più distanziati fra loro. Se questo, per ragioni pratiche, non è possibile, nel commentare i dati bisogna sempre tener presente qual è la distanza minima che si può ritenere significativa.

Va infine controllato se la **distribuzione** dei propri dati è simile a quella dei dati normativi (cioè dei dati su cui è calcolata la taratura standard): questo è relativamente facile quando si utilizzano test collettivi su grandi campioni, mentre richiede una paziente accumulazione di dati per i test individuali. Le norme date dai manuali sono abitualmente espresse in punti "normalizzati", cioè distribuiti secondo la curva normale. Se i soggetti su cui si lavora riportano punteggi con marcate anomalie di distribuzione (le più frequenti sono: asimmetrie e distribuzioni plurimodali), la forzata "normalizzazione" dei loro punteggi può introdurre distorsioni che tradiscono la realtà. Ad esempio, se un test di profitto viene somministrato a studenti del I anno della scuola secondaria superiore in settembre, per valutare lo stato dei "prerequisiti" per l'apprendimento di una certa area disciplinare, è molto probabile che si ottenga una distribuzione con la marcata asimmetria a sinistra propria dei test "difficili": subito dopo le vacanze è noto che i ragazzi non ricordano molto di quel che avevano appreso l'anno precedente. Se questi

punteggi vengono standardizzati in riferimento a norme calcolate su alunni al termine della III media (riferimento razionale per quel che riguarda i contenuti della disciplina, ma non appropriato sotto l'aspetto didattico e psicometrico, dato che al termine di un ciclo di apprendimento di solito la distribuzione è normale e simmetrica), tutta la distribuzione viene distorta: le distanze fra risultati superiori alla media vengono artificialmente amplificate e le distanze tra punteggi inferiori alla media artificialmente ridotte.

d) Punteggi non espressi in scale a intervalli

Nell'interpretare i punteggi va infine tenuta ben presente la caratterizzazione metrica delle scale in cui sono espressi, ricordando le proprietà dei diversi livelli di misura (cfr. cap. III). Qui appresso vengono evidenziate alcune implicazioni importanti per la pratica.

- Quando i punteggi sono **espressi in frequenze o in percentuali** (scala nominale: p.es. quasi tutti i test proiettivi) va tenuto presente che le differenze tra punteggi vanno stimate in base all'errore tipico della misura, che di solito è elevato (cfr. cap. VI).

- Le misure in **centili** vanno interpretate come graduatorie, ricordando che gli intervalli fra un valore e l'altro non sono né uguali né costanti: p.es., il primo e il secondo arrivato in una corsa possono essere ben distanziati o molto vicini, e non è detto che siano altrettanto vicini tra loro quanto lo sono il decimo e l'undicesimo.

Le graduatorie non esprimono le distanze tra punteggi ma solo il rapporto "più" e "meno": fra il centile 49 e il centile 50 non c'è la stessa distanza che passa fra centile 1 e centile 2.

- Quando i punteggi sono **ipsativi**, cioè derivati da scelte forzate (p.es. questionari di Kuder, EPPS, Myers-Briggs, descritti nei cap. XIV e XVI) va tenuto presente che esprimono una graduatoria interna alle preferenze del soggetto, limitata agli oggetti di scelta che sono stati proposti; non possono essere considerati indici del gradimento di un oggetto "in assoluto" né possono essere interpretati in riferimento al gradimento che su quell'oggetto esprimono altri soggetti.

Non è corretto neanche interpretare come misure della stessa intensità d'interesse i punteggi identici assegnati a due soggetti diversi.

- I punteggi dei **test "riferiti a criteri"** quantificano la padronanza in un'area di materia in riferimento a come questa è stata operativamente analizzata e definita da quel particolare test. Hanno uno stretto riferimento ai contenuti del test e non sono interpretabili al di fuori di questi.

- Nei **test a risposta aperta**, la **frequenza** con cui appaiono certi contenuti (p.es. nel TAT: cfr. cap. XVIII) è ritenuta da alcuni indicativa dell'**intensità** con cui sono presenti nel soggetto motivazioni o caratteristiche strutturali, il che è messo in dubbio da altri autori. Vanno quindi cercate nel Manuale prove esplicite a sostegno della validità del tipo di quantificazione proposta e, in assenza di queste, è preferibile interpretare in modo molto informale la quantificazione dell'intensità.

(B) INTERPRETARE TENENDO CONTO DELLE CARATTERISTICHE E DEL COMPORTAMENTO DEI SOGGETTI

L'interpretazione dei risultati dei test può esser compiuta a diversi livelli di organizzazione: si può interpretare il risultato di un singolo test (p.es.: un $QI = 100$ può essere interpretato come "livello d'intelligenza normale"), si può interpretare il "pro-fi-lo" risultante dalla somministrazione di una "batteria" più o meno ampia di test, cercando di strutturare tutte le informazioni in un insieme organico, e infine si possono interpretare i risultati di uno o più test mettendoli in rapporto con caratteristiche dei soggetti diverse dalle prestazioni al test, quali ad esempio il sesso, l'età, la scolarità, la professione del soggetto o dei suoi genitori, i dati anamnestici (si vedano al riguardo, fra le molte trattazioni su problemi psicodiagnostici generali e particolari: Cohn, 1988; Green, 1988; Groppo e Liverta Sempio, 1988; Jansky, 1988; Korchin, 1977; Kron e Kestenbaum, 1988; Sundberg, 1977; Wedell e Raybound, 1976). Quando i risultati di uno o più test vengono interpretati con esclusivo riferimento alle prestazioni fornite in risposta agli stimoli, si parla di "interpretazione cieca". L'interpretazione cieca è considerata da tutti un requisito desiderabile nelle ricerche sui test, dato che la conoscenza di alcune caratteristiche personali può produrre risultati

distorti (si rammentino l'"effetto di alone" e le altre distorsioni tipiche della conoscenza interpersonale "ingenua": cfr. cap. I). Nelle applicazioni pratiche invece la maggior parte degli Autori consiglia di tener conto di tutti i dati disponibili, per arrivare ad un "rapporto psicologico" il più possibile aderente alla realtà. Per i test più complessi, e in particolar modo per i test proiettivi, le cui modeste proprietà psicometriche possono indurre in errore su variabili importanti, la maggior parte degli AA. consiglia di articolare l'interpretazione in due fasi: a) interpretazione basata esclusivamente su dati desunti dalle risposte al test, organizzati in modo da ottenere una descrizione psicologica del soggetto il più possibile completa e senza contraddizioni interne (Exner, 1986; Rossel, Husain, Merceron, 1985/86); b) reinterpretazione della descrizione psicologica in funzione di altre caratteristiche del soggetto, quali età, sesso, cultura, anamnesi, ecc. (cfr. Koppitz, 1982). La necessità della seconda fase è evidente soprattutto quando si esaminano soggetti che non siano giovani adulti normali, di cultura medio-superiore, cioè caratterizzati in modo diverso da quelli a cui in genere si riferiscono gli studi di validazione. Ad esempio, la quantità delle risposte al Rorschach notoriamente aumenta in funzione dell'età e dello sviluppo intellettuale, inducendo come conseguenza un aumento nella varietà delle categorie di risposta e una frequenza diversa da zero per alcune categorie relativamente "rare" (movimento, originali...); se un protocollo - soprattutto di un bambino o di un anziano - viene interpretato senza conoscere l'età del soggetto, dati puramente evolutivi possono esser considerati, erroneamente, come indicatori di psicopatologie.

5. Interpretare in funzione dell'età dei soggetti

La maggior parte delle caratteristiche umane rilevanti ai fini della valutazione psicologica si evolve nel tempo: non solo le abilità cognitive progrediscono e si differenziano qualitativamente dall'infanzia alla maturità per poi declinare nell'età senile, ma anche le caratteristiche strutturali e dinamiche della personalità cambiano con il tempo: il quadro fattoriale ottenuto esaminando bambini, adolescenti e adulti è differenziato in alcune sue componenti (cfr. Cattell, cap. XV), alcuni segni psicopatologici si accentuano nell'età avanzata (p.es.: depressione), e così via.

Per lo più, l'evoluzione delle caratteristiche umane attraverso il tempo viene controllata scegliendo test e norme appropriati all'età del soggetto esaminato. Conviene tuttavia aver presenti alcune tendenze evolutive riguardo almeno alle strutture e alle dinamiche più comunemente esaminate.

a) Leggere criticamente gli studi sull'evoluzione delle caratteristiche personali

Nell'esaminare qualsiasi studio riguardo all'evoluzione di caratteristiche personali con l'età, va fatta molta attenzione alla metodologia con cui lo studio è stato condotto. La maggior parte delle ricerche sono **trasversali**, cioè confrontano le prestazioni di soggetti di diversa età esaminati simultaneamente. Negli studi trasversali è pressoché impossibile distinguere l'influsso del fattore età dall'influsso dei fattori ambientali. Soggetti dai 20 ai 70 anni esaminati, poniamo, nel 1990, vengono da esperienze personali e da formazioni scolastiche molto diverse: i ventenni sono stati istruiti da una scuola dell'obbligo unitaria, largamente e capillarmente diffusa, per lo più frequentata fino a 14 anni, seguita per moltissimi da un'istruzione secondaria e universitaria pochissimo selettive; chi ha cinquant'anni o più è stato istruito da una scuola dell'obbligo estesa solo fino agli 11 anni, meno capillarmente diffusa, gestita in situazioni di grande difficoltà durante la guerra e nel dopoguerra, seguita da una scuola secondaria e da un'istruzione universitaria selettive. Essendo notorio l'influsso sistematico della qualità dell'istruzione sullo sviluppo delle abilità, sarebbe sorprendente se persone sottoposte ai due diversi trattamenti ne uscissero con gli stessi risultati, anche a parità di condizioni genetiche e ambientali generali. Non si possono peraltro ipotizzare stabili neanche le condizioni ambientali generali: gli anziani del 1990 hanno vissuto la guerra e il dopoguerra da adulti, i cinquanta-sessantenni da bambini; i quarantenni hanno avuto la loro infanzia nel "boom" degli anni cinquanta, e così via; la famiglia, negli ultimi cinquant'anni, è passata da una struttura semi-patriarcale, di "famiglia allargata", a una struttura nucleare, in cui i nonni, gli zii e in parte le mamme sono stati sostituiti da asili nido e "baby-sitter". La diversità nelle esperienze precoci e nella strutturazione dei rapporti affettivo-sociali notoriamente induce differenze individuali: dobbiamo quindi presumere che anche i campioni che rappresentano fasce d'età diverse siano differenziati nelle caratteristiche affettive e sociali. Gli studi

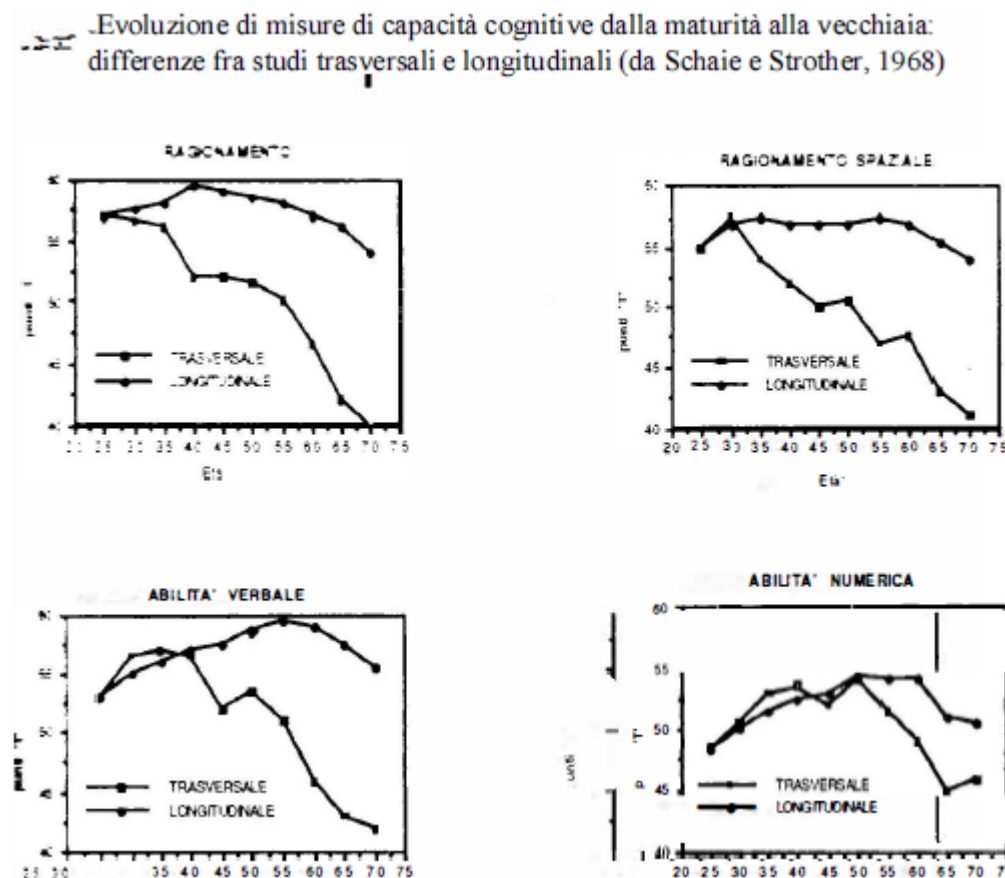
longitudinali danno risultati più chiaramente interpretabili, in quanto confrontano i risultati successivamente ottenuti dagli stessi soggetti. Sono però poco numerosi, dato l'impegno organizzativo che richiedono per fronteggiare il progressivo assottigliamento dei campioni, per elaborare test equivalenti adatti a tutta la gamma di età considerata e per ridurre gli effetti di apprendimento e memoria e le distorsioni che i soggetti più o meno consciamente introducono nel tentativo di evitare ripetizioni. Quando vengono pubblicati, inoltre, sono in gran parte "vecchi", dato che i primi dati raccolti sono sempre più lontani nel tempo.

Nell'interpretare i dati sulla stabilità delle misure personali bisogna anche far attenzione alla "finezza" delle misurazioni usate: se gli Autori impiegano misure sommarie (p.es.: si limitano a distinguere fra soggetti con punteggi superiori o inferiori alla media) la stabilità appare sopravvalutata (cfr. Clarke e Clarke, 1972). E' anche importante considerare la diversa "qualità" delle variabili in rapporto agli strumenti di misura e al criterio. P.es. (McClelland, 1985) distinguere tra misure "operanti", che si riferiscono a pensieri o azioni prodotti spontaneamente o in risposta a stimoli ambigui (p. es. caratteristiche del modo di parlare, momenti espressivi, successo professionale, bisogni manifestati nelle storie del TAT, determinanti delle risposte al Rorschach...) e misure "rispondenti", riferite a pensieri o azioni prodotti in risposta a stimoli chiaramente identificati (p.es. risposte a questionari, tempi di reazione, indipendenza dal campo, ecc.).

Gli studi sulla stabilità delle caratteristiche personali presentano dati di due tipi: informazioni su eventuali cambiamenti nei valori medi da un anno all'altro (evidenziati mediante t di Student, AVA, ecc.) e informazioni sulla stabilità delle differenze individuali, per lo più ottenute calcolando coefficienti di correlazioni fra test e retest. Questi due tipi d'informazione sono qualitativamente diversi e vanno integrati l'uno con l'altro.

FIG. XX.1

Evoluzione di misure di capacità cognitive dalla maturità alla vecchiaia: differenze fra studi trasversali e longitudinali (da Schaie e Strother, 1968)



b) Evoluzione di caratteristiche "cognitive"

Per quel che riguarda i test di **abilità generale** è comunemente accettato che i risultati abbiano un incremento medio graduale dall'infanzia alla maturità, un declino dalla maturità alla vecchiaia e notevole stabilità nel periodo compreso fra la tarda adolescenza e l'inizio dell'invecchiamento. La predittività di misure prese nell'infanzia (inizio scuola elementare) è stata trovata molto buona per i più dotati e per i meno dotati (si vedano in particolare gli studi di Terman, sintetizzati anche in Cronbach, 1984, Anastasi, 1982, Bloom, 1965). I problemi pratici da affrontare nell'interpretazione dei risultati riguardano le ipotesi: (a) sull'età cronologica a cui, per la maggior parte delle persone, termina lo sviluppo; (b) sull'età a cui comincia il declino; (c) sulla progressione dell'evoluzione e del declino, che potrebbe avvenire lungo un continuum o per "stadi"; (d) sulle differenze di stabilità fra abilità diversamente caratterizzate (p.es.: fra intelligenza "fluida" e "cristallizzata").

L'età a cui termina lo sviluppo dell'abilità generale misurata dalle scale Stanford-Binet e Wechsler (cfr. cap. IX) è di solito individuata intorno ai 20 anni (dai 17 ai 25, a seconda degli studi: cfr. Bradway e Thompson, 1962) in base a studi trasversali. Il QI "terminale", tenendo conto di un insieme di studi, ha un coefficiente di correlazione (corretto per l'attenuazione) $r = 0,65$ con il QI misurato a 3 anni, $r = 0,80$ con il QI misurato a 5 anni, $r = 0,90$ con il QI misurato a 8 anni e tra $r = 0,90$ e $r = 1$ con misure prese dopo questa età (Bloom, 1965). In base a studi longitudinali i risultati ai test tradizionali aumentano progressivamente fino ai 55 anni (Schaie e Strother, 1968). I test utilizzabili con bambini molto giovani (fino ai due anni) danno risultati poco predittivi anche perché necessariamente basati su attività elementari, scarsamente comparabili con quelle stimulate dai test di abilità generale (cfr. Anastasi, 1982; Cattell, 1987). L'impiego di misure riferite a singole abilità (p.es.: memoria di riconoscimento) non sembra finora aver dato risultati migliori rispetto all'uso delle "scale per bebè" (cfr. McCall, 1981), che a livello di gruppo consentono di predire con esattezza il 70-80% dei risultati raggiunti nella fanciullezza (cfr. Siegel, 1989). L'applicazione di nuovi approcci metodologici intermedi fra il modello longitudinale e quello trasversale potranno forse facilitare la raccolta di dati pertinenti ed utili per le applicazioni future (cfr. Chown, 1989). In base ai dati attuali comunemente si ritiene che fino all'età di cinque anni sia imprudente far predizioni su quale sarà il livello terminale di sviluppo e, conseguentemente, prendere decisioni che implicino questa previsione.

Inoltre, va tenuto presente che questi dati si riferiscono alla "media" e possono quindi essere inappropriati ai singoli casi, sui quali potrebbero influire ritmi di sviluppo peculiari e influssi ambientali in opposte direzioni (cfr. Clarke e Clarke, 1972). Ad esempio, un gruppo di 25 adolescenti ai limiti inferiori della norma per rendimento intellettuale generale venne esaminato due volte (Clarke, Clarke, Reiman, 1958). La correlazione fra i QI ottenuti a distanza di due anni e tre mesi fu $r = 0,90$ e l'incremento medio dei QI fu di 9,71 punti. La distribuzione dell'incremento però variava notevolmente a seconda dei soggetti: per uno il QI aumentò 25 punti, sei ebbero un aumento di 15 o 16 punti, uno di 13, due di 12 e i rimanenti 15 ebbero un aumento di 11 punti o meno. Di solito, le variazioni individuali dipendono da fattori fisici o ambientali identificabili (cfr. casi commentati in Cronbach, 1984, pp. 240-241). Conseguentemente, previsioni fatte in riferimento ai soli valori medi, ignorando le situazioni personali che possono influire sui cambiamenti, sono necessariamente poco accurate.

Abilità generali diversamente caratterizzate rispetto ai contenuti su cui operano o ai processi posti in atto hanno caratteristiche evolutive diverse, evidenziate anche da una diversa composizione del quadro fattoriale: p.es. (Cattell, 1987) l'abilità numerica ha saturazioni egualmente importanti in gf e gc a 5-6 anni, mentre è prevalentemente satura in gc a 13-14 anni e la correlazione tra gc e gf è abbastanza elevata tra ragazzi in età scolare (intorno a $r = 0,45$), bassa tra adulti ($r = 0,16$). L'**intelligenza cristallizzata** (cfr. cap. X) continua ad aumentare fino a un'età abbastanza elevata (intorno ai 50 anni: cfr. Cattell, 1987 e alcuni studi riferiti in Chown, 1972) e declina molto lentamente dalla maturità alla vecchiaia, mentre l'**intelligenza fluida** termina il suo sviluppo nella tarda adolescenza e comincia a declinare rapidamente fin dall'inizio della maturità (cfr. Eysenck, 1973a, p. 79): ad esempio, il rendimento alle PM 38 - considerate tra le più valide misure del fattore g - comincia a declinare intorno ai 20-25 anni (cfr. J.C. Raven, Court, J. Raven, 1986b e alcuni dati italiani in Spangaro, 1970); per i soggetti di 65 anni il Manuale inglese riporta valori medi uguali a quelli dei ragazzi di 9 anni (J.C. Raven, Court, J. Raven, 1986b). Mostrano un'evoluzione analoga anche abilità più particolari, ma fortemente sature nell'uno o nell'altro di questi

due fattori "generalisti". Ad esempio, gli stili di apprendimento (che dipendono in gran parte dall'abilità di elaborare l'informazione a livelli di concettualizzazione diversi) continuano a svilupparsi durante il periodo universitario, evolvendo in molti casi da un'elaborazione "superficiale" dell'informazione, basata principalmente sulla memorizzazione di contenuti, a un'elaborazione "profonda", in cui i contenuti vengono analizzati, esaminati criticamente e organizzati in una struttura personale prima di essere memorizzati (cfr. Schmeck, 1983). Sembra anche che il declino nelle prestazioni ai test "non culturali" (in specie: alle PM) sia tanto più lento quanto più i soggetti hanno un livello elevato di intelligenza "cristallizzata", il che suggerisce l'influsso mediatore di variabili che influiscono su ambedue i costrutti (p.es. livello socioculturale e salute, che peraltro potrebbero essere connessi fra loro: cfr. Chown, 1972). Il declino nel rendimento intellettuale degli anziani appare anche sensibile a variabili ambientali: ad esempio, è stato osservato un accentuarsi del declino nel periodo immediatamente successivo a eventuali istituzionalizzazioni (cfr. Chown, 1972).

Per quel che riguarda le applicazioni all'orientamento, per la fascia d'età compresa fra i 14 e i 19 anni i risultati dei test di abilità generale, caratterizzati culturalmente (Gc) o no (Gf) mostrano una buona stabilità; rispetto alle valutazioni degli insegnanti, i punteggi dei test sono in alcuni casi meno predittivi (soprattutto i test di Gf), ma più stabili (cfr. Aubret, 1987).

La rapidità, che secondo alcune interpretazioni è una dimensione importante dell'intelligenza (cfr. $\approx$ 8) appare evidentemente soggetta al declino dopo la maturità. Secondo alcuni AA. (cfr. Chown, 1972) gli anziani sono più lenti nella soluzione dei problemi cognitivi perché impiegano molto tempo a prendere decisioni (cfr. Griew, 1959) e perché insistono più a lungo dei giovani nel tentar di risolvere problemi difficili per loro (Chown e Davies, 1969, cit. in Chown, 1972). Nell'interpretare i risultati di test cognitivi somministrati a soggetti anziani può quindi essere utile anche esaminare il rapporto errate/esatte/omesse, considerando un normale segno dell'età i comportamenti che ad età inferiori sarebbero interpretati come eccesso di accuratezza.

Le ricerche sulle caratteristiche evolutive delle **abilità specifiche** sono meno numerose. Uno studio di L.L. Thurstone (Thurstone, 1955; cfr. anche Eysenck, 1973a) dà le seguenti stime sul livello terminale dello sviluppo di alcune abilità specifiche, basandosi su dati raccolti con metodo trasversale: Rapidità percettiva 12 anni, Ragionamento spaziale e Ragionamento astratto 14, Ragionamento numerico e memoria 16, Comprensione verbale 18, Fluidità verbale 20. Dati raccolti con metodo longitudinale (Schaie e Strother, 1968; Cattell, 1987) evidenziano che molte capacità raggiungono il loro massimo livello a 50-60 anni (Fluidità ideativa e associativa, Nozioni meccaniche, Valutazione sperimentale, Rapidità in aritmetica), altre lo raggiungono intorno ai 35-40 (Abilità spaziale, Visualizzazione, Ragionamento, Rapidità di strutturazione), altre intorno ai 21-28 (Flessibilità figurativa adattiva, Fluidità verbale e, in genere, abilità primarie connesse con la rapidità mentale e psicomotoria), altre ancora intorno ai 18-20 (Flessibilità di strutturazione, Orientamento spaziale, Relazioni figurative, Rapidità intellettuale, Induzione). Com'era prevedibile, è stato provato l'influsso specifico del fattore "giovinezza" sulle abilità motorie (cfr. Highmore, 1949, in P.E. Vernon, 1961). Le caratteristiche cognitive patologiche hanno stabilità variabile a seconda della categoria considerata. In una ricerca statunitense su 523 soggetti portatori di handicap lievi, il 24% risulta classificato in almeno due modi diversi durante il suo periodo di scolarità, fra le categorie considerate in questa ricerca, la meno stabile era "impedimenti nel linguaggio" e la più stabile "disturbato emotivamente" (Wolman, Thurlow, Bruininks, 1989).

Riguardo alle abilità specifiche oltre ai problemi sopra accennati se ne pone un altro, importante per alcune applicazioni (p.es. per l'orientamento): l'ipotesi sull'età in cui le singole abilità cominciano a evidenziarsi, differenziandosi rispetto all'abilità generale. Nel corso del dibattito fra sostenitori della teoria del fattore generale e sostenitori della teoria multifattoriale (cfr. cap. X) venne formulata l'ipotesi che g "tendesse a differenziarsi in abilità più specializzate durante l'adolescenza e l'inizio dell'età adulta" (cfr. P.E. Vernon, 1961, pp. 29-30). Fra il 1930 e il 1950 furono condotti numerosi studi per verificare questa ipotesi, senza che si arrivasse a conclusioni univoche. L'ipotesi venne successivamente riformulata per tener conto dell'interferenza di fattori evidenziati dalle ricerche. Fra questi, ha un ruolo molto importante l'organizzazione scolastica, curricolare e didattica, che a seconda dell'impostazione attuata appare connessa con un accentuarsi delle differenziazioni attitudinali (in scuole tecniche o professionali) o, viceversa (in scuole simili al nostro Liceo classico) con un accentuarsi dell'integrazione delle conoscenze tra loro: "non c'è una tendenza generale alla differenziazione, eccetto forse nella prima infanzia, e tutto

dipende dal tipo di formazione scolastica e professionale" (P.E. Vernon, 1961, pp. 30-31).

E' stata anche segnalata l'importanza di alcuni problemi metodologici: ad esempio, i campioni di adulti considerati dalle ricerche sono quasi sempre campioni più omogenei per abilità, il che automaticamente riduce il peso del fattore g; quando i test sono facili viceversa aumenta il peso del fattore generale. Alcuni fattori sono stati più studiati sotto l'aspetto evolutivo. Ad esempio, l'abilità spaziale è già evidente, e distinta dal fattore generale, a 11-13 anni e anche prima, ed ha un forte incremento durante l'adolescenza, specie nei maschi (cfr. P.E. Vernon, 1961, pp. 67-8 e 120; Pellegrino e Goldman, 1983). Studi longitudinali hanno evidenziato che i risultati di test verbali e di memoria somministrati in età prescolastica (da 2;6 a 5;6 anni) predicono i risultati ottenuti in età adulta meglio di quanto non avvenga per i risultati dei test non verbali (cfr. Bradway e Thompson, 1962).

Il fatto che la maggior parte degli autori consideri costante, almeno dalla seconda infanzia alla maturità, il peso relativo di g e dei fattori specifici non significa peraltro che si ipotizzi una stabilità nel tempo dei processi posti in atto dai soggetti per rispondere agli stessi tipi di test (cfr. P.E. Vernon, 1961, p. 140). Ad esempio, i test di abilità numerica basati sull'esecuzione e il controllo di addizioni e sottrazioni sono saturi in fattore numerico a bassi livelli di età (ragazzi della scuola dell'obbligo) e dipendono invece quasi esclusivamente dalla rapidità in campioni di adolescenti e di adulti che ormai padroneggiano completamente quelle tecniche di calcolo. Differenze nei processi cognitivi posti in atto a età diverse sono stati evidenziati in riferimento all'abilità nella lettura (Perfetti, 1983), a problemi di analogie verbali (Pellegrino e Goldman, 1983) e di ragionamento spaziale (Pellegrino e Goldman, 1983).

Lo sviluppo cognitivo, com'è stato documentato da numerosi studi, può utilmente essere interpretato come una successione di **stadi** qualitativamente differenziati. La successione fra gli stadi tuttavia non è, come alcuni vorrebbero, così rigida da consentire l'individuazione di "leggi naturali" ad alto livello di generalizzabilità (cfr. Epstein, 1974a, b, e una raccolta di critiche in McCall, 1988). Per di più, la regolarità della successione degli stadi non è tale da eliminare le differenze individuali di sviluppo, che possono riguardare sia il ritmo con cui gli stadi si susseguono, sia lo schema generale in cui si organizzano; è anche stato provato che tutti i fattori ambientali che influiscono sullo sviluppo cognitivo influiscono anche sulla durata e l'organizzazione interna degli stadi (cfr. una rassegna di studi in Fischer e Silvern, 1985). Nell'interpretazione dei risultati dei test cognitivi espressi in stadi (cfr. cap. IX) si dovrà quindi evitare di riferirsi crudamente agli studi piagetiani, tenendo conto invece sia di tutti i dati empirici disponibili (in specie: risultati ottenuti in quella prova da campioni di soggetti simili ai propri), sia di tutti i fattori di cui si parla in questa sezione del manuale.

Nell'esame degli **anziani** bisogna essere attenti a distinguere il declino "normale" delle prestazioni dai segni psicopatologici. Ad esempio, è stato ripetutamente segnalato che prestazioni normali per l'età possono essere erroneamente interpretate come indicative di patologie neuropsicologiche, soprattutto se gli anziani hanno basso livello culturale (cfr. cap. XIII e Ziskin e Faust, 1988, pp.529-31).

c) Evoluzione di caratteristiche "non cognitive"

Sull'evoluzione delle caratteristiche affettive e delle psicopatologie le ricerche sono più carenti. Tuttavia, al momento attuale sono stati almeno accertati alcuni fatti importanti per l'interpretazione dei risultati dei test.

Nel periodo compreso fra la tarda adolescenza e l'inizio della senescenza le caratteristiche personali, così come sono misurate dai test, appaiono notevolmente stabili, sia nella media (differenze tra medie dei punteggi a distanza di tempo), sia come costanza individuale (correlazioni fra test e retest).

Non tutte le variabili sono però ugualmente **stabili**: le correlazioni nei retest a distanza di 4 anni, per il 16 PF forma A vanno da una mediana $r = 0,30$ per il fattore N a un valore $r = 0,64$ per il fattore H in un campione maschile (cfr. Cattell, Eber, Tatsuoka, 1970); a distanza di tre anni (ultimi tre dell'istruzione secondaria) la stabilità degli interessi scientifici è stata quantificata in $r = 0,12$ e quella degli interessi letterari in $r = 0,22$ (Aubret, 1987). La stabilità varia a seconda della fascia d'età considerata: ad esempio, tratti come l'estroversione e la motivazione al successo appaiono già ben fondati nella prima infanzia (cfr. Emmerich, 1964), mentre tratti come la fiducia in se stessi, il conformarsi all'autorità e l'indipendenza dal gruppo evolvono durante tutta l'adolescenza (cfr. Costa e McCrae, 1986; Kagan e Moss, 1962; Tuddenham, 1959). Il confronto tra i questionari di Cattell adatti a diverse fasce d'età evidenzia nell'insieme una notevole stabilità del quadro fattoriale per i fattori designati con le lettere che vanno

da A a J, cioè per i fattori che spiegano la maggior proporzione di varianza (Cattell e Kline, 1977, 1982). La corrispondenza con il quadro fattoriale dell'età adulta (16 PF), com'è ovvio, è tanto maggiore quanto più i soggetti si avvicinano all'età adulta (cfr. fig. XX.2).

La stabilità delle variabili è diversa per i due sessi. Ad esempio, nelle risposte al 16 PF forma A (cfr. Cattell, Eber, Tatsuoka, 1970, p. 31) appare diversa la stabilità di uomini e donne relativamente ai fattori A ("affettotimia": $r = 0,49$ per gli uomini e $r = 0,62$ per le donne), H (timidezza-audacia: $r = 0,49$ per gli uomini e $r = 0,64$ per le donne), I ("durezza": $r = 0,63$ per gli uomini e $r = 0,53$ per le donne), N ("ingenuità": $r = 0,39$ per gli uomini, $r = 0,21$ per le donne). L'influsso differenziale dei ruoli sessuali sulla stabilità dei tratti è stato peraltro esplicitamente segnalato da vari autori: ad esempio la "dipendenza" appare più stabile fra le donne, in cui viene incoraggiata, che non fra gli uomini, in cui viene scoraggiata (Kagan e Moss 1962; cfr. fig. XX.3).

FIG. XX.2

I fattori di R.B. Cattell nelle varie fasce d'età (da Cattell e Kline, 1982)

Fattore	PSQ	ESPQ	CPQ	HSPQ	16PF
A	+	+	+	+	+
B	+	+	+	+	+
C	+	+	+	+	+
D	+	+	+	+	+
E	?	+	+	+	+
F	+	+	+	+	+
G	?	+	+	+	+
H	?	+	+	+	+
I	+	+	+	+	+
J	+	+	+	+	+
K			?	?	+
L					+
M					+
N	+	+	+		+
O	+	+	+	+	+
P					+
Q ₁					+
Q ₂				+	+
Q ₃			+	+	+
Q ₄	+	+	+	+	+
Q ₅					+
Q ₆					+
Q ₇					+

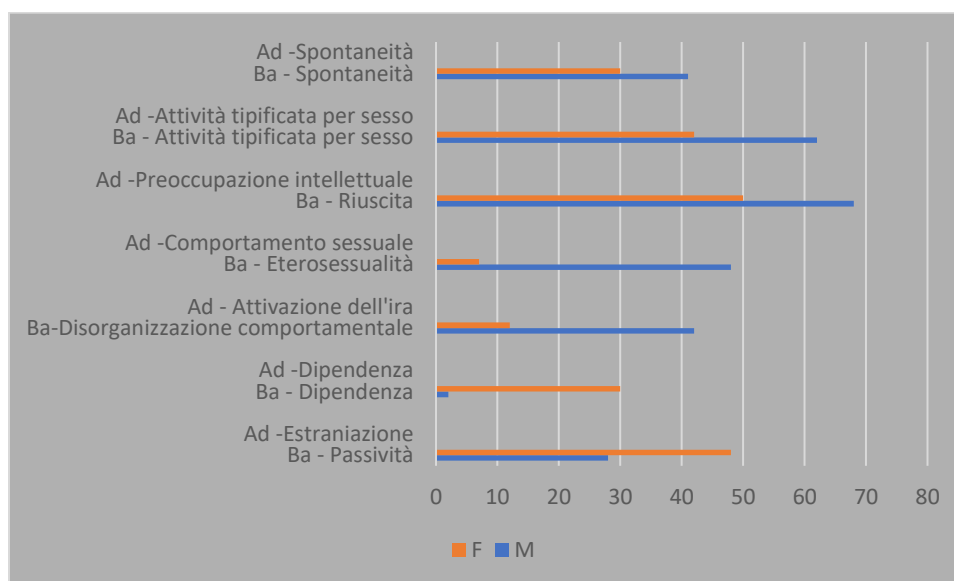
I cambiamenti sistematici in direzione di una maggior conformità agli stereotipi sessuali sociali è evidente nel passaggio dall'adolescenza alla maturità (cfr. Block, 1971) e documentata da studi empirici: ad esempio (Stein, Newcomb, Bentler, 1986) fra i 14 e i 22 anni aumenta più per i maschi che per le femmine l'ambizione e l'estroversione, più per le femmine che per i maschi l'oggettività, l'essere ordinate e l'autoaccettazione. Tutti i dati finora citati sono relativi a popolazioni statunitensi. Vale però anche per la popolazione italiana l'ipotesi che la progressiva differenziazione tra sessi avvenga in funzione degli stereotipi culturali, anche se questi sono in Italia parzialmente diversi da quelli statunitensi e quindi potrebbero comportare l'incremento differenziale di variabili diverse da quelle citate come esempio.

La stabilità delle caratteristiche personali varia attraverso le diverse fasi della vita.

L'importanza delle esperienze avute nella **prima infanzia** è comunemente riconosciuta, ma la maggior parte degli autori considera semplicistico ipotizzare una connessione deterministica fra caratteristiche personali nella prima infanzia e nella vita adulta : "la comune convinzione che le esperienze dei primi anni di vita determinano lo sviluppo di una persona per il resto della sua vita ... è generalmente un mito" (Kohlberg, Ricks, Snarey, 1984; cfr. anche Clarke e Clarke, 1972).

FIG. XX.3

Correlazione fra caratteristiche personali nell'infanzia (Ba = Bambini) e nell'età adulta (Ad), per i due sessi (Blu = Maschi, Arancio = Femmine; da Kagan e Moss, 1962)



La stabilità delle caratteristiche della prima infanzia appare ovviamente migliore se il criterio è la seconda infanzia. Misurando le capacità di bambini piccoli (da pochi mesi di vita a 2 anni di età) con le Bayley Infant Scales (Siegel, 1989) sono state ottenute predizioni di gruppo abbastanza accurate (esatte per il 70 o 80% dei casi) dei livelli cognitivi raggiunti nella fanciullezza (fino a 8 anni).

Tra la fanciullezza e l'adolescenza i tratti mostrano una certa stabilità, ma inferiore a quel che si osserva all'interno dell'età adulta: sono state osservate correlazioni intorno a $r = 0,50$ fra valutazioni date dagli insegnanti agli stessi ragazzi, a 10 e 13 anni (Backteman e Magnusson, 1981). Tra misure prese nel corso dell'adolescenza le correlazioni sono un po' più elevate: p.es., intorno a $r = 0,65$ tra valutazioni di aggressività e accettazione dei pari date agli stessi ragazzi a 13 ed a 16 anni (Olweus, 1977), intorno a $r = 0,76$ per altre variabili (cfr. Block, 1971).

E' stato osservato invece che **gran parte delle caratteristiche dell'età adulta sono già presenti nell'adolescenza**. Fra i 14 e i 22 anni, usando un questionario composto di 15 scale molto brevi, sono stati osservati coefficienti la cui mediana è $r = 0,31$, ma la stabilità delle caratteristiche appare maggiore ($r = 0,56$) se si confrontano le misure prese sugli stessi soggetti a 18 e a 22 anni (Stein, Newcomb, Bentler, 1986). Altri studiosi (cfr. Block, 1971) hanno confrontato misure prese durante la scuola secondaria superiore e a 30 anni, trovando coefficienti intorno a $r = 0,55$. Block segnala anche alcuni "tipi" che manifestano una notevole stabilità nel tempo. P.es., ragazzi che nella scuola media erano attendibili, produttivi, ambiziosi, brillanti, con interessi intellettuali e filosofici, gradevoli, interessati a molte cose, introspettivi, aperti al contatto sociale e soddisfatti di sé, intorno ai 45 anni possedevano ancora, sostanzialmente, l'insieme di queste qualità, anche se era diminuito il loro calore umano (Block, 1985). Considerando l'insieme dei dati, sembra ragionevole concludere che una discreta approssimazione all'età adulta si osserva intorno ai 18 anni. Per alcune variabili i cambiamenti che si producono durante l'adolescenza sono significativi: in particolare, tra l'inizio e la fine dell'adolescenza aumenta il conformismo (cfr. Kagan e Moss, 1962;

Tuddenham, 1959) e la stima di sé (Kagan e Moss, 1962; O'Malley e Bachman, 1983; Petersen et alii, 1984; Tuddenham, 1959).

In ricerche (trasversali) su campioni di **anziani** si è visto che misure di estraversione e di soddisfazione per la vita diminuiscono progressivamente a partire dai 69 anni (cfr. Clément, 1980). Questi dati lasciano qualche dubbio interpretativo se sono esaminati alla luce di altre informazioni: ad esempio, è stato osservato che il cambiamento nelle risposte all'EPI di Eysenck (estraversione e nevroticismo) a distanza di un mese (Gold e Andres, 1985) era significativamente correlato con una misura di benessere, il che fa pensare a un influsso non trascurabile dello "stato" momentaneo, e con la scolarità, il che fa pensare a un influsso della scarsa comprensione dei quesiti.

Sull'evoluzione nel tempo delle caratteristiche personali influiscono sia fattori genetici sia fattori ambientali, di cui non è facile dire il peso relativo. In questo paragrafo è stata esposta una sintesi di studi in cui, per lo più, non sono state compiute distinzioni fra i due tipi d'influsso: nella misura in cui l'evoluzione dipende da fattori culturali, può darsi quindi che i risultati di ricerche compiute all'estero non siano generalizzabili ai nostri soggetti.

Per di più, non tutte le persone sono ugualmente stabili nelle loro caratteristiche personali, e in alcune persone è anche difficile evidenziare la presenza di "tratti comuni", il che produce un profilo "piatto" (cfr. Tellegen, 1988).

In linea di massima, tuttavia, si può tener presente che: a) **la predittività è tanto maggiore quanto più breve è l'intervallo di tempo** a cui ci si riferisce; b) la stabilità delle caratteristiche è massima durante la maturità, minima per le misure prese nella prima infanzia; c) si possono trovare singoli individui con variazioni elevate anche quando le correlazioni fra test e retest sono alte; d) la variabilità individuale è massima nei periodi in cui è forte la spinta al cambiamento derivante da fattori genetici (infanzia, senescenza) o ambientali (cambiamenti importanti nella vita: lutti, stress, trasferimenti, matrimoni, divorzi...). Si noti, riguardo a quest'ultima affermazione, che "attualmente, praticamente tutti gli Autori, incluso Jensen (1969) nel suo famoso articolo, accettano che il cambiamento da condizioni ambientali estremamente avverse a condizioni migliori tende ad essere seguito da cambiamenti desiderabili nelle caratteristiche comportamentali" (Clarke e Clarke, 1972).

Nel commentare i risultati dei test bisogna quindi fare molta attenzione a non interpretare in senso patologico valori che in media evolvono spontaneamente in direzione normale e a prendere in considerazione - per chi ne ha bisogno - soprattutto interventi "di sostegno", preferibilmente collettivi e non etichettati come "terapia", che aiutino lo sviluppo normale.

6. Interpretare tenendo conto del sesso dei soggetti e dell'anamnesi generale

a) Differenze tra i sessi nelle prestazioni ai test

Le proteste del movimento femminista, fra il 1960 e il 1970, attirarono l'attenzione sulle distorsioni che i test potevano introdurre a svantaggio delle donne. Fra i risultati notevoli del movimento si possono citare l'eliminazione della scala "interessi maschili - femminili" dal questionario d'interessi di Strong (descritto nel cap. XVI) e la pubblicazione delle Guidelines for assessment of sex bias and sex fairness in career interest inventories ("Guida per la valutazione delle distorsioni inerenti al sesso e dell'equità per i sessi negli inventari d'interesse per le carriere": cfr. Tittle, 1979). [Aggiornamento 2021. Attualmente gli studi si riferiscono al concetto di "genere" e di "carriera", spesso all'interno di ambiti professionali relativamente omogenei. Un lavoro recente sulla predittività nell'ambito di diverse carriere mediche è: Burns, S. (2014). Gender Differences in the Validity of Career Interest Inventories. *Psychology*, 5, 785-797]

Nell'**area cognitiva** sono state spesso segnalate differenze tra sessi (cfr. capp. X e XI), fin dall'inizio di questo secolo (cfr. Burt e Moore, 1912), soprattutto ri-guardo alle abilità spaziali-meccaniche (in cui eccellono i maschi) ed alle abilità specifiche (p. es. abilità manuali e capacità di ragionamento spaziale per i maschi, immaginazione, capacità verbale e ortografia per le femmine). Gli studi più recenti, condotti su campioni di ragazzi che frequentavano classi "miste", non evidenziano differenze significative tra i due sessi; nei sistemi scolastici in cui sono consentite opzioni tra materie c'è però una preferenza più accentuata dei maschi per la matematica e le materie scientifiche (cfr. Chipman, Brush, Wilson, 1985): si tratterebbe quindi di differenze tra interessi (o valori) più che tra abilità. Anche le differenze a favore dei maschi ripetutamente rilevate nelle abilità "spaziali" (cfr. cap. XI) potrebbero

essere in rapporto con preferenze per occupazioni del tempo libero in cui queste abilità vengono stimolate e potenziate: presumibilmente, giochi come il calcio stimolano le abilità spaziali più dei giochi con le bambole. Le differenze tra sessi andrebbero Burns, S. (2014). *Gender Differences in the Validity of Career Interest Inventories. Psychology, 5, 785-797* comunque studiate tenendo anche conto del diverso ritmo maturativo di Maschi e Femmine in età scolare e dell'interferenza di caratteristiche "non cognitive" a forte base genetica (p.es. la dominanza) che contribuiscono a differenziare i comportamenti dei due sessi anche sotto l'aspetto cognitivo.

Studi recenti, basati sul confronto tra dati normativi di test cognitivi largamente diffusi (DAT, PSAT, SAT: cfr. cap. XI) e su meta-analisi di ricerche, evidenziano che almeno negli USA le differenze cognitive tra sessi stanno scomparendo (Feingold, 1988; Linn e Hyde, 1989) e che, comunque, sono strettamente connesse con influssi culturali, soprattutto dell'istituzione scolastica (cfr. Linn e Hyde, 1989; Marsh, 1989).

Nei test di abilità generale di solito non si osservano differenze statisticamente significative per sesso; in qualche caso sono state osservate differenze a favore delle donne, che però con l'età tendono a scomparire (cfr. Bradway e Thompson, 1962). Le differenze tra sessi nelle risposte a test d'intelligenza fluida e cristallizzata non sono statisticamente significative (Cattell, 1987).

Per l'uso applicativo dei test è relativamente poco importante sapere se la base delle differenze è genetica o culturale, ma è fondamentale interpretare i punteggi in modo da non introdurre indebite distorsioni e discriminazioni.

A questo proposito, rispetto alla valutazione dell'**abilità generale** è stato osservato che anche utilizzando i test più distorti "sessualmente" (cioè quelli basati su attività tipiche solo dei maschi o tipiche solo delle femmine) si riesce a ottenere una stima non distorta dell'intelligenza generale (Roznowski, 1987), ovviamente a condizione di usare le norme appropriate. Questa è la ragione per cui i test d'intelligenza, e in specie le batterie multiattitudinali, presentano norme separate per sesso. In piccola parte le differenze tra sessi sembrano dovute a caratteristiche temperamentali differenziali che è possibile tenere sotto controllo con accorgimenti statistici: p.es. si è visto che correggendo i punteggi per tener conto delle "indovinate a caso" si riduce il vantaggio dei Maschi (Ben Shakar e Sinai, 1991).

L'adozione di **norme separate per sesso**, tuttavia, non è una decisione da prendere senza accurata riflessione, in quanto elimina differenze che nella realtà esistono. Ad esempio, se una ragazza educata in una scuola solo femminile ottiene in matematica o in ragionamento spaziale punteggi grezzi inferiori a quelli di un coetaneo educato in una scuola maschile, è "giusto" confrontarne le prestazioni con norme femminili e conseguentemente attribuirle punti standard pari a quelli del ragazzo, così da non penalizzare la diversa appartenenza istituzionale, o non è più sensato riferirsi a norme comuni, darle punti standard più bassi e commentare la differenza fra ragazza e ragazzo riferendosi alla differenza tra scuole?

Analogo è il problema per i **questionari d'interesse**. Se ci si riferisce a norme separate per sesso, l'intensità di un interesse professionale viene confrontata solo con quanto è espresso dai soggetti dello stesso sesso. Ad esempio, un numero relativamente basso di preferenze per items connessi con l'insegnamento (interesse tipicamente femminile) può indurre ad attribuire a un ragazzo un punteggio standard elevato in quest'area, e quindi a consigliargli questo settore occupazionale, mentre una ragazza con lo stesso punteggio grezzo verrebbe considerata poco interessata al settore. In realtà per i questionari d'interesse i problemi di equità più seri riguardano i contenuti degli items (l'operatore dovrebbe accertare al momento della scelta dei test se sono rappresentati contenuti e scale adatti ad ambedue i sessi; cfr. Cronbach, 1984; Tittle, 1979) e le procedure seguite per la validazione delle scale, dato che il separato trattamento di campioni maschili e femminili può di fatto comportare il perpetuarsi di differenze culturalmente indotte (cfr. Anastasi, 1982). Nel commentare i risultati dovrebbero essere presi in considerazione anche i punteggi bassi, segnalando come possibili settori in cui in soggetto non ha sviluppato interessi per mancanza di stimolazione (cfr. Tittle, 1979).

Le ricerche sulle **differenze comportamentali** connesse con differenze tra sessi hanno segnalato, negli ultimi decenni, una grande varietà di prestazioni diversificate (cfr. Gray e Drewett, 1977): riguardo alle soglie sensoriali, riguardo alle capacità percettive, motorie e linguistiche, riguardo alle prestazioni ai test d'intelligenza, riguardo al comportamento affettivo, riguardo al comportamento antisociale, in numerosi fattori di personalità misurati

mediante questionari. Per quel che riguarda la psicopatologia si discute se le donne siano più o meno vulnerabili degli uomini alle malattie mentali in genere e/o se siano più facilmente etichettate in senso psicopatologico (cfr. Williams, 1984). E' stato comunque osservato che psicopatologie specifiche assumono diversa configurazione e decorso a seconda del sesso del paziente (cfr. Copolov et alii, 1990; per la schizofrenia, cfr. Goldstein, 1988).

Alcune differenze tra esseri umani hanno una forte analogia con differenze intersessuali fra animali, il che ha indotto a formulare ipotesi sulla probabile dipendenza genetica e biologica delle differenze stesse. Per esempio, tra i mammiferi, i maschi di quasi tutte le specie sono più aggressivi delle femmine ed esprimono la loro aggressività principalmente in competizioni finalizzate a ottenere dominanza sociale (cfr. J.A. Gray, 1971; Gray e Buffery, 1971). Di fatto, ricerche comparate nell'uomo e negli animali hanno evidenziato una correlazione significativa fra ormoni androgeni (in specie: testosterone) e aggressività (cfr. Gray e Drewett, 1977).

Per la maggior parte delle differenze osservate, tuttavia, è difficile escludere l'influsso prevalente dell'apprendimento sociale, che precocemente stimola comportamenti e interessi diversificati per sesso. Il peso di questo influsso e le sue dinamiche variano a seconda del costrutto e a seconda delle caratteristiche della società: bisogna quindi studiare accuratamente i dati sulla validità dei singoli costrutti per dare interpretazioni accurate.

Ad esempio, gli studi sulla motivazione al successo hanno evidenziato nelle donne un costrutto chiamato "paura del successo" (cfr. Horner, 1972), che nell'esame della personalità femminile va tenuto presente accanto al costrutto - comune a uomini e donne (cfr. Farmer, 1987) - denominato "motivazione al successo" (achievement motive). Va anche tenuto presente che la motivazione al successo, nelle sue relazioni con altre variabili, si connota diversamente per uomini e donne (cfr. Farmer, 1987). Se questi problemi sono stati tenuti presenti al momento della scelta dei test è relativamente facile mettere insieme tutti i dati necessari per una corretta interpretazione; se invece l'interprete dispone, al riguardo, soltanto di dati adatti a comprendere dinamiche maschili, deve segnalare quali sono le informazioni che sarebbe necessario raccogliere per un'interpretazione adeguata.

b) Interpretare in funzione delle condizioni fisiche e dell'anamnesi generale

Gli studi sul rapporto tra condizioni fisiche e caratteristiche personali (cognitive o no) sono poco noti e, nei loro aspetti più deterministici, sono attualmente poco accreditati: ad esempio, i tentativi di individuare la "personalità di base" di diabetici, tubercolotici o portatori di handicap specifici non hanno dato risultati soddisfacenti e generalizzabili. In questo paragrafo vengono menzionati alcuni studi apparsi di recente nella letteratura internazionale, senza pretesa di completezza, soprattutto al fine di stimolare gli psicologi a non dimenticare di raccogliere informazioni riguardo a condizioni fisiche e biologiche dei soggetti che potrebbero essere d'aiuto per interpretare correttamente i dati.

Alcuni di questi problemi sono abbastanza banali. Ad esempio può avvenire che un soggetto abbia un **deficit sensoriale** (ipoacusia o deficit visivi) non precedentemente diagnosticato. Questa situazione, abbastanza frequente tra bambini in età prescolare (sono spesso gli insegnanti i primi a notare deficit di questo tipo) e tra anziani, comporta difficoltà nel decodificare gli stimoli e conseguentemente è associata a risposte inadeguate, ma per motivi che poco hanno a vedere con la struttura psicologica del soggetto. Questa è un'ipotesi interpretativa da considerare quando ci si trova di fronte a risultati inferiori alla media nei test cognitivi.

Nell'esame dei bambini va dedicata particolare attenzione alle informazioni sulla salute in generale e sull'ambiente familiare. È noto infatti che problemi sanitari non individuati tempestivamente possono indurre a ipotesi infondate sulle caratteristiche psicologiche del soggetto. Nei test cognitivi, un eventuale scarso rendimento può essere connesso, oltre che con i deficit sensoriali prima segnalati, con "assenze" dipendenti da una forma epilettica, con uno stato cronico di anoressia da problemi respiratori, con problemi di denutrizione (cfr. correlazioni tra peso corporeo e rendimento cognitivo riferite in Meili, 1967), con uno stato di debilitazione dovuto a un'improvvisa rapida crescita o al succedersi rapido, in un particolare periodo, di "malattie dell'infanzia". Più recentemente è stata segnalata la connessione fra bassi livelli di rendimento a test cognitivi e forme d'intossicazione, rivelate ad esempio dalla concentrazione di piombo nel sangue o nei denti (cfr. Fergusson et alii, 1988a; 1988b; Raab et alii, 1990). Per quel che riguarda l'ambiente familiare è noto il rapporto fra certi "stili educativi" e alcune psicopatologie (cfr. Green, 1988; Kron e Kestenbaum, 1988) o caratteristiche poco desiderabili

della personalità. E' anche noto che figli di genitori alcolizzati presentano problemi cognitivi, comportamentali ed emotivi con maggior frequenza rispetto ai figli dei non alcolisti (cfr. Bennett, Wolin, Reiss, 1988).

Se si rileggono i manuali di psichiatria o di psicologia evolutiva in un'ottica psicodiagnostica si possono ricavare indicazioni utili per l'interpretazione dei dati individuali e, soprattutto, per un'impostazione realistica degli interventi. In alcuni casi, i fattori fisico-genetici imm modificabili sono associati a caratteristiche psichiche su cui si può intervenire con successo. Ad esempio, è stato ripetutamente osservato che i primogeniti mostrano un maggior bisogno di affiliazione, sono più dipendenti e percepiscono come particolarmente stressanti le situazioni in cui non vengono loro dati i rinforzi affettivi che si aspetterebbero (cfr. Zigler e Child, 1969). E' ovvio che l'ordine di genitura non può essere modificato. Tuttavia, se un bambino primogenito riceve punteggi bassi ai test cognitivi e punteggi alti in una scala di Dipendenza, è opportuno interpretare questi dati alla luce di informazioni supplementari sullo stile educativo familiare (p.es.: cercando di sapere se la madre è possessiva e dominante, se c'è una situazione di iperprotezione, ecc.) e cercar di influire sulla situazione familiare o di offrire al ragazzo un ambiente educativo esterno diversamente caratterizzato. In ogni caso, a parità di condizioni, punteggi elevati in dipendenza o reazioni emotive spiccate a improvvise carenze di supporto affettivo vanno considerate meno patologiche in un primogenito.

Indipendentemente dall'ordine di genitura o dall'influsso di fattori fisici, è comunque noto che un livello modesto di rendimento ai test cognitivi (e a scuola) in alcuni casi è associato a rapporti affettivamente insoddisfacenti con i genitori (cfr. Zigler e Child, 1969) e con "altri importanti" (non ultimi gli insegnanti).

Nell'esame degli adulti può essere utile tener presenti i risultati degli studi sul rapporto tra le condizioni fisiche generali e alcune caratteristiche personali. Per quel che riguarda le condizioni generali, si può citare uno studio da cui risulta che il rendimento ai test cognitivi di anziani che non mostravano nessun problema a un'accurata visita medica appare significativamente superiore a quello di coetanei con uno stato di salute meno soddisfacente (cfr. Birren et alii, 1963). Più specificamente, sono state evidenziate connessioni fra risultati inadeguati ai test cognitivi e problemi quali arteriosclerosi e ipertensione arteriosa (cfr. Birren e Spieth, 1962; Spieth, 1964). Indicazioni indirette sul peso del cattivo funzionamento cardiovascolare vengono anche dai numerosi studi (cfr. una rassegna in Doan e Scherman, 1987) che segnalano miglioramenti sistematici nel concetto di sé, nell'ansietà, nella depressione e in altre caratteristiche misurate da questionari generali di personalità (i test più usati sono il Cattell 16 PF e l'EPI di Eysenck) ottenuti dopo interventi basati su esercizi fisici miranti al miglioramento dello stato cardiovascolare. Se si tiene conto anche delle ricerche sulla "personalità di tipo A" (cfr. cap. XIV) si può ragionevolmente concludere che informazioni sul funzionamento del sistema cardiovascolare di un soggetto (da chiedere eventualmente al "medico di famiglia") possono essere utili sia nell'interpretazione del profilo di personalità sia nell'impostazione dell'intervento.

7. Interpretare tenendo conto di fattori biologici e culturali

Non è possibile né utile riassumere articolatamente in questa sede le numerose ricerche sull'incidenza rispettiva dei fattori biologici e ambientali sullo sviluppo intellettuale, emotivo, sociale. Può essere utile però una sintesi dei termini più recenti della polemica, con riferimento alle posizioni di alcuni "genetisti" (Eysenck e Jensen) e "ambientalisti" (Kamin) e qualche cenno alle variabili "non cognitive" su cui sembra influire significativamente l'eredità genetica. Prima di entrare nei dettagli sono necessarie alcune precisazioni.

(a) Quando si parla di "**ereditarietà**" ci si riferisce alla proporzione della variabilità di un tratto (p.es. dell'intelligenza) che, all'interno di una popolazione data, può essere attribuita a fattori genetici (cfr. Stott, 1983). Allo stato attuale della scienza e della tecnica, è impossibile procurarsi misure dirette dell'ereditarietà, che viene "stimata" per lo più in base a correlazioni e confronti tra prestazioni di consanguinei e tra prestazioni di gruppi etnicamente diversi.

(b) Quando si parla di "**ambiente**" ci si riferisce a un'enorme varietà di fattori, che vanno dalle condizioni di vita intrauterina (p.es. in rapporto alla quantità e qualità del cibo o dei farmaci assunti dalla madre, alle eventuali malattie o stress emotivi di questa, ecc.) fino alle differenze fra culture di popolazioni geograficamente lontane e

senza contatti tra loro o vicine e in opposizione politico-sociale; conseguentemente i "fattori ambientali" vengono stimati in base a indicatori di vario tipo, che cambiano a seconda delle ricerche.

(c) Quando si parla di "**intelligenza**" o di "**rendimento cognitivo**" mentalmente ci si riferisce a un'astrazione che mantiene un significato genericamente costante nella mente di ognuno di noi, ma praticamente ci si riferisce a misure ottenute mediante test diversi, che spesso misurano variabili diverse da una ricerca all'altra. Sulle variabili "non cognitive" l'ambiguità è minore solo perché le ricerche sono poco numerose.

(d) **Nessun genetista**, per quanto accanito, sostiene che l'ambiente non interagisce con le caratteristiche genetiche: anzi, è ritenuta largamente accettabile un'interpretazione di "programmazione genetica" come condizionamento delle cellule a reagire in una data maniera alle singole circostanze ambientali (cfr. Oliverio, 1977: in particolare i contributi di Caspari e Denenberg).

(e) **Nessun ambientalista**, per quanto accanito, sostiene che intelligenza e attitudini cognitive siano le uniche caratteristiche umane su cui l'ereditarietà genetica non influisce.

La materia del contendere è quindi "quanto" influiscano rispettivamente i fattori genetici e quanto quelli ambientali. Il metodo della contesa è soprattutto la critica metodologica agli studi compiuti dai sostenitori delle tesi opposte. Critiche facilissime da fare, perché è impossibile - ad esempio - mettere insieme e contrapporre campioni casuali e adeguatamente numerosi di gemelli monozigoti (o di semplici fratelli) rispettivamente allevati nella stessa famiglia o in famiglie diverse per un periodo di anni adeguato e a partire da età molto giovane, è impossibile sintetizzare le migliaia di differenze "ambientali" in indici statisticamente maneggiabili, è impossibile usare test "perfettamente" validi e attendibili. Negli ultimi due decenni, sulla base dei dati raccolti, si è diffusa l'accettazione dell'idea che "tutte le funzioni umane, incluso il comportamento, devono essere il risultato di influenze sia ereditarie sia ambientali" (Baker e Clark, 1990). Di fatto, in campo applicativo questa convinzione teorica comporta l'attribuzione di un ruolo più importante alla componente genetica rispetto alle convinzioni comuni qualche decennio fa, il che induce vivaci discussioni teoriche e pratiche (cfr. il numero speciale dedicato alle "Origini genetiche del comportamento: implicazioni per i consiglieri" dal Journal of Counseling and Development nel 1990).

In questo paragrafo vengono prima affrontati problemi prevalentemente teorici e poi problemi prevalentemente applicativi.

a) Argomenti pro e contro l'ereditarietà genetica delle abilità cognitive

Con queste necessarie premesse, vediamo quali sono gli argomenti portati a sostegno delle tesi genetiste (formulate fin dall'inizio del secolo) e le obiezioni che ad essi sono state successivamente rivolte.

Il gruppo più rilevante di argomentazioni riguarda il **confronto tra prestazioni intellettuali di consanguinei**.

In specie, si sostiene che confrontando il livello intellettuale di **gemelli** raggruppati in funzione di maggiore o minore similarità genetica (monozigoti - dizigoti) e di maggiore o minore similarità ambientale (allevati nella stessa famiglia - in famiglie diverse) l'influsso del fattore genetico appare significativamente superiore a quello del fattore ambientale.

Le critiche degli ambientalisti alla metodologia con cui sono stati condotti i più importanti studi sui gemelli ha portato alla clamorosa identificazione di falsi e imprecisioni nelle ricerche di C. Burt su questo argomento (cfr. Kamin, 1974; Hearnshaw, 1979). Le critiche ad altri importanti studi sull'argomento (Newman, Freeman, Holzinger, 1937; Shields, 1962; Juel-Nielsen, 1965) sono più sottili e di tipo tecnico (cfr. Eysenck e Kamin, 1981; Stott, 1983), tali da indurre a conclusioni sfumate: non sembra ragionevole negare il peso del fattore genetico, ma è molto difficile quantificarlo, sia pure relativamente al peso del fattore ambientale.

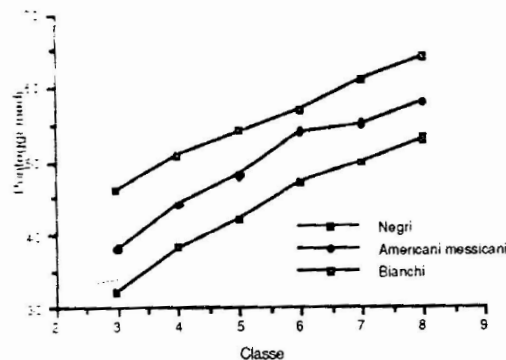
Le ricerche sui bambini adottati inducono a conclusioni dubbie, dato che in gran parte le famiglie adottanti sono scelte dal magistrato in modo da offrire ai bambini un ambiente simile a quello della famiglia d'origine. Lo stesso può dirsi per gli studi sui figli di genitori che erano cugini tra loro, dato che questi nuclei familiari avevano mediamente uno status socioculturale più modesto delle famiglie "di controllo" (cfr. Stott, 1983).

Altre ricerche sulle prestazioni di consanguinei ai test d'intelligenza hanno evidenziato coefficienti di correlazione tra genitori e figli che secondo alcuni studiosi (Erlenmeyer-Kimling, Jarvik, 1963; Eysenck, 1979) vanno da $r = 0,2$ a $r = 0,8$ (mediana: $r = 0,5$), ma che vanno soltanto da $r = 0,1$ a $r = 0,4$ (mediana: $r = 0,33$) secondo altri (Kamin, 1974). Per di più, è difficile attribuire queste correlazioni solo all'ereditarietà e non anche all'ambiente, data la convivenza dei nuclei familiari. Analizzando un insieme di ricerche sulla famiglia, l'ereditabilità in senso lato è stata stimata intorno a 0,51 (Chipner, Rovine, Plomin, 1990).

L'analisi ha evidenziato anche che l'influsso di uno stesso ambiente è diverso per tipi di parenti diversi (p.es. gemelli, fratelli, fratellastri, cugini) e che è importante tener conto dei fattori selettivi nella scelta del partner (assortative mating). Studi recenti sulle **"tendenze secolari" dell'ereditabilità dell'intelligenza** segnalano, inoltre, che alcune variazioni ambientali-culturali sembrano modificare in qualche misura il peso dell'ereditarietà (cfr. Sundet et alii, 1988).

FIG. XX.4

Punteggi medi alle PM di alunni di scuola elementare statunitensi (da Eysenck e Kamin, 1981, p. 80)



Un altro gruppo di ricerche condotte dai sostenitori dell'ipotesi genetica riguarda le **differenze tra sessi**. Il ragionamento soggiacente a queste ricerche, riassunto qui appresso, è meno ovvio dei precedenti. E' stato segnalato (Lehrke, 1978) che la distribuzione dei QI tende ad avere una gamma più ampia per i maschi che per le femmine. Per questo fenomeno è stata formulata da Lehrke un'ipotesi esplicativa riproposta da H.J. Eysenck: "La differenziazione sessuale negli animali superiori dipende dall'abbinamento fra cromosomi sessuali: due cromosomi X per le femmine e, per i maschi, un X e un Y. Il cromosoma X nell'uomo è di media grandezza, contiene circa il 5 o il 6% del materiale genetico e trasporta all'incirca la stessa proporzione d'informazione genetica, inclusi i geni che sappiamo influire su ogni sistema corporeo più rilevante. Il cromosoma Y, d'altro lato, è uno dei più piccoli e, per quel che sappiamo, trasporta soltanto istruzioni genetiche per la mascolinità. Mentre tutti gli altri cromosomi operano in coppia, nel caso dei cromosomi sessuali questo abbinamento avviene solo nelle femmine: nei maschi il cromosoma Y non si appaia al cromosoma X. Ne risulta che nei maschi i geni contenuti nel cromosoma X, qualunque essi siano, troveranno la loro espressione senza nessuna interferenza da parte dei cromosomi Y, mentre nelle femmine il secondo cromosoma X ridurrà l'impatto di qualsiasi informazione genetica sia contenuta nel primo. Questo rende più probabile geneticamente che i maschi mostrino un qualche eccesso all'una o all'altra estremità della scala e dà conto della loro maggior variabilità, come ha sottolineato Robert Lehrke. Ne consegue un'ipotesi in termini operativi, verificabili: "se i geni di primaria importanza relativi all'intelligenza sono contenuti nel cromosoma X, ci si può aspettare che siano molto simili le correlazioni tra punteggi dei test per madre-figlia, padre-figlia e madre-figlio, perché in ciascuno di questi casi genitore e figli hanno un cromosoma X in comune. Le correlazioni tra padre e figlio dovrebbero invece essere più basse, perché non hanno un cromosoma X in comune, e le correlazioni fratello-sorella dovrebbero essere intermedie, dal momento che hanno un cromosoma X in comune in metà di casi." (Eysenck e Kamin, 1981).

I dati che Eysenck porta a sostegno di questa ipotesi si riferiscono a un solo studio, su 51 famiglie; Kamin lo controbatte con più numerosi dati in contrasto. La polemica è tuttora in corso ed è probabile che continuino a prodursi ricerche al riguardo. Gli argomenti riguardanti i confronti tra gruppi etnici risalgono al testing di massa operato sulle reclute della prima guerra mondiale e si concretizzano principalmente nel confronto tra statunitensi di ceppo anglosassone e di altra origine: negri, neolatini, irlandesi, ebrei, giapponesi e cinesi (questi ultimi tre gruppi ottengono in genere risultati superiori anche agli anglosassoni nei test d'intelligenza). Neanche gli ambientalisti mettono in dubbio che il rendimento dei negri americani ai test d'intelligenza (inclusi quelli "non verbali": cfr. fig. XX.4) sia inferiore di circa 15 punti di QI - una deviazione standard nelle scale Wechsler - al rendimento medio dei bianchi, qualunque siano i test usati (si veda la rassegna di ricerche in: Loehlin,

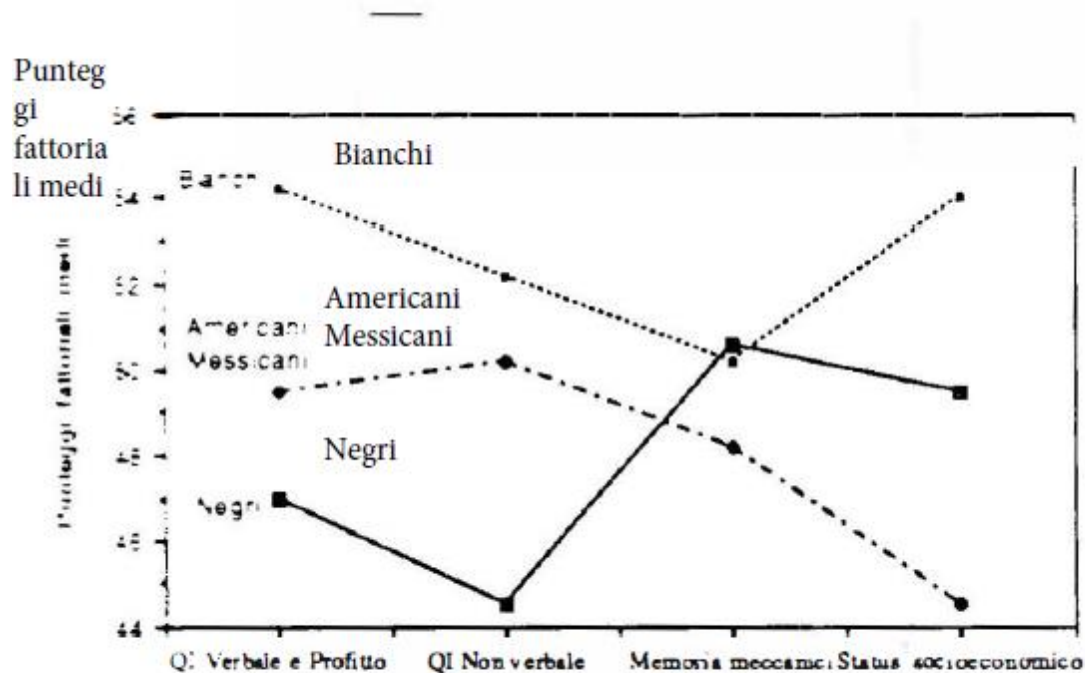
Lindzey, Spuhler, 1975).

Le **ipotesi interpretative sulla qualità delle differenze** si basano su ricerche più complesse. Ci limitiamo a darne qualche esempio. La fig. XX.5 riproduce i risultati di una ricerca di A. Jensen mirante a sostenere l'argomento che le prestazioni ai test d'intelligenza dei negri sono inferiori per ragioni genetiche e non socioculturali (Jensen, 1980).

I punteggi in ciascuna variabile sono calcolati dopo aver escluso, con un procedimento statistico, l'influsso delle altre variabili. Il rendimento ai test "culturali" differenzia nettamente le tre etnie (media = 55 per gli anglosassoni, 49 per i messicani-americani, 46 per i negri americani), il rendimento ai test non verbali, una volta escluso l'influsso delle componenti socioculturali e della memoria, è uguale per anglosassoni e messicani (media: 50) e inferiore per i negri (media: 47), il rendimento alla prova di memoria meccanica è simile per anglosassoni e negri (media: 50-51) e inferiore per i messicani (media: 43) e infine, a parità di rendimento ai test cognitivi lo status socioeconomico è più alto per gli anglosassoni (media: 54), di livello intermedio per i negri (media: 49) e più basso per i messicani (media: 44). Questi dati inducono Jensen e Eysenck a considerare i messicani come "culturalmente deprivati" ma pari agli anglosassoni nelle prestazioni più caratterizzanti il patrimonio genetico, mentre i negri non sarebbero svantaggiati sotto l'aspetto sociale ("status socioeconomico"), ma soltanto sotto l'aspetto cognitivo.

FIG. XX.5

Punteggi fattoriali medi di scolari (5a elementare) in variabili diverse (adattato da Eysenck e Kamin, 1981, p. 81)



Altri studiosi evidenziano invece che, fatto pari lo status socio-culturale, le differenze nelle prestazioni intellettuali (QI alla scala Wechsler) di bianchi e negri statunitensi si riducono della metà (Cronbach, 1984, p. 322). Viene inoltre sottolineato che è praticamente impossibile "rendere pari" per status socio-culturale gruppi di bianchi e di negri statunitensi: il titolo di studio non è indicativo, perché è probabile che i bianchi abbiano frequentato scuole migliori, la qualifica professionale non è indicativa, perché una stessa qualifica (p.es.: "imprenditore") riferita a un negro indica solitamente uno status socioeconomico inferiore, e così via (Cronbach, 1984). A questo va aggiunto un argomento di cui è difficile valutare appieno il peso: i negri statunitensi sono stati deportati come schiavi in America (il che fra l'altro implica che non sono arrivati lì per iniziativa e per scelta di un male minore o aspirazione a un benessere maggiore, come gli emigranti dall'Europa o dall'Asia) e furono mantenuti schiavi per secoli. E' un'esperienza di massa unica, non confrontabile con altre nella storia contemporanea: chi può dire quali siano gli effetti che il senso di estraneità, di rifiuto reciproco a forte carica emotiva, di sentimenti conflittuali tra la classe dominante e gli schiavi o ex-schiavi ha suscitato a livello cognitivo nei negri? (cfr. Butcher, 1972b). Questa situazione è difficile da comprendere per un europeo. Per capire meglio l'entità e la complessità delle differenze si pensi alla minoranza etnica che

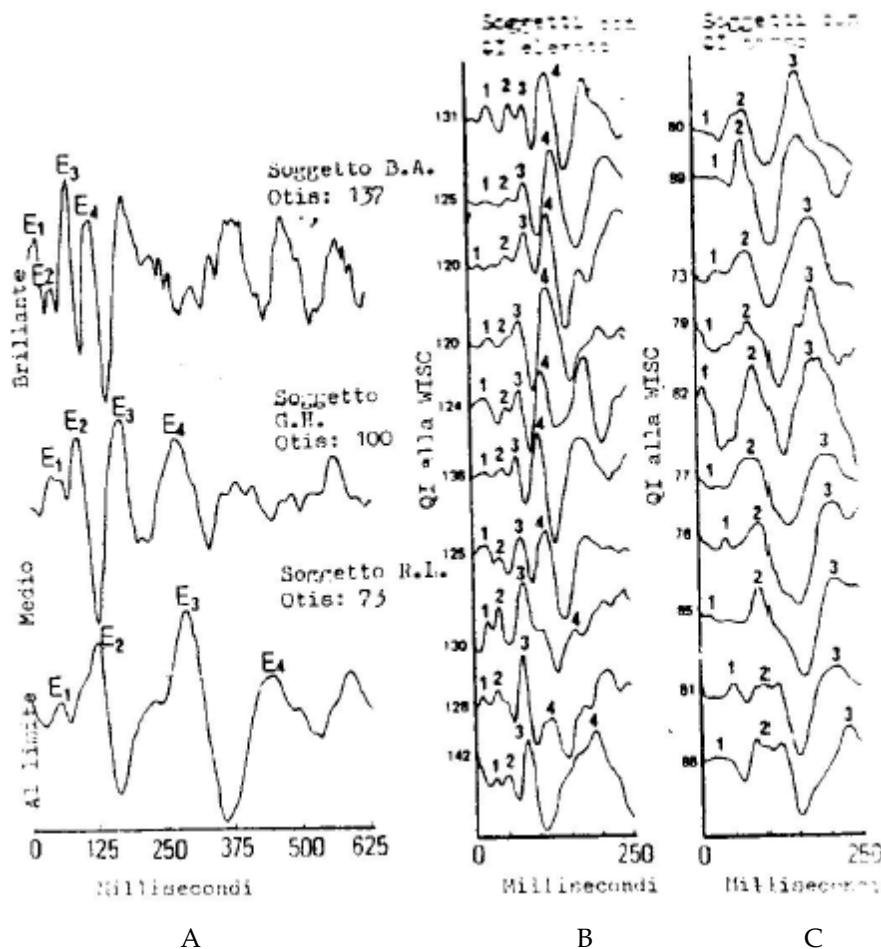
in Italia - secondo le ricerche psicosociali - si colloca al gradino più basso della scala socio-culturale: gli zingari. Gli zingari in Italia non sono mai stati schiavi, sono sempre andati e venuti a loro piacere, hanno sofferto persecuzioni razziali (insieme ad altre minoranze) in periodi di tempo circoscritti, sono emarginati socialmente ma non legalmente. Tuttavia, se venissero pubblicate ricerche da cui risultasse un minor rendimento degli zingari ai test d'intelligenza, pochi lettori sarebbero sorpresi. Questo ipotetico risultato peraltro non demolirebbe lo stereotipo dello zingaro "furbo", in grado di imbrogliare anche chi ottiene un buon QI ai test d'intelligenza. I risultati cioè si presterebbero ad essere interpretati secondo le convinzioni personali di chi li legge.

Di fatto, le numerose ricerche su quanto la differenza dipenda da fattori genetici e quanto da fattori ambientali non hanno condotto a nessun risultato pienamente accettabile dalle due parti. Commenta Kamin: "Non c'è modo di fornire una risposta conclusiva al quesito sulle differenze tra bianchi e neri finché e se non siamo capaci di costruire una società in cui bianchi e neri siano esposti ad ambienti parimenti favorevoli e non-discriminatori. L'ironia è che se riusciamo a costruire una tale società, nessuno avrà più interesse a rispondere al quesito." (Eysenck e Kamin, 1981).

Oltre a sostenere i propri argomenti, Eysenck critica pesantemente le ricerche sul **rapporto ambiente-intelligenza**, in particolare quelle che evidenziano un incremento nel QI di bambini trasferiti, abbastanza precocemente, ad un ambiente più arricchente. Come esempio, riassume una ricerca condotta a Milwaukee, in cui 20 soggetti sperimentali rigorosamente appaiati a 20 di controllo, vennero seguiti dalla nascita fino al primo anno di scuola elementare con interventi di stimolazione sui bambini (estesi ad abilità sensoriali, motorie, linguaggio e pensiero) e interventi "educativi" per le madri. Lo sviluppo di ambedue i gruppi venne misurato periodicamente fino al nono anno di età. A questa età, il gruppo di controllo aveva un QI medio intorno a 80, mentre la media del gruppo sperimentale era intorno a 104 (Heber, Garber, 1977). Eysenck ritiene i risultati non probanti, sia perché il QI dei bambini potrebbe ancora variare prima dell'età adulta sia perché "il profitto non è maggiore di quel che ci si potrebbe aspettare in base al modello genetico" (Eysenck e Kamin, 1981): argomenti che a più di un lettore sembrano poco oggettivi: la misurazione su cui è basato il confronto è avvenuta tre anni dopo la fine dell'intervento, il che è una sufficiente garanzia di stabilità (cambiamenti ulteriori potrebbero anche essere determinati da fattori ambientali successivamente intervenienti) e la differenza fra i due gruppi è statisticamente significativa e di entità rilevante almeno quanto quella delle differenze fra bianchi e negri riportate prima. E' comunque documentato da numerosi studi che interventi di stimolazione appropriati, condotti simultaneamente sull'ambiente familiare e su quello scolastico, coinvolgendo in un ruolo attivo genitori e insegnanti, producono un aumento significativo delle prestazioni ai test cognitivi (di "pensiero convergente" o di creatività) e un miglioramento nel concetto di sé, su ragazzi di qualsiasi livello cognitivo appartenenti a famiglie di basso status socioculturale (cfr. AA.VV., 1984; Karnes e Johnson, 1987; Wirth, 1978).

FIG. XX.6

Potenziali evocati medi di soggetti diversi per QI (da Eysenck, 1979)
A Stesso test (Otis) e QI diversi – B Stesso test (WISC IQ) 10 Ss con QI elevati
C Stesso test (WISC IQ) 10 Ss. con QI basso



Nei grafici B e C il QI di ciascun Soggetto è indicato nell'asse a sinistra

Le prove più convincenti della base genetica dell'intelligenza dovrebbero comunque essere ricercate nel **rapporto diretto tra misure psicofisiologiche del funzionamento del s.n.c. e rendimento nei test d'intelligenza**. I dati disponibili attualmente sono pochi e lacunosi. Qui appresso vengono riassunti brevemente alcuni risultati riguardanti i tempi di reazione e i "potenziali evocati" (Eysenck, 1979; Eysenck e Kamin, 1981), insieme ad alcune critiche rivolte alle ricerche da cui sono desunti.

(a) Ricerche sui **tempi di reazione** sono state condotte da Jensen (Jensen, 1980), che riferisce solo una correlazione "all'incirca - 0,40" calcolata su un campione di 39 ragazze quattordicenni, mentre per altri sei studi riporta in forma grafica le pendenze della retta di regressione, in modo tale da render difficile una stima dell'entità delle correlazioni. Kamin afferma di aver chiesto più volte a Jensen i dati grezzi delle sue ricerche, sembrandogli improbabile che i negri avessero tempi medi di reazione più bassi dei bianchi pur eccellendo in sport come la pallacanestro e il pugilato, e di non averli mai ricevuti (Eysenck e Kamin, 1981). Per di più, gli esperimenti di Jensen comportano una modalità di risposta particolare: i soggetti non devono limitarsi a premere automaticamente un bottone quando appare lo stimolo, ma devono premere un pulsante che si accende, in ordine casuale, nell'ambito di una serie (fino a otto pulsanti alternativi). Questa attività è ovviamente più complessa della reazione semplice e verosimilmente implica processi cognitivi più complessi e numerosi (Stott, 1983).

(b) Altre ricerche riguardano i **potenziali evocati**, ossia i tracciati che si ottengono all'EEG quando uno stimolo visivo o uditivo viene presentato improvvisamente. Alcune ricerche (cfr. Eysenck, 1979; Eysenck e Kamin, 1981; Ertl, 1968; Ertl, Schafer, 1969) mostrano che i potenziali evocati variano a seconda del QI dei soggetti (cfr. fig. XX.6) e che sono correlati positivamente con la riuscita in test di G_r.

Un'ipotesi esplicativa (formulata da A. Hendrikson) è che "quando un messaggio passa da un neurone all'altro attraverso la corteccia (...) possono capitare degli errori. Quanto maggiore è la probabilità che capitino un errore, tanto maggiore è la difficoltà di un soggetto nel risolvere problemi cognitivi. Il potenziale evocato rivelerebbe il numero di errori che capitano durante la trasmissione. Si rammenti che un potenziale evocato è in realtà il potenziale evocato medio di numerose trasmissioni. Errori di trasmissione, postula Hendrikson, avrebbero l'effetto di rendere l'onda più "liscia", tanto da perdere molti dei "picchi"

caratteristici di una trasmissione priva di errori. Finché non ci sono errori, quanto più un problema è difficile tanto più numerosi saranno gli "alti e bassi" di un'onda." (Eysenck e Kamin, 1981, p. 72).

Sia i dati sia l'interpretazione sono suggestivi, ma se si esamina l'insieme delle ricerche disponibili si trovano praticamente altrettanti risultati a favore e contro l'ipotesi (cfr. Rust, 1975).

In aggiunta a questi due gruppi di ricerche, viene anche citata (Eysenck, 1979) una ricerca di dottorato di W.L. Bastendorf, da cui risultano correlazioni significative tra QI e conduttanza palmare in campioni di bambini (9 e 12 anni). Questa ipotesi sembra tuttavia ancora ad uno stadio embrionale: necessita per lo meno di chiarificazioni riguardo alla possibile interferenza di tratti e stati emotivi (p.es.: ansia).

Gli studi sull'ereditarietà delle abilità cognitive sono prevalentemente stati condotti in riferimento a misure di abilità generale ("QI"). E' presumibile che i risultati sarebbero più articolati in riferimento a un quadro multifattoriale: è ragionevole presumere che la varianza in alcuni fattori sia prevalentemente di origine genetica e in altri sia prevalentemente effetto di influssi ambientali. Le ricerche in quest'ottica non sono però ancora in grado di offrire elementi utili allo psicologo applicativo (cfr. Carroll, 1983).

Questa rapida rassegna di studi sul problema eredità-ambiente non si può concludere senza un accenno all'importanza dell'interazione fra i due aspetti considerati. P.E. Vernon conclude una rassegna critica di studi con le riflessioni seguenti, che molti possano condividere. "La cosiddetta covarianza genetico-ambientale è stata trascurata dalla maggior parte degli autori di analisi sull'ereditabilità. Questo termine si riferisce all'ovvia probabilità che genitori intelligenti, che trasmettono geni superiori ai loro discendenti, solitamente provvedono anche ambienti superiori alla media. (...) Si noti anche che la percentuale genetica è molto al di sotto dell'80% che Jensen sostenne nel 1969 in base alle cifre dei test collettivi di Burt. Io penso che molto probabilmente, se potessimo testare campioni più ampi in condizioni più controllate, la percentuale genetica potrebbe cadere al di sotto del 60%. Ma la cifra esatta non importa, purché si riconosca che sia i geni sia l'ambiente hanno un effetto molto sostanziale sul QI del bambino e che la componente genetica sia probabilmente la più ampia, il che non possiamo permetterci di ignorare." (P.E. Vernon, 1979).

b) Ipotesi sull'eredità genetica di caratteristiche personali "non cognitive"

Un quadro generale che sintetizza vari studi sull'eredità genetica di caratteristiche personali "non cognitive" è riprodotto nella tab. XX.2 (da Loehlin, 1977). La maggior parte delle ricerche sono state svolte nell'ambito delle teorie di Eysenck e di Cattell (accennate nel. cap. XV).

Come si vede, la mediana delle correlazioni calcolate fra gemelli identici, in una varietà di studi e di questionari, è nettamente superiore ($r = 0,48$) alla mediana ($r = 0,24$) dei coefficienti calcolati fra gemelli dizigoti. Questi dati sommari dicono che gli effetti della componente genetica sono evidenti anche tra le variabili "non cognitive", nonostante la modesta attendibilità dei questionari di personalità rispetto ai test d'intelligenza, le differenze di ereditabilità tra una variabile e l'altra e le differenze fra metodi d'indagine (si vedano descrizioni di vari metodi in Loehlin, 1977 e in Fulker, 1981).

Le informazioni sul diverso grado di ereditabilità delle caratteristiche personali servono allo psicologo per interpretare correttamente alcuni insiemi di dati (p.es. su gruppi familiari), per non omettere la raccolta di informazioni importanti nel fare l'anamnesi dei soggetti, per formulare previsioni più realistiche sulla presumibile evoluzione spontanea di alcune caratteristiche (p.es. nella diagnosi di bambini e adolescenti) e, soprattutto, per identificare obiettivi d'intervento realistici e mettere in atto strategie efficaci.

TAB. XX.2

Correlazioni intraclassa tra gemelli identici (I) e fraterni (F) in varie scale di questionari di personalità (da Loehlin, 1977)

Studio	Numero di coppie		Correlazione mediana		Numero delle scale	Questionario*
	I	F	I	F		
Carter (1935)	40	43	.60	.33	4	Benreuter
Newman, Freeman, & Holzinger (1937)	50	50	.56	.37	1	Woodworth
Vanderberg (1962)	45	35	.55	.00	7	Thurstone

Gottesman (1963)	34	34	.45	.21	31	HSPQ,MMPI
Wilde (1964 a)	88	42	.50	.34	4	Amsterdam-BV
Gottesman (1966)	79	68	.48	.31	18	CPI
Partaanen, Bruun & Markkanen (1966)	157	189	.26	.18	4	Speciale
Nichols (1965)-Maschi	207	126	.53	.25	24	CPI
Nichols (1965)-Femmine	291	193	.50	.35	24	CPI
Vandenberg et al. (1967)	111	92	.47	.24	12	Comrey
Schoenfeldt (1968)-Maschi	150	53	.40	.36	7	Fattori
Schoenfeldt (1968)-Femmine	187	103	.49	.33	7	Fattori
Canter	39	44	.37	.15	31	EPI,Foulds,16PF
Eaves & Eysenck (1975)-Maschi	120	59	.42	.21	2	PEN
Eaves & Eysenck (1975)-Femmine	331	198	.42	.15	2	PEN
Horn Plomin, & Rosenman (1976)	99	99	.46	.18	18	CPI
Mediana	2028	1428	.48	.24		

* HSPQ = High School Personality Questionnaire; CPI = California Psychological Inventory; MMPI = Minnesota Multiphasic Inventory; BV = Biographical Questionnaire; EPI = Eysenck Personality Inventory; PF = Personality Factor; PEN = Psychoticism, Extraversion, Neuroticism.

Riguardo ai **fattori di Eysenck** vengono riferiti un certo numero di studi, per lo più condotti su un ristretto numero di soggetti.

Confrontando i coefficienti di correlazione tra punteggi di **neuroticismo** di gemelli monozigoti cresciuti nella stessa famiglia o cresciuti in famiglie diverse sono state trovate coppie di valori quali: $r = 0,53$ rispetto a $r = 0,38$ (Shields, 1962: 42 e 43 coppie), $r = 0,56$ rispetto a $r = 0,58$ (Newman, Freeman, Holzinger, 1937: 50 coppie e 19 coppie); mediana, tra campioni maschili e femminili, $r = 0,53$ rispetto a $r = 0,25$ (Loehlin e Nichols, 1976: per i monozigoti, 197 coppie di maschi e 284 di femmine, per i dizigoti, 122 coppie di maschi e 190 di femmine). Due studi condotti su gruppi familiari utilizzando il test di Bernreuter (Carter, 1933; Crook, 1937) evidenziano in modo peculiare il peso dell'affinità genetica: la correlazione tra genitori (79 coppie) è $r = 0,07$ (consanguineità nulla e correlazione nulla), tra genitori e figli (134 coppie) è $r = 0,31$, tra fratelli o sorelle allevati insieme (101 coppie) è $r = 0,25$, fra gemelli dizigoti di sesso diverso ($N = 34$) è $r = 0,18$, fra gemelli dizigoti dello stesso sesso ($N = 44$) è $r = 0,32$, fra gemelli monozigoti ($N = 55$) è $r = 0,63$ (massimo di patrimonio genetico comune e massimo di correlazione).

I coefficienti di correlazione riguardanti l'**estroversione** trovati nello studio di Shields già citato sono $r = 0,61$ tra gemelli monozigoti cresciuti nella stessa famiglia, rispetto a $r = 0,42$ fra gemelli monozigoti cresciuti in famiglie diverse (Shields, 1962: 43 coppie); in un altro studio, la mediana dei coefficienti di correlazione, fra campioni maschili e femminili, era $r = 0,60$ per gemelli monozigoti e $r = 0,24$ per gemelli dizigoti (Loehlin e Nichols, 1976: per i monozigoti, 197 coppie di maschi e 284 di femmine, per i dizigoti, 122 coppie di maschi e 190 di femmine). Studi condotti sottoponendo ad analisi della varianza i risultati ottenuti al questionario di Eysenck da gemelli monozigoti e dizigoti evidenziano un effetto statisticamente significativo dell'ereditarietà, quantificato nel 42% (57% se corretto per la scarsa attendibilità).

In base ai dati su nevroticismo ed estraversione Fulker (1981) conclude che, tenendo conto dell'imperfetta

attendibilità delle misure, l'ereditarietà ha un peso di circa il 70%.

Anche sui **fattori di Cattell** sono state trovate indicazioni analoghe.

I risultati di uno studio condotto con una vecchia edizione dell'HSPQ di Cattell su coppie di gemelli adolescenti (Cattell, Blewett, Beloff, 1955), rianalizzati da Fulker (1981), evidenziano: (a) per quel che riguarda il **nevroticismo** un'ereditabilità quantificata in 0,31 (0,60 se s'introduce una correzione per la scarsa attendibilità) per il fattore C e in 0,42 (0,76 se s'introduce una correzione per la scarsa attendibilità) per il fattore Q₄; (b) per quel che riguarda l'estroversione un'ereditabilità quantificata in 0,37 (1 se s'introduce una correzione per la scarsa attendibilità) per il fattore A e in 0,40 (0,83 se s'introduce una correzione per la scarsa attendibilità) per il fattore Q₃.

Studi sulla **personalità nevrotica** condotti con la **batteria di test oggettivi di Cattell** (descritta nel cap. XVII) evidenziano (cfr. Delhees, 1977) una dominanza della componente genetica nell'**inibizione** (U.I. 17), nell'**indipendenza** (U.I. 19) e nell'**esuberanza** (U.I. 210, ma soprattutto nell'**intelligenza "cristallizzata"** (U.I. 1) e nel **conformismo** (U.I. 20).

Una rassegna di studi riferiti ai **fattori di primo ordine** definiti da R.B. Cattell segnalano (Delhees, 1977) una spiccata ereditabilità dei fattori **D** ("**eccitabilità**": non è presente nei questionari per adulti), **E** ("**dominanza**") e **Q₃** ("**sentimento di sé**"), ma soprattutto **B** ("**intelligenza cristallizzata**") e **H** ("**parmia**", **audacia**). In un contributo più recente, R.B. Cattell conclude che "Il rapido avanzamento della genetica del comportamento nell'ultimo decennio mostra che una quantità di tratti d'origine, come l'**intelligenza fluida**, **gf**, l'**abilità spaziale**, **S**, la **surgeny** [variabilità di umore, n.d.T.] **F**, e la **premsia** [sensibilità emotiva protetta, n.d.T.] **I**, hanno un'**ereditabilità** così **alta** che le loro unità devono in gran parte emergere dalla maturazione interna coordinata e forse dalla dipendenza da radici fisiologiche, come le dimensioni della corteccia per **gf** o un qualche stimolatore ("pacemaker") chimico per **F**. Tuttavia, altri tratti di origine, come la **forza del super-Io G**, la **tendenza al senso di colpa contrapposta all'autostima O**, il **radicalismo-conservatorismo Q₁** e la **tensione ergica Q₄**, che hanno **ereditabilità relativamente trascurabile**, devono originarsi dall'esposizione all'apprendimento in istituzioni unitarie, p.es. l'addestramento domestico ecc., che simultaneamente nell'insieme costruiscono un intero anello di comportamenti specifici" (Cattell, 1986a, pp. 7-8).

Anche i risultati degli studi su **variabili psicopatologiche** evidenziano un influsso della componente genetica. Ad esempio, in una ricerca statunitense (McClearn e DeFries, 1973) si è trovato che il valore medio delle correlazioni tra 120 coppie di gemelli monoigoti, per le 10 scale standard del MMPI, era $r = 0,40$, mentre il valore medio delle correlazioni tra 132 coppie di gemelli dizigoti, per le stesse scale, era $r = 0,20$. Com'è noto, nella letteratura è segnalato in particolare un influsso genetico per la **schizofrenia** (cfr. DeFries e Plomin, 1978; Fulker, 1981). Applicando l'analisi della varianza a punteggi di **Psicoticismo** ottenuti all'EPQ da gemelli monoigoti e dizigoti (Eaves e Eysenck, 1975) è stato trovato un effetto statisticamente significativo dell'ereditarietà, quantificato in 46% (81% se corretto per la scarsa attendibilità).

Nell'insieme, i dati finora disponibili sulla componente ereditaria di alcune caratteristiche personali autorizzano a fondare ipotesi di studio, ma non consentono di avere certezze su casi singoli. Per di più, i coefficienti di correlazione sono di solito modesti (intorno a $r = 0,40$: corrisponde al 16% di varianza in comune) anche tra consanguinei che hanno il massimo di patrimonio genetico (e spesso di condizioni ambientali) in comune. Conseguentemente, anche se i membri di una stessa famiglia mostrassero punteggi molto simili e "anormali" in certe variabili, non sarebbe giustificato ipotizzare che quelle caratteristiche sono di origine genetica e quindi immutabili: anche se la componente genetica è realmente presente, il suo peso non è tale da impedire modificazioni significative del comportamento nella maggior parte dei casi.

Di fatto, attualmente incontra molto favore l'approccio pragmatico suggerito da alcuni Autori (cfr. Angoff, 1988a), che si può sintetizzare negli argomenti seguenti: sapendo che molte caratteristiche ereditate si possono modificare e, viceversa, che molte caratteristiche indotte dall'ambiente sono estremamente resistenti ai cambiamenti, sotto l'aspetto pratico la problematica più importante è la modificabilità delle caratteristiche in funzione di benefici personali e sociali e non la proporzione della loro ereditabilità.

c) Il problema dei test "culturalmente equi"

La polemica teorica riguardo all'influsso rispettivo di "eredità" e "ambiente" sul rendimento ai test comporta, a livello applicativo, alcuni corollari importanti, da tener presenti sia nella selezione degli strumenti (cfr. cap. XIX) e nell'interpretazione dei risultati, sia nelle decisioni di politica scolastica e di politica delle assunzioni e promozioni in ambito lavorativo. Qui vengono esposte solo alcune considerazioni sulle implicazioni riguardanti l'interpretazione dei risultati dei test somministrati nel modo "standard", senza affrontare la trattazione approfondita degli aspetti politici dell'uso dei test, che esigerebbe un contesto diverso da quello di questo manuale. Dei test d'intelligenza somministrati in modo "dinamico", cioè in stretta interazione con interventi di modificazione cognitiva, si è parlato nel cap. XIX e in quella sede si è anche accennato alle modalità

d'interpretazione di quei risultati.

Gli stimoli di cui sono costituiti i **test d'intelligenza**, i criteri per l'identificazione delle risposte esatte, l'assunto motivazionale alla base della situazione di testing (si suppone che ciascuno cerchi di dare quante più soluzioni esatte, il più rapidamente possibile: i test cognitivi valutano la "prestazione massima") sono il prodotto di quella che genericamente si chiama "cultura occidentale". La familiarità con gli stimoli impiegati (lingua standard, figure geometriche astratte, ecc.: cfr. cap. X e XI) è anch'essa un aspetto della culturalità, ma non è certo l'unico. I tipi di problema da risolvere e le soluzioni esatte sono connessi con la logica aristotelica (blandamente) e con l'empirismo filosofico anglosassone; l'assunto motivazionale - a noi mediterranei - sembra prevalentemente connesso con l'etica protestante, che assegna un ruolo fondamentale all'impegno nel lavoro e all'economia del proprio tempo; le modalità di svolgimento dei compiti sono tipiche di strutture sociali in cui è portatore di responsabilità l'individuo come singola persona, e non solo la famiglia o il gruppo sociale. Anche i test "non verbali", sotto questi aspetti generici ma fondamentali, sono "culturali": è prevedibile che a un qualsiasi test ("cognitivo" o "non cognitivo") ottengano risultati medi migliori i membri della cultura di cui i test sono l'espressione, e che i risultati medi siano tanto più "scadenti" quanto più la cultura di appartenenza differisce da quella espressa dal test nel modo di intendere l'"intelligenza", nelle convinzioni riguardanti la logica, nell'approccio motivazionale al lavoro, nella gerarchizzazione del valore socialmente attribuito ai vari tratti di personalità.

Qualche esempio concreto può chiarire meglio la natura di queste differenze.

Il concetto di intelligenza non è detto che sia identico da una cultura all'altra. Anche dando per scontato che siano universali connotazioni importanti (p.es.: "capacità di risolvere problemi per utilizzare o modificare l'ambiente a proprio vantaggio"), è probabile che ambienti molto diversi abbiano indotto le subculture (polinesiani, eschimesi, bantu, europei di città, paese, montagna) a selezionare come importanti alcuni problemi e strategie di soluzione e a ignorarne altri. E' intuitivo che i comportamenti "intelligenti" che un bambino deve apprendere o "inventare" per sfruttare al meglio il suo ambiente se vive in un villaggio del Ciad sono diversi da quelli utili per un bambino altrettanto povero che vive in una baracca a Calcutta o per un bambino, indiano come lui, ma che appartiene ad una famiglia benestante di casta elevata o per il figlio di un bancario residente a New York. La selezione delle strategie di soluzione più adeguate si è cristallizzata talvolta in sistemi logici veri e propri: p.es. studenti universitari provenienti dalla Cina o dal Giappone hanno difficoltà nello studio dei nostri maggiori sistemi filosofici, non riescono a considerare "logico" quel che noi consideriamo "logico", e presumibilmente lo stesso avverrebbe a noi se provassimo a studiare i loro scritti filosofici tradizionali. Sono molto divulgati gli esempi di soluzioni considerate "esatte" a test come le PM o i cubi di Kohs da popolazioni africane. Gli esempi nella fig. XX.7 si riferiscono al test dei cubi di Kohs (cfr. cap. IX, paragrafi riguardanti le scale di Wechsler); il tratteggio indica le delimitazioni fra cubi. Il perché degli "errori" diventa comprensibile se si sa che i soggetti non consideravano nemmeno la possibilità di "costruire" un triangolo rosso più grande mettendo vicine le due parti triangolari rosse di cubi diversi, e così via (Wober, 1967 in P.E. Vernon, 1969). Le differenze qualitative e quantitative nelle risposte al test di Kohs e ad altri test non verbali sono state osservate anche in situazioni di relativa prossimità culturale (p.es.: tra bambini di famiglie francesi e di famiglie algerine immigrate in Francia, pari per livello socioeconomico, all'interno delle stesse classi elementari). Gli AA. che hanno segnalato queste differenze formulano un'ipotesi esplicativa interessante (Ionescu, Jourdan-Ionescu, Toselli-Toschi, 1983), basata sulla teoria della differenziazione psicologica (cfr. cap. X): in alcune culture vengono privilegiate pratiche educative che favoriscono lo sviluppo dell'autonomia, dell'atteggiamento analitico, dell'ingegnosità, il che stimola lo sviluppo dell'"indipendenza dal campo" e delle capacità percettive, mentre altre società non propongono ai bambini giochi che favoriscono lo sviluppo dell'abilità motoria, delle capacità percettive, dell'analisi, non premiano l'eccellenza dei risultati individuali (la coesione del gruppo è fortissima) e puniscono la violazione delle regole tradizionali (quindi, l'innovazione). E' prevedibile che persone appartenenti a culture differenziate sotto gli aspetti socio-educativi prima descritti ottengano risultati significativamente diversi ai test sensibili alle capacità percettive e alla differenziazione psicologica, cioè in pratica a tutti i test di fattore g.

Altrettanto note sono le ricerche sulle differenze motivazionali e temperamentali tra statunitensi anglosassoni e amerindi: per i membri di alcune popolazioni indiane la "fretta" con cui lavorano gli statunitensi è indice di squilibrio mentale e per i membri di altre popolazioni è estremamente sveniente riuscire meglio degli altri in

un compito (Erikson, 1966).

Se volessimo, a scopo di ricerca "pura", misurare l'intelligenza di popolazioni così disparate come quelle appena elencate, dovremmo quindi anzitutto identificare un gruppo di comportamenti ritenuti "intelligenti" in tutte le culture poste a confronto e un insieme di compiti ugualmente familiari e potenzialmente rivelativi di quei comportamenti: non appare facile.

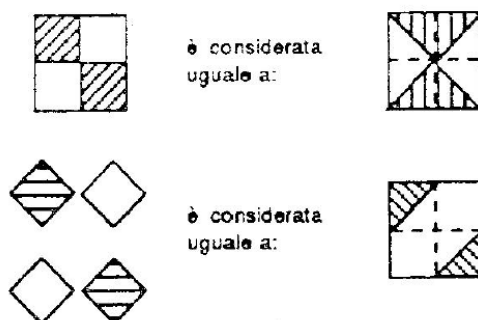
In secondo luogo dovremmo trovare modalità di somministrazione che consentano di interpretare costantemente, per tutte le popolazioni confrontate, i risultati come "**prestazione massima**" o, viceversa, come "**prestazione abituale**". La metodologia delle ricerche interculturali si è evoluta notevolmente nel corso di questo secolo; le tendenze di sviluppo più evidenti sono l'attribuzione di un ruolo sempre più importante alle modalità di somministrazione ed all'analisi dei contenuti degli stimoli e delle qualità dei processi mentali da essi implicati (cfr. Irvine e Carroll, 1980, dove si trovano anche molte indicazioni pratiche tuttora utili). In alternativa, si è pensato anche a misure psicofisiologiche "dirette" dell'intelligenza: in questo campo però la ricerca è ancora in una fase iniziale (cfr. paragrafi precedenti, cap. XIII e XVII), e non è escluso che evidenzii in futuro componenti "culturali" che attualmente nessuno ipotizza.

Di fatto, le ricerche "interculturali" in cui sono stati utilizzati test cognitivi erano spesso finalizzate a scopi pratici: individuare i giovani potenzialmente più adatti ad apprendere le tecnologie o le scienze applicate (p.es.: medicina, ingegneria) che i governanti di determinati paesi "in via di sviluppo" ritenevano prioritarie per accelerare la propria indipendenza dagli "occidentali". In questi casi, l'istanza politica può essere letta come: chi dei nostri ha maggiori capacità di apprendimento nella cultura "loro"? Il fatto di riuscire a risolvere problemi formulati da occidentali nel modo che gli occidentali ritengono esatto può fornire una stima dell'abilità richiesta ed essere quindi una risposta congruente con l'istanza politica.

Il problema è in parte analogo quando si considerano **differenze tra subculture** e **differenze fra strati sociali** all'interno di una stessa nazione. In questi casi si ipotizza in genere una trasformazione della società in tempi lunghi, una progressiva integrazione fra subculture, il contenimento della conflittualità sociale entro limiti che non disturbino lo sviluppo complessivo della nazione. Si ipotizza anche un incremento del benessere economico medio, il che attualmente comporta il ricorso sempre più largo alle tecnologie e quindi maggior richiesta di mano d'opera qualificata e di tecnici che non di mano d'opera non qualificata. Anche in questo caso sembrano funzionali agli obiettivi politici test di "intelligenza" che premiano persone caratterizzate da facilità ad apprendere le tecnologie e a lavorare con precisione e rapidità.

FIG. XX.7

Soluzioni ai cubi di Kohs considerate esatte da popolazioni africane (adattata da Vernon, 1969, p. 102)



I **test di profitto** sono sensibili anche a influssi culturali più sottili, esercitati dal micro-sistema costituito dall'insieme degli insegnanti e degli alunni di una particolare classe scolastica.

Si è già detto nel ≈ 3 che se i contenuti di un test di profitto corrispondono solo in parte ai contenuti che gli alunni

di una data classe hanno avuto l'opportunità di apprendere, i risultati del test sono "distorti" (cfr. trattazione sulla validità di contenuto, cap. IV). Un test può dare risultati distorti anche soltanto perché è stato somministrato molte settimane prima della fine dell'anno scolastico, quando alcune parti del programma non sono state ancora affrontate da alcuni insegnanti.

Per numerosi **test di personalità** è documentata una sostanziale **stabilità del quadro fattoriale** da una cultura all'altra, soprattutto nell'ambito delle "culture occidentali". Questo significa - fra l'altro - che l'insieme delle variabili misurate dal test è osservabile in tutte le culture poste a confronto. La stabilità del quadro fattoriale, in molti casi, è però associata a differenze statisticamente significative tra i valori medi ottenuti, per certe caratteristiche, da popolazioni diverse (cfr. S. Eysenck e H.J. Eysenck, 1969). Il problema delle diversità culturali, evidenziate dal quadro fattoriale o dalle differenze tra medie, si pone in termini diversi a seconda dello scopo per cui viene usato il test.

Se si tratta di **ricerche interculturali**, la stabilità del quadro fattoriale attraverso le variazioni culturali è sufficiente a garantire la validità dei confronti (ovviamente, devono essere state raccolte prove ripetute di stabilità fattoriale su campioni rappresentativi). Le eventuali differenze tra valori medi di campioni diversi vengono interpretate come variazioni di caratteristiche psicologiche "reali", dando per scontato che il significato psicologico di una variabile rimanga costante per tutti i campioni.

Se invece si vuole utilizzare a scopi applicativi, cioè per valutare le caratteristiche di singoli soggetti, un test elaborato all'interno di un'altra cultura, è necessario: (a) verificare la stabilità del quadro fattoriale, come nel caso precedente; (b) verificare la validità del contenuto, cercando di accertare che le caratteristiche misurate dal test abbiano simile connotazione di valore nelle due culture interessate; (c) verificare la validità dei costrutti nel nuovo ambito culturale; (d) elaborare norme statistiche appropriate, calcolandole su un campione della popolazione per cui il test viene adattato, diverso dal campione su cui viene studiata la validità del test.

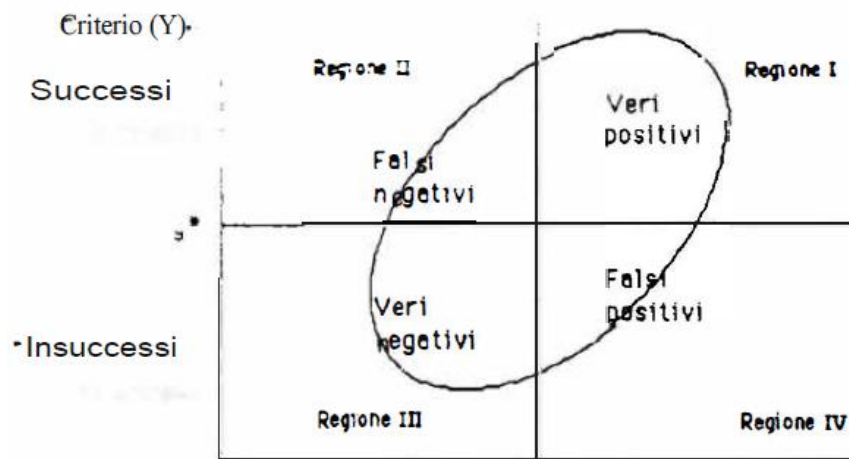
Le differenze di origine culturale o etnica sono di ampiezza diversa a seconda della caratteristica considerata e delle culture confrontate. Ad esempio, per l'EPI (descritto nel cap. XV) sono state trovate differenze statisticamente significative tra bianchi e negri statunitensi nella scala di estroversione (i bianchi hanno medie sistematicamente più elevate in E), ma non nella scala N (cfr. Lowe e Hildman, 1972).

Quando si utilizza uno stesso test per misurare caratteristiche ("cognitive" o "non cognitive") di soggetti appartenenti a popolazioni eterogenee, **l'equità del test può essere analizzata con metodologie diverse**, ciascuna delle quali comporta differenze qualitative nell'interpretazione. In particolare, è stata segnalata l'importanza e la diversità qualitativa dei seguenti approcci metodologici (cfr. Drasgow, 1987; Petersen e Novick, 1976).

(1) **Equivalenza di misurazione**: individui che appartengono a popolazioni diverse e possiedono con pari intensità il tratto misurato dal test dovrebbero ottenere gli stessi punteggi. Non si tiene conto di criteri predittivi (successo o insuccesso nell'attività per cui si viene selezionati). I test che non possiedono questa proprietà danno misure distorte (biased).

FIG. XX.8

Rappresentazione schematica dei principali tipi di "distorsione"
(da Petersen e Novick, 1976)



L'equivalenza di misurazione viene stimata con strategie diverse, le più citate delle quali sono le seguenti (per una discussione più approfondita sotto l'aspetto probabilistico e statistico, cfr. Ironson e Subkoviak, 1979; Petersen e Novick, 1976).

- **Rapporto costante:** chi assume questo criterio dice che il test è equo se la proporzione dei soggetti che superano (o, viceversa, non superano: cfr. Petersen e Novick, 1976) una selezione compiuta con quel test è uguale in tutte le subpopolazioni (Thorndike, 1971a). Alcuni AA. confrontano l'intera distribuzione delle risposte date da due campioni, verificando l'ipotesi nulla mediante il chi quadro (cfr. Scheunemann, 1975, cit. in Ironson e Subkoviak, 1979). L'approccio del "rapporto costante" è criticato perché non distingue tra distorsione della misura e differenze tra popolazioni riguardo alla caratteristica misurata (cfr. Hulin, Drasgow, Parsons, 1983). Gli AA. che fanno questa critica suggeriscono tecniche statistiche più sofisticate, fra cui quelle descritte qui appresso oppure basate su trasformazioni algebriche o trigonometriche dei punteggi, stime probabilistiche, analisi della varianza, tecniche di analisi multivariata e altro ancora (cfr.: Tittle, 1979).

- **Struttura fattoriale uguale in tutti i gruppi** (cfr. Ironson e Subkoviak, 1979).

- **Curve caratteristiche dell'item uguali in tutti i gruppi** (Lord, 1977): viene calcolato, separatamente per le popolazioni da confrontare, il rapporto tra il livello di abilità degli esaminati e la probabilità di dare la risposta esatta (questo rapporto è rappresentato graficamente nella "curva caratteristica dell'item": cfr. cap. I); se, dopo aver "equalizzato" gli items (cioè, dopo aver posto i valori dei parametri nella stessa scala), le curve calcolate su popolazioni diverse non coincidono, l'item viene considerato "distorto"; vengono usati a questo scopo sia i modelli a un parametro sia quelli a tre parametri, con riferimento al modello di Rasch o a quelli di Lord (cfr. cap. I; Lord, 1977; 1980; Hulin, Drasgow, Parsons, 1983; Ironson e Subkoviak, 1979; Petersen, 1980; Scheuneman, 1980, 1982), applicando programmi di elaborazione il più diffuso dei quali è LOGIST.

- **Discriminazione degli items uguale in tutti i gruppi** (cfr. Ironson e Subkoviak, 1979; Ozenne, Van Gelder, Cohen, 1974): gli indici di discriminazione, basati sulla correlazione tra item e totale (cfr. cap. VIII), vengono calcolati separatamente per le popolazioni da confrontare; gli items che sono tra i 50% più discriminanti per una popolazione e tra i 50% meno discriminanti per un'altra vengono considerati "distorti" (cfr. Ironson e Subkoviak, 1979).

I metodi sopra descritti possono dare risultati diversi l'uno dall'altro sotto l'aspetto decisionale. Ad esempio, sulle risposte date a sei test cognitivi da un campione statunitense di oltre 1500 studenti bianchi e uno di oltre 1500 studenti negri di 12th grade (fine scuola secondaria superiore), vennero calcolati quattro diversi tipi di indice di distorsione, basati rispettivamente su: rapporto costante (difficoltà), rapporto costante (chi quadro fra intere distribuzioni), discriminazione, curva caratteristica degli items (Ironson e Subkoviak, 1979). Gli items identificati dai quattro metodi come "distorti" coincidevano solo in piccola parte (il metodo più indipendente dagli altri era quello basato sugli indici di discriminazione). Risultati analoghi sono stati ottenuti da altri ricercatori statunitensi (cfr. Ironson e Subkoviak, 1979).

Quando si leggono studi basati sull'analisi degli items (e non sull'analisi dei punteggi totali delle scale), si tenga presente che la presenza di items "distorti" culturalmente non autorizza a concludere che i punteggi nelle scale siano "distorti": in molti casi, di fatto, le distorsioni a favore dell'una o dell'altra cultura si compensano. Ad esempio, alcuni studi compiuti sui singoli items

del CPI (descritto nel cap. XIV) hanno evidenziato differenze statisticamente significative tra bianchi e negri statunitensi (specie tra donne) in circa il 60% degli items (cfr. Cross e Burger, 1982). Se però si considerano i punteggi nelle scale, le differenze tra bianchi e negri non appaiono altrettanto nette (cfr. Cross, Barclay, Burger, 1978).

(2) **Equivalenza relazionale:** l'intensità della correlazione tra la variabile misurata dal test e un'altra variabile pertinente (p.es.: tra i risultati di un test d'intelligenza e una misura di profitto scolastico) è uguale in subpopolazioni diverse o, in altri termini, la predizione differenziale è stabile nonostante il variare delle popolazioni: p. es., un test d'intelligenza X predice ugualmente bene la riuscita scolastica o professionale di bianchi e di negri statunitensi, di italiani del nord e del sud, di uomini e di donne, ecc.. Anche l'equivalenza relazionale (cfr. Petersen e Novick, 1976) può essere studiata con vari approcci.

- **Analisi della regressione:** in questo caso si dice che "il test è distorto (biased) se il punteggio-criterio predetto dalla linea di regressione comune è sistematicamente troppo alto o troppo basso per i membri di un sottogruppo"; un test che non risponda a tale requisito in misura soddisfacente può esser chiamato "ingiusto" (unfair: cfr. Cleary, 1968).

Se il criterio viene ridotto a una dicotomia, il modello della regressione si semplifica, e può essere analizzato in riferimento a:

- **Probabilità condizionale:** a tutti i candidati che sono potenzialmente capaci di aver successo nell'attività-criterio dovrebbe esser data una pari opportunità di essere selezionati, indipendentemente dal gruppo a cui appartengono (Cole, 1973); il criterio può essere formulato anche con riferimento al polo opposto, cioè all'insuccesso (cfr. Petersen e Novick, 1976).

- **Uguale probabilità:** con riferimento non alla riuscita futura (come nel modello della probabilità condizionale), ma alla situazione dei soggetti nel momento in cui vengono esaminati, dovrebbe esser garantita a tutti i candidati uguale probabilità di essere selezionati, indipendentemente dalla loro appartenenza di gruppo (Linn, 1973); il modello denominato uguale rischio (Einhorn e Bass, 1971) può esser considerato una variante di questo.

Ai fini pratici, prima di interpretare i risultati di un test conviene cercare, nel Manuale o nella letteratura specializzata, se ci sono dati che fanno sospettare una "non equità" del test per soggetti appartenenti a culture diverse. Questi dati vanno letti analiticamente e criticamente, con particolare attenzione alla qualità delle variabili demografiche e socioculturali che, sommandosi, contribuiscono a differenziare le risposte dei soggetti. Ad esempio, nonostante una letteratura abbastanza consistente segnali la sensibilità del MMPI alle differenze etniche fra cittadini statunitensi, uno studio sul MMPI (Bertelson, Marks, May, 1982) in cui adolescenti e adulti bianchi e negri statunitensi erano accuratamente abbinati per sesso, età, luogo di residenza (rurale o urbano), anni di scolarità, stato civile, gravità della patologia (soggetti ospitalizzati o no) e altre variabili, ha concluso che la differenza etnica non contribuiva significativamente a spiegare le differenze tra punteggi nelle scale standard.

In fase di scelta del test questi elementi vanno accuratamente vagliati, per evitare di trovarsi alle prese con test che danno risultati difficili da valutare equamente. In fase interpretativa, se lo psicologo ha elementi di fatto per essere certo che i punteggi nelle scale non discriminano ingiustamente gli appartenenti a culture diverse, ma teme accuse basate su dati a suo parere criticabili, è bene che all'inizio del "rapporto psicologico" esponga gli elementi su cui fonda le proprie convinzioni.

Va anche tenuto presente che le differenze tra culture non sono di per sé generalizzabili: non è detto che le differenze tra statunitensi bianchi e negri corrispondano a differenze tra bianchi e negri residenti in Italia. Se i dati reperibili su un particolare test si riferiscono soltanto a popolazioni non italiane e indicano chiaramente che i risultati ottenuti da appartenenti a culture diverse sono differenti, la conclusione più ragionevole da trarre è che il test potrebbe non essere valido per la popolazione residente in Italia e che quindi non può essere utilizzato a fini applicativi se mancano prove della sua validità per l'Italia. In concreto, se un test dà risultati che ingiustamente penalizzano gli statunitensi di origine ispanica rispetto a quelli di origine anglosassone, probabilmente penalizza ancor di più popolazioni, come quella italiana, che sono non solo geneticamente ma anche geograficamente e culturalmente lontane dagli statunitensi bianchi anglosassoni. Se invece i dati sulla "equità rispetto alle culture" si riferiscono a popolazioni o subpopolazioni residenti in Italia, bisogna: (a) valutare analiticamente l'analogia fra i propri soggetti e le subpopolazioni che potrebbero essere ingiustamente discriminate dal test (p.es.: uno psicologo che risiede a Roma ha spesso occasione di valutare scolaresche in cui si mescolano alunni originari del Centro, del Sud e delle Isole e quindi cerca informazioni sull'equità dei test per queste culture; gli sono poco utili dati sull'equità di un test per soggetti italofoeni e germanofoni dell'Alto Adige); (b) ragionare sul significato e le

implicazioni psicologiche degli indici usati per valutare l'equità del test: p.es., inquadrare diversamente le informazioni sulla possibile discriminazione al momento dell'esame o in vista della riuscita accademica e professionale.

Quel che si è detto in questo paragrafo si riferisce al fatto che tutti i test, in quanto prodotto di persone appartenenti a una cultura, sono "culturali". Resta da vedere quali sono i problemi particolari su cui la "culturalità" dei test può incidere provocando effetti personali o politici indesiderati. Questo passaggio dal generale al particolare comporta anche un'analisi più fine della "culturalità" dei test, una distinzione tra gradi di "culturalità" o, meglio, fra componenti del compito comuni o estranee alle culture confrontate. Dato il carattere applicativo di questo manuale, ci limitiamo a considerare alcuni problemi in ambito nazionale e quindi ristretti al confronto tra subculture.

d) Alcuni problemi applicativi in ambito nazionale

In Italia coesistono culture diverse etnicamente (francofoni, germanofoni, ladini, albanesi...) e/o storicamente (l'unità politica d'Italia risale al 1861/1870). Ci sono inoltre differenze socioculturali che "attraversano" quelle etniche e storico-geografiche. Le diversità non hanno però in genere la consistenza e la drammaticità di quelle studiate negli Stati Uniti e brevemente riassunte all'inizio di questo paragrafo. Si possono comunque distinguere: (1) problemi connessi con l'**appartenenza a un gruppo "allofono" o "alloglotta"**: soggetti cioè che da bambini imparano a parlare una lingua o un dialetto molto diversi dall'italiano standard e che apprendono l'italiano a scuola, di fatto come "seconda lingua"; questa caratteristica è spesso associata a differenze culturali più profonde, soprattutto se i Soggetti appartengono a nuclei di recente immigrazione;

(2) problemi connessi con l'**appartenenza a strati sociali in cui non c'è familiarità con l'espressione scritta** (incluso il disegno);

(3) problemi connessi con un **atteggiamento verso il lavoro diverso da quello ipotizzato dai test cognitivi** ("rendimento massimo" in tempi brevi, responsabilità personale);

(4) problemi connessi con l'**atteggiamento verso categorie sociali come i professionisti che somministrano i test**.

Raramente questi problemi si presentano isolatamente l'uno dall'altro, soprattutto nelle categorie socialmente più "svantaggiate", più estranee alla cultura "standard". E' però utile separarli mentalmente, per trovare le soluzioni più appropriate. Vediamo qualche esempio con riferimento a problemi pratici.

L'utilizzazione dei **test cognitivi in campo scolastico**, nella maggior parte delle situazioni italiane (fatta eccezione cioè per le zone con forti concentrazioni di immigrati extracomunitari), non presenta problemi pari a quelli delle ricerche interculturali: le differenze culturali sono ridotte da una convivenza di lunga data e dall'incidenza sempre più forte dei mezzi di comunicazione di massa e della scolarità d'obbligo, con lingua e programmi miranti all'unificazione. Quando i test cognitivi vengono utilizzati in funzione della *programmazione didattica* o per *individuare soggetti a rischio di problemi di apprendimento* si può ipotizzare una strategia di corretta utilizzazione dei test che tenga conto dei problemi elencati all'inizio.

(1) Se ci sono **problemi culturali e linguistici** e i soggetti parlano una lingua in cui esistono test cognitivi validati, è ritenuta buona norma sottoporre al soggetto un test i cui stimoli e le cui istruzioni siano in "lingua materna" (cfr. cap. XIX). Quando ciò è impossibile (p.es. perché la "madre lingua" è un dialetto), a seconda della competenza dei soggetti in italiano standard si decide se è preferibile usare nelle istruzioni l'italiano (incoraggiando chi non comprende bene a chiedere spiegazioni) o il dialetto. Avrebbe grande interesse pratico disporre di istruzioni standardizzate nei dialetti. In ogni caso, se i soggetti non hanno familiarità con l'italiano standard sarebbe opportuno somministrare un insieme di test tale da rivelare la riuscita differenziale in prove "culturali" e "non culturali" i cui compiti siano simili. I risultati di prove d'intelligenza verbale-scolastica somministrate in italiano a questi soggetti vanno interpretati come misure del profitto ottenuto nell'apprendimento di una seconda lingua. Per quel che riguarda i problemi più generali di differenziazione culturale, soprattutto per gli immigrati, sono state costruite **scale di acculturazione** che possono dare un'indicazione del grado di familiarizzazione degli immigrati con la cultura del Paese di arrivo (si basano su indici quali l'età del Soggetto al suo arrivo nel Paese, il tempo trascorso dal suo arrivo, il suo livello di identificazione etnica, ecc.: cfr. Puente, 1990).

(2) **Se si presume che i soggetti non abbiano familiarità con l'espressione scritta** (p.es. bambini inferiori a 8 anni

provenienti da strati socio-culturali modesti) è consigliabile farli prima esercitare, a lungo, su compiti analoghi: si stima che un esercizio di mezz'ora/un'ora riduca gran parte delle differenze spurie imputabili a questa componente (cfr. cap. XIX). L'esercizio può avvenire il giorno precedente, per non affaticare i soggetti. In fase di interpretazione dei risultati, va riscontrato sul verbale di somministrazione quanto tempo è stato dedicato agli esercizi preliminari (cfr. cap. XIX).

(3) **Un lungo esercizio previo è utile** anche perché tutti apprendano il corretto approccio al compito per quel che riguarda l'importanza di esattezza e rapidità. In ambienti particolari può essere opportuno far precedere la seduta di testing da uno o più incontri informali in cui l'esaminatore abbia chiaramente un ruolo di supporto, non minaccioso. P.es., se i test vengono somministrati nell'ambito di un progetto di orientamento, è opportuno che siano somministrati dalle stesse persone che tengono incontri preliminari informativi. La presenza o assenza di eventuali incontri "di familiarizzazione" va tenuta presente in fase interpretativa.

(4) In alcuni casi, un lungo esercizio preliminare è sufficiente a **far familiarizzare i ragazzi con la figura del somministratore**.

Nonostante queste precauzioni, è probabile che i ragazzi acculturati nella classe "egemone" abbiano un rendimento migliore ai test di abilità generale e di fattore verbale-scolastico, mentre i test di fattore g e di "intelligenza fluida" (Gf) evidenzino meno le differenze rispetto ai ragazzi appartenenti a subculture minoritarie. **Se il risultato dei test deve essere utilizzato per la programmazione didattica o l'orientamento "continuo", "formativo",** di cui gl'insegnanti sono i principali utenti, la soluzione più produttiva sembra il ricorso a un insieme di test (cfr. cap. XIX) che consentano all'insegnante di fare un progetto d'intervento articolato. Accanto a questi test specifici possono anche essere usati test "di livello", che in questo caso dovrebbero includere almeno un test di "intelligenza fluida" (Gf) e uno di "intelligenza cristallizzata" (Gc). Se la misura di Gf risultasse superiore (tenuto conto dell'attendibilità delle misure: cfr. cap. VI) a quella di Gc si potrebbe commentare questo dato dicendo che il ragazzo sembra aver appreso di più dalla vita quotidiana che dall'istituzione scolastica; si può anche cercar di individuare i fattori (motivazionali, emotivi, ecc.) che potrebbero favorire un miglior apprendimento scolastico. Il riferimento puro e semplice al "livello di intelligenza" o di "abilità generale" potrebbe deresponsabilizzare l'insegnante e quindi nuocere al ragazzo, senza peraltro essere più corretto scientificamente.

Nell'**orientamento scolastico episodico** è implicata soprattutto la predizione del successo a breve termine: l'intervento (per lo più compiuto da uno psicologo o da un "consigliere" che ha una preparazione psicologica) consiste nel chiarire ai soggetti qual è il livello di prestazioni che sono in grado di dare rispetto a compiti cognitivi differenziati e nell'aiutarli a utilizzare al meglio le loro risorse, che si suppone siano adeguatamente rispecchiate dal rendimento ai test e che non siano suscettibili di sostanziali modificazioni nell'arco di uno-due anni.

Notoriamente, gli strumenti che offrono informazioni più attendibili nelle previsioni a lungo termine sono i test di abilità generale (descritti nel cap. IX; cfr. cap. XIX), che sono molto correlati allo status socio-culturale e alla padronanza della lingua standard, come peraltro lo è il rendimento scolastico (a parte i problemi specifici delle regioni bilingui). In questo caso i problemi linguistici e sociali vanno considerati come un condizionamento di fatto, di cui conviene stimare l'entità in modo da prevedere eventuali difficoltà al riguardo. E' più produttivo individuare e segnalare le differenze che cercar di ridurle con lunghi esercizi preparatori. Un alto livello di Gf, una forte motivazione e una buona maturità personale generale possono compensare carenze di abilità generale e verbale-scolastica, di solito accompagnate da carenze nelle acquisizioni scolastiche precedenti. La stima del livello generale di abilità è utile, ma va interpretata nel quadro d'insieme e comunicata ai soggetti con le cautele indicate più oltre, nel paragrafo sul "rapporto psicologico".

Anche nelle **assunzioni a posti di lavoro** i quattro problemi segnalati all'inizio di questo paragrafo sono importanti, ma assumono connotazioni diverse a seconda del tipo specifico di lavoro che le persone dovranno svolgere.

(1) La padronanza della lingua italiana standard è rilevante se si tratta di lavori impiegatizi che comportano corretta comprensione e trasmissione di messaggi in lingua materna: in tal caso è accettabile che i test di abilità generale e verbale-scolastica abbiano un peso rilevante, anche se il livello per cui viene fatta l'assunzione è modesto (uscire, dattilografo, centralinista). Se invece il compito è di altra natura (p.es.: autista, tecnico esecutivo, addetto alla manutenzione di macchine) l'utilizzazione di test che privilegiano i membri della cultura dominante

non troverebbe giustificazione.

(2) Considerazioni strettamente analoghe alle precedenti si possono fare per quel che riguarda la familiarità con l'espressione scritta.

(3) Nelle situazioni concorsuali si può dare per certo che ognuno cerchi di rispondere esattamente a quante più domande possibile nel minor tempo possibile. E' anche probabile però che una parte dei concorrenti abbia informazioni prelieve sui test usati per la selezione (libri divulgativi, fuga d'informazioni, partecipazione ad altre selezioni). Notoriamente il retest comporta un incremento nella media dei punteggi e quindi rende iniquo il confronto tra chi affronta il test per la prima volta e chi già lo ha sostenuto. Questo squilibrio può essere almeno in parte compensato dalla presenza di esercizi preliminari lunghi, purché le condizioni ambientali siano tali da non consentire ai più furbi, una volta riconosciuto il test, di omettere gli esercizi preliminari e di passare all'esecuzione della prova prima che venga dato il "via".

(4) Se i candidati appartengono a strati socioeconomici diversi, può avvenire che i soggetti di condizione medio-superiore si ritengano pari o inferiori a chi somministra i test, mentre i soggetti di condizione inferiore alla media si sentono a disagio (il che, ad esempio, comporta maggiori esitazioni nel chiedere chiarimenti, maggiore ansietà, ecc.). In una situazione di selezione, peraltro, il somministratore deve mantenere una certa impersonalità nel rapporto e deve anche far capire, con l'insieme del proprio comportamento, che ci sono regole da rispettare (soprattutto riguardo alle responsabilità individuali): non può quindi creare un clima "troppo" incoraggiante, che fra l'altro rischierebbe di trasformare in sfrontatezza la sicurezza di alcuni. Può soltanto cercar di essere molto chiaro nelle istruzioni, osservare personalmente come i soggetti rispondono agli esempi e dare spiegazioni aggiuntive in pubblico ogni volta che nota carenze o errori nell'interpretazione. Chi interpreta i risultati, in caso di marcata differenza tra punteggi medi di appartenenti a strati socioculturali diversi, dovrebbe informarsi (fra l'altro) anche sulle caratteristiche personali dei somministratori e sul tipo di rapporto stabilito fra loro e i soggetti. Un ultimo esempio riguarda l'uso dei test cognitivi nella **diagnostica clinica**, dove non è accettabile il rischio di considerare come sintomi patologici deficit di prestazione che hanno origine socioculturale, fatto peraltro ripetutamente accertato in riferimento sia a colloqui, sia a questionari (cfr. cap. XIV, in specie per quel che riguarda il MMPI), sia a test proiettivi (cfr. cap. XVIII, con particolare riferimento al Rorschach). In questo caso la padronanza della lingua orale (1) e la familiarità con la lingua scritta (2) vanno stimate con misure generali di fattore *v:ed* o *V* e interpretate in rapporto alle condizioni socioculturali e a misure di fattore *g*, oppure, se è evidente un danno in queste aree, vanno analizzate con test appositi (cfr. cap. XIII). Anche l'assunto che il soggetto possa e voglia rispondere esattamente al maggior numero di quesiti nel minor tempo (punto 3) va verificato esplicitamente e accuratamente, in quanto l'impegno potrebbe variare in rapporto a fattori particolari o da un momento all'altro, come avviene in chi ha disturbi dell'attenzione. Infine, per quel che riguarda l'atteggiamento verso il testista (che in questo caso ha un ruolo assimilato a quello del medico, quindi connotato da molto potere e di per sé altamente ansiogeno) vanno tenute presenti le raccomandazioni date per la somministrazione individuale dei test (cfr. cap. XIX).

8. Interpretare tenendo conto delle interferenze tra variabili cognitive e altre caratteristiche della personalità

I test vengono abbastanza rigidamente suddivisi in "cognitivi" e "non cognitivi" per ragioni prevalentemente storiche e, in parte, metodologiche: i test "cognitivi" cercano di stimolare le prestazioni ottimali, quelli "non cognitivi" cercano di raccogliere le reazioni abituali. Questa distinzione però non va interpretata come mancanza d'interazione, dato che nessuna teoria della personalità considera l'emotività e le dinamiche personali indipendenti dalle caratteristiche cognitive e che, comunque, alcune importanti caratteristiche personali misurate dai test sfuggono a questa dicotomia (si pensi ai test di "stile cognitivo" e ai test oggettivi descritti nel cap. XVII).

L'importanza teorica dell'interazione tra caratteristiche "cognitive" e "non cognitive" è documentata dall'elaborazione di modelli che interpretano le caratteristiche cognitive privilegiando le "propensioni" o "stili cognitivi" in contrapposizione all'esame delle "capacità" (cfr. Baron, 1982) e che, viceversa, interpretano le

caratteristiche "non cognitive" in termini di "capacità" (cfr. introduzione alla p. IV) ed è testimoniata anche dalla chiara affermazione della inscindibilità delle componenti cognitive e affettiva enunciata da Autori contemporanei quali Cattell (cfr. cap. XV), Eysenck (cfr. Eysenck, 1967; Eysenck e White, 1964 e quanto riportato più oltre in questo paragrafo), Neisser (cfr. la definizione degli "schemi" in Neisser, 1976), Piaget (Piaget, 1951, p. 205) e Zazzo (cfr. Zazzo, 1988). Sotto l'aspetto pratico è stato ripetutamente segnalato che il rendimento a molti test "cognitivi" riflette in misura non irrilevante la presenza di caratteristiche personali "non cognitive" (p.es.: test di disegno, cap. IX) e che peraltro la validità delle risposte a test "di personalità" può essere condizionata da caratteristiche cognitive: p.es., soggetti di basso livello verbale-scolastico possono apparire psicopatologici anche se non lo sono. Gli studi fattoriali hanno segnalato la stretta interconnessione tra abilità "cognitive" e "non cognitive" in alcune strutture di notevole interesse applicativo (cfr. Cattell, 1987, p.442ss.): (1) un fattore di **Inibizione Generale** (U.I. 17), la cui componente cognitiva è costituita da lentezza percettiva (p.es. nella strutturazione di Gestalt), lentezza nell'adattamento all'oscurità, ecc., ed è associata alla tendenza a evitare letture disturbanti, a lentezza nel reagire alla minaccia, ecc.; (2) un fattore di **Attivazione corticale** (U.I. 22, denominato "Cortertia" da Cattell), che include abilità percettiva, rapidità nel vedere molti oggetti in disegni non strutturati, rapidità di reazione, accanto a una forte associazione negativa con il nevroticismo; (3) l'**indipendenza dal campo** include abilità cognitivo-percettive nel dissociare figura da sfondo e simili (cfr. cap. X e XVII), nel mantenere operative più istruzioni contemporaneamente, nel seguire le istruzioni, ecc., ma è anche fortemente associata con il fattore di secondo strato **Indipendenza** (QIV di Cattell, componente della U.I. 19) ed è predittivo della distinzione tra nevrotici e normali; (4) l'**Esuberanza** (U.I. 21) ha una componente cognitiva caratterizzata dal fattore di secondo strato **g** (memoria e fluidità generale, connesso quindi anche con il "pensiero divergente") ed appare associata con la componente temperamentale F (*surgency*, oscillazione dell'umore), con basse probabilità di nevroticismo, rapidità nei giudizi sociali, ecc.; (5) la **Forza dell'Io competitiva** (U.I. 16) associa rapidità di percezione e di esecuzione di compiti ad assertività temperamentale; (6) il fattore **Mobilità-Regressione** (U.I. 23), fortemente associato con il **Nevroticismo** definito da Eysenck, al polo negativo ("Regressione") è associato a caratteristiche cognitive di rigidità percettiva, suscettibilità alla suggestione, rapida diminuzione dell'accuratezza con l'aumento della velocità, scarsa memoria immediata, scarsa riuscita nei test d'intelligenza tradizionali.

La distinzione artificiosamente introdotta tra variabili "cognitive" e "non cognitive" al momento della scelta dei test va superata al momento dell'interpretazione dei risultati se si vuole ottenere un'immagine del soggetto realistica, valida. Di fatto, le stesse ragioni epistemologiche che giustificano la separazione fra i due ambiti al momento dell'esame implicano il superamento della distinzione al momento dell'interpretazione dei risultati. Questa operazione di sintesi delle informazioni segue le norme generali della psicodiagnosi e implicano una competenza psicologica generale e clinica. In questa sede vengono solo riferite informazioni specificamente utili a chi deve elaborare informazioni raccolte mediante test.

a) Variazioni di caratteristiche personali in funzione del livello intellettuale e culturale

L'influsso delle caratteristiche cognitive sulle prestazioni ai test "non cognitivi" è il meno studiato tra i problemi delineati in questo paragrafo. In un'ottica applicativa l'argomento è peraltro importante: qui vengono evidenziati i punti più critici della problematica.

E' noto che uno **criteri per decidere sull'opportunità di somministrare questionari di personalità per adulti** è l'adeguatezza del livello cognitivo generale-verbale dei soggetti: i questionari principali sono inadatti a persone di livello scolastico inferiore alla terza media e di livello cognitivo-verbale medio-basso. Può darsi però che i test vengano somministrati "in batteria" e non sia possibile conoscere a priori il livello cognitivo di chi sostiene l'esame. In questo caso, nell'interpretare i risultati dei questionari di personalità si deve controllare anzitutto il livello cognitivo medio del soggetto. Per i soggetti il cui livello medio (specie nei test verbali) è inferiore alla norma, va controllato: (a) il numero dei quesiti omessi e non affrontati: se il soggetto non ha risposto a più del 2-3% dei

quesiti, è ragionevole pensare che abbia avuto difficoltà a comprendere il testo; (b) l'elevazione media del profilo: se i punteggi nelle scale sono quasi tutti medi, può darsi che a molti quesiti il soggetto abbia risposto "non so" o segnato risposte a caso perché non comprendeva il significato della domanda; (c) punteggi indicanti patologie: i soggetti meno sofisticati sotto l'aspetto cognitivo-verbale potrebbero essersi "difesi" meno bene degli altri, e quelli tra loro che provengono da ambienti socioculturali non borghesi potrebbero aver dato un certo numero di risposte che, nel loro ambiente, caratterizzano comportamenti non necessariamente patologici (p.es.: minor controllo delle emozioni e dell'aggressività fisica). In presenza di omissioni numerose, profilo piatto, punteggi patologici, è buona norma riesaminare il soggetto con tecniche adeguate a distinguere le caratteristiche cognitive e l'appartenenza socioculturale dalle caratteristiche "di personalità".

Anche nelle prestazioni ai test proiettivi è stato segnalato l'influsso delle caratteristiche cognitive: in particolare, le risposte al Rorschach e al TAT di soggetti appartenenti ad ambienti socioculturali non borghesi inducono più spesso lo psicologo a considerarli patologici (cfr. cap. XVIII).

Peraltro in **soggetti psicotici** l'analisi dei deficit cognitivi si è rivelata utile nella spiegazione di anomalie comportamentali clinicamente rilevanti (McGee, 1967) e in una prospettiva predittiva. Ad esempio, nella predizione dell'evoluzione a lungo termine (dieci anni di distanza) di bambini autistici si è visto che i soggetti con livello intellettuale relativamente più alto evolvono in modo più favorevole (Rutter, 1965). In un'altra ricerca su bambini con problemi psichiatrici di vario tipo si è constatato anche (com'era prevedibile) che i Soggetti più "competenti" in varie aree di funzionamento cognitivo sono anche percepiti come meno problematici dai loro genitori e insegnanti (Cohen, Kershner, Wehrspann, 1988).

Per quel che riguarda i Soggetti a **rischio di comportamento antisociale** (p.es. ragazzi che vivono in ambienti in cui il crimine è abituale) sembra che i Soggetti con QI più elevato resistano meglio alle stimolazioni antisociali dell'ambiente familiare (Burt, 1925; Kandel et alii, 1988).

Quanto detto finora induce a ritenere fondamentale la misurazione del livello cognitivo con test specifici e distinti da quelli di personalità, anche per l'interpretazione corretta degli stessi test di personalità.

b) Influsso di fattori motivazionali su prestazioni cognitive

I primi studi sul rapporto tra variabili "cognitive" e "non cognitive" ebbero in gran parte origine dall'esigenza di spiegare quella parte della varianza del profitto scolastico che non poteva essere spiegata dai fattori attitudinali (cfr. P.E. Vernon, 1961). Non sorprende che i primi fattori evidenziati come importanti fossero fattori motivazionali (cfr. Harris, 1940).

Attualmente, con il graduale prevalere della psicologia cognitiva e della ricerca di laboratorio, gli studi sul rapporto tra motivazione e cognizione hanno assunto un diverso taglio e prodotto conclusioni meno direttamente utilizzabili dallo psicologo applicativo.

Il ruolo dei **fattori motivazionali** nelle attività cognitive è riconosciuto dai maggiori Autori contemporanei (cfr. Haugen, 1989; si veda anche il numero speciale dedicato alle "Influenze motivazionali e sociali sul rendimento e sulla percezione di sé" dal Journal of Educational Psychology nel 1990), ma viene interpretato in modo diverso. Ad esempio, Neisser ritiene che "le motivazioni non sono forze estranee che danno vita a sistemi che altrimenti sarebbero passivi; sono soltanto sistemi più generali che accettano le informazioni e dirigono l'azione su una scala più ampia" (Neisser, 1976).

Piaget fa riferimento al ruolo di diversi costrutti motivazionali: (a) **affettività** ("L'affettività può essere considerata la forza energetica delle condotte, la cui struttura definisce le funzioni cognitive": Inhelder e Piaget, 1955, pp. 309-10); (b) **valore** ("è l'affettività che attribuisce un valore alle attività e ne regola l'energia": Piaget, 1967, p. 78); (c) **interesse** ("Dal punto di vista fisiologico, l'interesse è un processo causale che può essere descritto mediante la regolazione delle energie: l'interesse facilita l'azione liberando le energie disponibili. Dal punto di vista della coscienza, questo significa che certi valori vengono attaccati a certi oggetti, persone o azioni. Nessuno di questi valori è determinato dalla legge di causa ed effetto, ma da relazioni implicite": Piaget, 1954, p. 144).

Interpretando Piaget, Mischel vede nel concetto di equilibrizzazione la "molla" che spiega l'attività intellettuale: l'attività intellettuale si mette in moto quando una persona si rende conto di qualche conflitto o discrepanza tra i suoi schemi e una situazione nuova (Mischel, 1971, p. 333). In questo modo, di fatto si afferma che l'attività cognitiva viene intrapresa perché il soggetto desidera farlo, pur negando il concetto di "bisogno cognitivo". Da questa posizione dissente Ausubel (1971), che afferma l'esistenza di un bisogno cognitivo connesso con la tendenza all'esplorazione dell'ambiente e alla curiosità e sostiene la possibilità di apprendere anche senza motivazione (p.es.: apprendiamo anche dalla pubblicità ascoltata senza motivazione e con poca attenzione).

D'altro canto, gli studi sulle **motivazioni** sono andati progressivamente privilegiando la componente cognitiva. Ad esempio, nella teoria di McClelland le motivazioni sono apprese (non innate) ed hanno, oltre a una componente affettiva, una componente cognitiva costituita dalle "aspettative" basate su passate esperienze (cfr. McClelland et alii, 1953).

Le analogie fra cognizione e motivazione vengono spinte tanto oltre da alcuni Autori che la distinzione fra i due costrutti appare poco più che nominale: "L'interpretazione degli scopi (goals) come intenzioni e anticipazioni ci porta a vedere che la regolazione motivazionale e quella cognitiva probabilmente non sono processi diversi. Emozioni specifiche, legate alle attività, si sviluppano come risultato del raggiungimento o del mancato raggiungimento di uno scopo" (Hacker, 1985). Altri Autori vedono differenze solo negli effetti delle attività, ma non nelle attività in se stesse: "Il punto cruciale per decidere se un'attività debba essere etichettata cognizione o motivazione è quindi non il tipo di attività in se stessa, ma se l'attività influisca o no sul sistema del sé della persona, con le connesse disposizioni affettive" (Haugen, 1989).

In un'ottica applicativa, questo dibattito apparentemente astratto implica **corollari** notevoli: (a) quando le prestazioni cognitive non sono soddisfacenti, bisogna chiedersi se, nell'immediato, il soggetto era adeguatamente motivato al compito proposto dal test, e se, nel corso della sua vita, ha motivazioni cognitive sufficientemente elevate; per i bambini, è opportuno chiedersi se gli "altri importanti" (soprattutto i genitori e gli insegnanti) stimolano o, al contrario, reprimono, motivazioni a conoscere, tendenze a esplorare e manipolare l'ambiente; (b) nell'esame degli interessi e dei valori (soprattutto in ambito di orientamento) ci si deve sempre chiedere se, come e quanto i soggetti conoscono le attività professionali o accademiche rispetto alle quali misuriamo il loro interesse: un oggetto non può aver valore per una persona se la persona non lo conosce almeno un po' (cfr. discussione in Haugen, 1989). Nella letteratura internazionale si trovano anche proposte per un'interpretazione organica delle strutture cognitive e conative che intervengono nell'apprendimento scolastico (cfr. Snow, 1989).

c) Variazioni di prestazioni cognitive in funzione di caratteristiche personali: personalità, apprendimento e memoria

La relazione tra fattori di personalità e profitto scolastico è stata oggetto di numerosi studi fin dall'inizio del secolo. E' riconosciuto universalmente che il successo scolastico non può essere spiegato solo in base a caratteristiche cognitive. A seconda del livello scolastico, della disciplina e della quantità e qualità delle caratteristiche personali prese in esame l'incidenza dei fattori "non cognitivi" varia, fino a raggiungere il 50% della varianza (cfr. Griffiths, 1974; Rivas, 1977; Secadas, 1952); in qualche caso i fattori "di personalità" spiegano una proporzione di varianza maggiore di quella spiegata dai fattori cognitivi (cfr. Pelechano, 1972; Repetto, 1984). Variazioni ancor più ampie si notano a livello individuale. Soprattutto quando si opera con particolari categorie di soggetti, come i bambini handicappati, "è difficile in pratica tracciare una distinzione netta tra fattori cognitivi e non cognitivi in un dato bambino... Il dispiegamento effettivo delle capacità intellettive dipende non solo dal potenziale genetico e da esperienze ambientali appropriate, ma da forze motivazionali, sia interne al bambino, sia operanti su di lui attraverso l'incoraggiamento e gli atteggiamenti dei genitori" (Mittler, 1974).

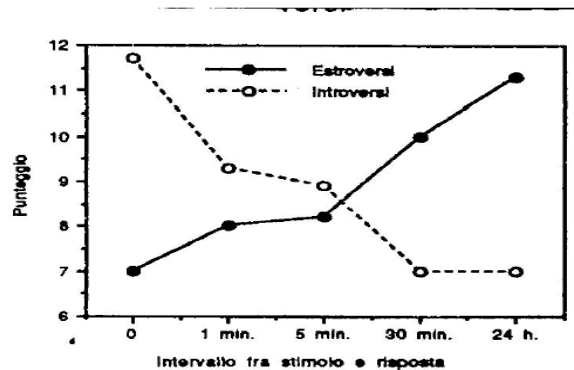
Nell'interpretazione dei risultati "cognitivi" solitamente si prendono in considerazione i fattori "non cognitivi" quando: (a) i risultati scolastici sono oggettivamente scadenti; (b) i risultati scolastici sono sufficienti o superiori alla sufficienza, ma non così buoni come ci si potrebbe aspettare dai risultati dei test di livello cognitivo. Nel corso di questo paragrafo si accenna anche ai risultati di "ricerche di base" sui rapporti fra apprendimento o memoria e fattori di personalità, per l'apporto che possono dare alla comprensione dei problemi applicativi.

Gli studi più antichi sulla dipendenza di processi "cognitivi" da fattori "non cognitivi" riguardano probabilmente i rapporti tra emozioni e oblio e tra apprendimento scolastico e motivazione.

Molti studi sul rapporto tra **oblio** e qualità dei contenuti affettivi memorizzati (cfr. Martins, 1981) risalgono al secolo scorso. All'inizio di questo secolo vennero formulate ipotesi esplicative edoniste di diversa natura: i ricordi spiacevoli sarebbero rimossi nell'inconscio dall'Io, che in tal modo difende la propria integrità (Freud) oppure il richiamo dei ricordi spiacevoli sarebbe estinto perché spiacevole (legge dell'effetto di Thorndike). Si riferiscono pure all'influsso della qualità affettiva dei ricordi ipotesi quali l'"effetto Zeigarnik" (i compiti a cui non si è trovata soluzione si ricorderebbero meglio degli altri) e il "quadro di riferimento" (si memorizzano meglio contenuti riconducibili a quadri di riferimento noti). Altri Autori hanno formulato ipotesi sull'influsso esercitato sulla memoria dall'intensità della valenza affettiva (cfr. Dutta e Kanungo, 1975). Questi studi testimoniano una lunga tradizione di studio sulla problematica di cui si occupa questo paragrafo, anche se sono di remota applicabilità all'interpretazione dei risultati dei test.

FIG. XX.9

Risultati di prove di apprendimento di coppie di parole ottenuti da estroversi e introversi

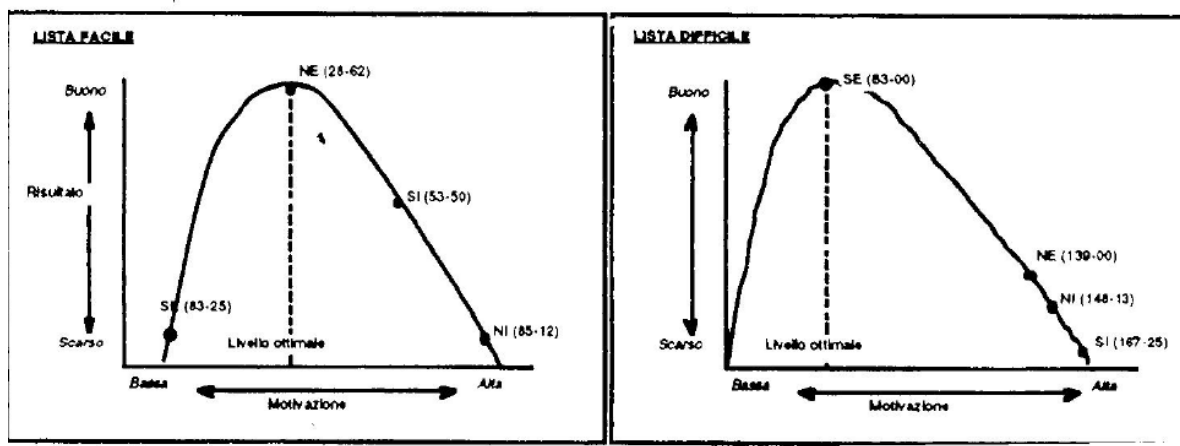


Più recentemente, i **processi cognitivi** posti in atto da soggetti con caratteristiche personali diverse sono stati studiati, nell'ambito della teoria della personalità di Eysenck, mediante ricerche di laboratorio per lo più centrate su compiti di **apprendimento** di coppie di parole o, in genere, su prove di **memoria** (cfr. Martins, 1981). Gli estroversi ottengono risultati migliori degli introversi (cfr. fig. XX.9) quando si chiede di rispondere subito (2 minuti d'intervallo tra stimolo e risposta), gli introversi invece superano gli estroversi quando la risposta viene differita (24 ore di distanza).

Studi più recenti, compiuti usando l'EPQ (descritto nel cap. XV) si riferiscono a una definizione operativa delle variabili parzialmente diversa (p.es.: dall'Estroversione sono stati eliminati gli indicatori connessi con l'impulsività). In analogia con gli studi precedenti, evidenziano che: le risposte a prove di memoria di riconoscimento (parole e disegni) date da introversi sono peggiori in condizioni di minaccia; le risposte di soggetti con punteggio alto in nevroticismo e in psicoticismo sono migliori quando viene promessa una ricompensa; i soggetti con punteggio alto in psicoticismo danno risposte peggiori in condizioni di elevata rumorosità (cfr. Stelmack et alii, 1984).

FIG. XX.10

Risultati di prove di apprendimento di coppie di parole in funzione della difficoltà del compito e dell'intensità della motivazione



Di più immediato interesse per chi deve interpretare i risultati di un insieme di test sono le ricerche sui rapporti tra caratteristiche personali "non cognitive" e rendimento scolastico (che nelle ricerche anglosassoni è sempre misurato mediante test di profitto).

In una rassegna degli studi pubblicati fra il 1930 e il 1937 sui **rapporti tra "personalità" e apprendimento scolastico** (Harris, 1940) la **motivazione** veniva segnalata come il fattore più importante. Tuttora i fattori motivazionali (la "voglia di apprendere") sono ritenuti rilevanti (cfr. Bruner et alii, 1966); a differenza di quel che avveniva nelle prime ricerche vengono però studiati anche in rapporto a fattori sociali e ipotizzando il ruolo di mediazione dei genitori (cfr. Marjoribanks, 1987).

Successivamente, l'attenzione si spostò sul ruolo dell'"**adattamento**" (Gough, 1953; Gough, 1959c; Secadas, 1952; 1962) e di variabili "disturbanti", quali l'ansia o il nevroticismo. Per i rapporti, complessi, fra **ansia** e apprendimento, si veda quanto è stato detto sull'ansia da esame nel cap. XIX (cfr. anche una rassegna di studi di psicologia generale in: Martins, 1981). Nell'interpretare i risultati, va tenuto presente che gli appartenenti a categorie "socialmente deboli" (p.es., nelle ricerche statunitensi, donne e negri) sono in alcuni casi più vulnerabili dall'ansia d'esame (cfr. Payne, 1984).

Rapporti significativi tra apprendimento e variabili "non cognitive" sono stati evidenziati anche in riferimento ad altre caratteristiche. Ad esempio, studi su quattordicenni con problemi comportamentali hanno evidenziato che i ragazzi più "chiusi" reagivano meno bene all'insegnamento rispetto ai loro compagni con sindrome ostile-aggressiva (Marshall, 1967). Altre ricerche prendono in considerazione variabili quali il concetto di sé, lo stile cognitivo, la risonanza, la riflessione (cfr. Repetto, 1984).

L'insieme delle ricerche segnala anzitutto la rilevanza dei fattori "non cognitivi" in genere rispetto al rendimento scolastico.

In uno studio statunitense su adolescenti (Cattell e Butcher, 1968) il peso relativo di 16 caratteristiche personali rispetto a un criterio di profitto scolastico venne misurato mediante analisi della regressione. La variabile di maggior peso (0,50) risultò, com'era prevedibile, l'intelligenza cristallizzata (fattore B), seguita dalla "forza del Super Io" (G: 0,25), dall'indipendenza personale (Q2: 0,20), dalla buona integrazione del concetto di sé (Q3: 0,20) e da altri fattori con pesi inferiori (fra 0,10 e 0,15). Una panoramica di 25 studi compiuti nell'ambito della teoria di Cattell (Warburton, 1968, sintetizzato in Entwistle, 1972) segnala una connessione significativa tra fattori di personalità e profitto fino alla fascia 15-17 anni. A questa età i risultati degli studi cominciano a discordare tra loro, fatto che, di per se stesso, denuncia il diminuire del loro peso.

In una ricerca su un campione di studenti universitari italiani (Boncori, 1985b; 1986), sottoponendo ad analisi delle componenti principali 72 misure di caratteristiche "cognitive" e "non cognitive" della personalità, si è trovato che il primo fattore è bipolare, caratterizzato al polo positivo da buoni risultati nei test d'intelligenza, da interessi teoretici e scientifici, realistici e tecnici, da tratti di personalità quali Forza dell'Io, Dominanza, Gusto per il cambiamento e al polo negativo da un insieme di indici psicopatologici (soprattutto attinenti a nevroticismo e ansia), da interessi convenzionali, da preferenza per uno status dipendente. L'importanza dei fattori "non cognitivi" è segnalata dall'associazione di variabili "cognitive" e "non cognitive" all'interno del primo fattore, quello che spiega la maggior proporzione della varianza (12%).

Nel proseguimento della ricerca (Boncori, 1987), si è trovato che le differenze statisticamente significative tra un campione di maturandi e un campione di studenti universitari del terzo anno riguardavano più spesso aspetti non cognitivi che cognitivi: gli universitari del terzo anno ottengono in media punteggi più alti dei maturandi ai test d'intelligenza, punteggi più bassi alle scale psicopatologiche e punteggi più alti in "radicalismo". Differenze "cognitive" e "non cognitive" sono ugualmente significative anche nel caratterizzare campioni di studenti che seguono corsi di studi qualitativamente diversi (p.es.: arti e scienze; cfr. Saville e Blinkhorn, 1976).

Numerose ricerche (per lo più afferenti ai quadri teorici di Cattell e di Eysenck) evidenziano anche alcune interessanti qualità del rapporto tra fattori "cognitivi" e "non cognitivi", che per praticità vengono qui appresso sintetizzate con riferimento ai fattori di secondo ordine ansietà-nevroticismo ed estroversione-introversione (Cattell offre anche equazioni per massimizzare la predittività del profitto utilizzando i dati del 16 PF e dell'HSPQ: cfr. Cattell, 1987, p. 471 ss).

I rapporti tra **estroversione-introversione e risultati ai test di profitto** variano a seconda del livello scolastico considerato:

- gli estroversi (nell'accezione di Eysenck) tendono a ottenere risultati scolastici migliori nella scuola elementare (correlazioni intorno a $r = 0,22$ in un campione di 4000 alunni al termine della scuola elementare: Eysenck e Cookson, 1969);
- gli introversi (nell'accezione di Eysenck) ottengono risultati scolastici migliori ai livelli superiori, in specie all'Università (Entwistle, 1972; Eysenck, 1972; Eysenck e Cookson, 1969; Lavin, 1967; Savage, 1962).

Anche nell'ambito della teoria di Cattell (v. panoramica di studi in Entwistle, 1972) le variabili attinenti al fattore di secondo ordine Estroversione/Introversione hanno lo stesso tipo di correlazione con il profitto. I fattori di primo ordine attinenti all'introversione hanno correlazioni sistematicamente negative con il profitto fino ai 17 anni; dopo questa età prevale (sei studi su dieci) la tendenza opposta. I 18 studi riguardanti la coscienziosità (G+) hanno lo stesso andamento d'insieme. Altre variabili invece hanno una stabile correlazione positiva con il profitto, ma solo a partire da una certa età: la "tenerezza" (I+, N-) dai 12 anni in su, il "radicalismo" (M+, Q1+) dai 17 anni in su.

La spiegazione del fenomeno potrebbe essere connessa con il diverso tipo di richiesta posto dalla scuola primaria (dove la socializzazione è un importante obiettivo educativo e il clima di lavoro è molto informale) rispetto ai livelli d'istruzione successivi; potrebbe però anche essere connessa con un diverso ritmo di sviluppo cognitivo negli introversi e negli estroversi (Eysenck e Cookson, 1969).

Il rapporto con il profitto è meno chiaro per quel che riguarda il **nevroticismo** (cfr. Entwistle, 1972; Eysenck e Cookson, 1969): probabilmente, data la connessione fra questa variabile e l'ansietà, le ricerche dovrebbero essere condotte in modo da distinguere tra ansia facilitante e debilitante (cfr., nel cap. XIX, il paragrafo sull'ansia da esame).

Studi condotti nell'ambito della teoria di Cattell evidenziano che i fattori primari connessi con il fattore di ordine superiore **ansietà-nevroticismo** (C-, D+, L+, O+, Q3-, Q<+) sono correlati negativamente con il profitto fino ai 14 anni; per la fascia 15-17 anni, sette su nove degli studi recensiti evidenziano una correlazione positiva e per studenti sopra i 17 anni 11 studi su 19 evidenziano correlazioni positive (cfr. anche altri dati in Cattell, 1987). Questi dati vanno nella stessa linea di alcuni risultati riguardanti il rapporto tra Nevroticismo e profitto nella teoria di Eysenck.

Nel loro insieme, le misure della personalità e della motivazione spiegano una proporzione di varianza del successo scolastico che secondo alcuni AA. è pari o superiore a quella spiegata dalle misure di abilità. Secondo Cattell e Butcher (1968) la varianza spiegata dalle misure di abilità fluida è il 20%; i 13 fattori "non cognitivi" dell'HSPQ ne spiegherebbero il 18% e i 15 fattori psicodinamici dello SMAT il 16%. Usando congiuntamente predittori "cognitivi" e "non cognitivi" si spiegherebbe il 60% o più della varianza del criterio.

Le connessioni fra riuscita scolastica e caratteristiche della personalità pongono però problemi di ricerca complessi. I risultati appaiono molto differenziati a seconda delle discipline e del curriculum; è anche evidente che il quadro delle correlazioni tra caratteristiche personali e profitto varia a seconda del criterio di profitto usato (valutazioni globali dell'insegnanti, saggi, test di profitto...), dell'età e del livello di abilità generale dei soggetti, del sesso, dell'area geografica, dell'organizzazione scolastica, dei metodi d'insegnamento e della personalità dell'insegnante (cfr. Entwistle, 1972), delle caratteristiche socioculturali dell'ambiente. L'insieme delle variazioni connesse con differenze curricolari e culturali e con i metodi di misurazione del profitto consiglia estrema cautela nell'applicare alla situazione italiana informazioni specifiche desunte da studi stranieri. In presenza di deficit di profitto inspiegabili in base al quadro attitudinale è ragionevole però cercare fra le caratteristiche motivazionali o strutturali della personalità un'ipotesi esplicativa, tenendo conto anche di quanto risulta dalle ricerche svolte su popolazioni diverse.

d) Livello cognitivo, creatività e caratteristiche personali

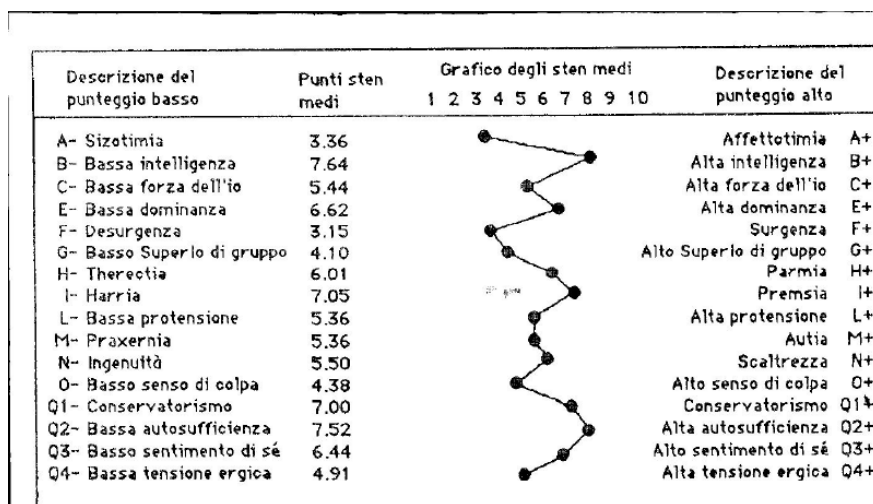
Numerosi test possono dare informazioni, al tempo stesso, su caratteristiche "cognitive" e "non cognitive" di una persona (cfr. P.E. Vernon, 1953): p.es. il Rorschach e il TAT (cfr. cap. XVIII) valutano principalmente le strutture e

le dinamiche emotive, ma consentono anche di stimare il livello e la qualità del funzionamento intellettuale (cfr. Koppitz, 1982, per l'uso dei test proiettivi come test informali di abilità), le scale Wechsler misurano il livello intellettuale ma sono state usate anche per valutare strutture e dinamiche emotive della personalità (cfr. Schafer, 1971; Rapaport, Gill, Schafer, 1975 e, per un approccio quantitativo, Henrichs, Klauskopf, Amolsch, 1982). E' noto però che quando si misurano molte variabili utilizzando un solo strumento si ottengono spesso convergenze artificiali, dovute all'interdipendenza tra stimoli e allo scarso numero delle risposte disponibili per misurare ciascuna delle caratteristiche a cui si è interessati.

Le ricerche più affidabili, condotte con strumenti indipendenti, riguardano soprattutto soggetti dalle caratteristiche eccezionali: molto o poco intelligenti, portatori di psicopatologie, ecc. Qui appresso vengono segnalati solo alcuni studi, senza pretesa di completezza.

FIG. XX.11

Profilo medio al 16 PF di eminenti ricercatori (N = 144) in fisica, biologia o psicologia (da Cattell e Butcher, 1968, p. 290)

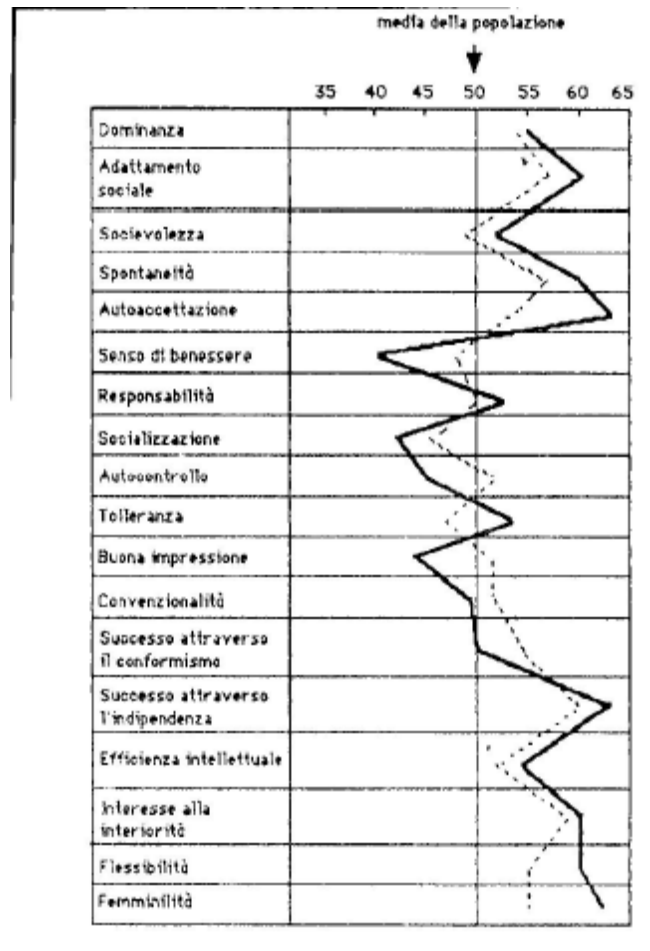


Fin dall'inizio di questo secolo numerose ricerche hanno lavorato su ipotesi riguardanti la connessione tra **adattamento personale e livello intellettuale elevato**. Le conclusioni di questi studi non sono univoche. Per citare soltanto un paio di autori famosi: la Hollingworth (1942) sostiene che i ragazzi "superdotati" tendono ad avere problemi di socializzazione e di adattamento personale, mentre gli studi longitudinali di Terman evidenziano che ragazzi molto intelligenti (QI > 140 alla Stanford-Binet) sono meglio adattati della media (Terman, 1925-1959). Le diverse conclusioni sono probabilmente da attribuire a differenze metodologiche: p.es., la diversa definizione operativa di "superdotato" e di "adattamento" e, soprattutto, alla presenza o assenza del campione di controllo e alla sua caratterizzazione. E' stata anche suggerita la possibilità che le correlazioni tra intelligenza e personalità siano curvilinee (Cattell, 1987, p. 455).

Una rianalisi dei dati della famosa ricerca di Terman con il metodo della *path-analysis* (Tomlison-Keasy e Little, 1990) ha evidenziato netti rapporti di causalità rispetto a quattro distinti criteri di riuscita: (1) **Risultati scolastici** (i migliori predittori sono: Responsabilità sociale, Socievolezza, Istruzione dei genitori); (2) **Capacità intellettuale** (miglior predittore: Determinazione intellettuale); (3) **Adattamento personale** (miglior predittore: Armonia familiare); (4) **Risultati occupazionali** (i migliori predittori sono Risultati scolastici e Capacità intellettuale).

Studi recenti evidenziano che, tenendo sotto controllo variabili quali l'attendibilità di chi risponde ai questionari e la coesione familiare, i dati sono piuttosto a favore dell'ipotesi di una correlazione positiva tra adattamento e intelligenza in bambini con QI > 120 (secondo ciclo della scuola elementare: Grossberg e Cornell, 1988). La tipologia dei disturbi dei ragazzi "superdotati", secondo i dati di cliniche specializzate, è varia e non sembra caratterizzata in modo peculiare (cfr. Prat, 1979). Nel caso si trovino segni di disadattamento in ragazzi "superdotati", sembra quindi ragionevole cercare le cause del disadattamento in fattori diversi dall'elevato livello intellettuale. Clinici specializzati nella terapia di ragazzi intellettualmente dotati segnalano

Profilo medio di scrittori "creativi" e "di confronto", rilevato con il CPI



Studi condotti con test proiettivi rivelano dati qualitativamente più articolati. Ad esempio, una ricerca condotta su alunni italiani (Andreani e Orio, 1972) evidenzia che i soggetti con alta creatività sono caratterizzati da maggiore ansia e aggressività specialmente al Rorschach, cioè a livello più profondo, pur avendo un buon adattamento sociale (rivelato, p.es., da risposte costruttive al PFS di Rosenzweig).

Gli studi fin qui riferiti sono stati condotti con una pluralità di strumenti e/o con metodologie complesse. Quando la metodologia è molto semplificata, le conclusioni sono più difficili da interpretare. Ad esempio, in uno studio su un campione di detenuti per omicidio volontario i soggetti che ricevevano i punteggi più bassi ai test d'intelligenza (sia verbale sia non verbale) tendevano ad avere profili MMPI più patologici (cfr. Holcomb e Adams, 1983). Si può pensare sia che c'è una correlazione negativa tra intelligenza e psicopatologia, sia che i più intelligenti sono anche più abili nel dissimulare eventuali patologie. La stessa tendenza si nota peraltro anche nelle risposte ai test proiettivi (cfr. cap. XVIII: in specie riguardo al Rorschach), che dovrebbero rendere la dissimulazione assai più difficile.

In un'altra ricerca su campioni di **anziani** è stato trovato che misure di efficienza intellettuale correlano significativamente (anche se debolmente: intorno a $r = -0,20$) con misure di rigidità-ossessività (cfr. Clément, 1980). Sembra quindi ragionevole tener conto anche del livello intellettuale prima di segnalare come patologici eventuali indicatori di rigidità-ossessività negli anziani. Si può concludere sia che rigidità e ossessività intralciano le prestazioni cognitive, sia che le persone meno intelligenti diventano più facilmente ossessive e rigide, sia che il deterioramento dovuto all'età danneggia in pari misura caratteristiche cognitive e non cognitive.

Altre ricerche evidenziano connessioni più specifiche, d'interesse ristretto o da approfondire ulteriormente.

Ad esempio, uno studio longitudinale su un campione di studenti californiani (Ozer, 1987) ha evidenziato correlazioni statisticamente significative (coefficienti per lo più fra $r = 0,30$ e $r = 0,40$) tra le prestazioni a un test di *abilità spaziale* (rotazione mentale) e numerose caratteristiche personali, diverse però a seconda dell'età e del sesso dei soggetti. Ad esempio, a 3 anni l'abilità spaziale correla soprattutto con "attenzione e capacità di concentrarsi" ($r = 0,48$), "tendenza a modificare i limiti" ($r = -0,40$), "ricerca di contatto fisico con gli altri" ($r = -0,38$); da 7 a 14 anni con "sembra avere un'elevata capacità intellettuale"

(mediana: $r = 0,38$), "pone a se stesso elevati standard di rendimento" (mediana: $r = 0,33$); a 18 anni le ragazze presentano i coefficienti di correlazione più elevati con variabili quali "è sottilmente negativista" ($r = -0,43$), "è suscettibile alle critiche" ($r = -0,42$); i ragazzi con le variabili: "è imbroglione, manipolativo, opportunist" ($r = -0,38$), "estrapunitivo: tende a trasferire o proiettare le colpe" ($r = -0,38$); unica variabile correlata significativamente con il rendimento al test spaziale per maschi e femmine è "sensibilità alle richieste" ($r = -0,40$ per le femmine, $r = -0,31$ per i maschi). Probabilmente, come ipotizza l'autore, il costrutto che media fra l'abilità spaziale e l'insieme delle caratteristiche comportamentali evidenziate dal test di personalità usato (California Q-set) è l'indipendenza dal campo: l'ipotesi però è ancora da sottoporre a verifica.

Attualmente mancano ricerche che consentano di dire, soprattutto alla psicologia italiana, se la correlazione tra punteggi nei test d'intelligenza e nelle scale psicopatologiche risponde a una connessione reale tra questi due ordini di variabili oppure è un artefatto metodologico (p.es.: chi è più intelligente si difende più efficacemente dal rischio di ricevere un'etichetta psicopatologica). Data la delicatezza del problema, nell'interpretare i risultati di soggetti che hanno bassi punteggi ai test d'intelligenza generale si dovrà usare più cautela del solito prima di attribuire una psicopatologia. In particolare, sarebbe opportuno integrare i dati dei test con l'osservazione sistematica del comportamento del soggetto.

e) Variazioni di prestazioni cognitive in funzione di caratteristiche personali: rapidità di reazione e intelligenza

L'importanza e la complessità dei rapporti fra rapidità e intelligenza sono state già segnalate riguardo sia alla scelta sia alla somministrazione dei test. Qui l'argomento viene ripreso in funzione dell'interpretazione dei risultati.

Sir Francis Galton riteneva che l'intelligenza fosse basata su processi biologici semplici e considerava buoni indicatori dell'intelligenza i tempi di reazione misurati in riferimento a semplici compiti di discriminazione sensoriale o di percezione. L'ipotesi galtoniana sull'importanza della correlazione tra rapidità di reazione e intelligenza ha continuato a generare ipotesi di ricerca e conferme parziali (cfr. Larson, Merritt, Williams, 1988 e bibliografia ivi citata) nonostante alcune clamorose smentite empiriche (p.es. Wissler, 1901; una breve sintesi di più recenti polemiche si trova in: Lindley, Smith, Thomas, 1988). Nella clinica applicativa il problema della rapidità è stato di fatto ignorato, adottando la prassi di lasciar lavorare ciascun soggetto al ritmo preferito. Nella maggior parte degli altri campi applicativi (lavoro, orientamento, scuola, ricerche fattoriali sulla struttura dell'intelligenza) la tradizione binetiana, che privilegiava la stimolazione dei processi mentali "superiori" e quindi la complessità del compito, è stata abbinata alla tradizione galtoniana della misura della rapidità di reazione senza riflessioni teoriche preliminari. Questo "ibrido" ha prodotto una complessità nei risultati che solo di recente è stata denunciata come problema interpretativo: "qualsiasi punteggio ottenuto in test a tempo limitato con test somministrati nel modo convenzionale, è probabilmente funzione di componenti separate: rapidità e livello di padronanza" (Carroll, 1983).

Nell'ambito dei test di ragionamento spaziale la ricerca è andata anche oltre, segnalando distinzioni tra fattori imputabili al diverso ruolo della rapidità di soluzione (Egan, 1978, 1981): i test di Visualizzazione sono poco sensibili alla rapidità, mentre i test di Relazioni spaziali e di Orientamento spaziale lo sono molto (Lohman, 1979). La rapidità, inoltre, influisce di più quando gli items sono in forma di "scelta multipla" (cfr. Carroll, 1983).

Come si è detto nel paragrafo sulla scelta dei test, alcuni recenti modelli dell'intelligenza e del *problem solving* (Cattell, Eysenck, Furneaux: cfr. cap. X) assegnano un ruolo importante alla rapidità.

La correlazione tra rapidità e intelligenza peraltro non è elevata. I coefficienti di correlazione tra misure di rapidità di reazione e misure d'intelligenza raramente superano la "soglia" di $r = 0,3$ (cfr. Baxter, 1941; Hunt, 1980), il che significa che la rapidità di reazione spiega, all'incirca, non più del 10% della varianza dei risultati ai test d'intelligenza. Va tenuto presente però che la maggior parte degli studi di laboratorio coinvolgono piccoli campioni, che proprio per le loro ridotte dimensioni sono spesso molto omogenei e quindi non possono dare origine a coefficienti di correlazione elevati.

L'intensità della relazione fra rapidità e risultati varia a seconda degli strumenti usati per misurare la rapidità e l'intelligenza.

Ad esempio, se si somministra un test con tempo ridotto e con istruzioni che incoraggiano la rapidità di esecuzione, la rapidità può determinare anche il 60% della varianza dei risultati (Baxter, 1941: test Otis Self, livello superiore, somministrato con tempo abbreviato). Ancora maggiori sono le differenze tra misure di rapidità, dato che queste sono molto differenziate e poco correlate fra loro (cfr. Buckhalt e Jensen, 1989: la mediana dei coefficienti di correlazione fra una dozzina di misure di rapidità è $r = 0,06$). Se la rapidità viene stimata misurando i tempi di reazione in compiti complessi o in riferimento ad attività molto simili a quelle stimulate dal test d'intelligenza che fa da criterio, i coefficienti di correlazione tra rapidità e punteggio al test superano i valori abituali. Ad esempio un coefficiente $r = 0,50$ è stato ottenuto correlando i risultati delle PM 47 "avanzate" con i risultati di una prova complessa, Counters ("Contatori"), in cui il soggetto deve controllare simultaneamente sullo schermo di un calcolatore tre contatori in cui i numeri cambiano rapidamente e in ordine casuale; quando appare un quadratino, il Soggetto deve aggiungere "1" al contatore il cui numero si trova sotto il quadratino (Larson e Saccuzzo, 1989). Ci si chiede se i compiti di ispezione e di apprendimento implicati da questa prova di rapidità non siano troppo simili ai compiti implicati dalla soluzione delle Matrici Avanzate. Alcuni AA. interpretano i risultati di ricerca finora disponibili come un insieme unitario, ritenendo di secondaria importanza i fattori specifici connessi con le peculiarità di compiti impiegati nei diversi esperimenti. Concludono che le correlazioni fra rapidità di elaborazione mentale e risultati di test d'intelligenza sono prove a sostegno della teoria del fattore g e propongono modelli interpretativi neurofisiologici (cfr. Haier et alii, 1988; Larson e Saccuzzo, 1989). Altri AA. invece cercano di chiarire i rapporti tra rapidità di reazione e intelligenza sia tenendo conto di fattori interagenti (p.es.: estroversione, impulsività), sia analizzando i processi e le componenti della rapidità. In quest'ultima prospettiva, ad esempio, sono stati messi a punto strumenti elettronici per la registrazione automatica dei tempi (Roth, 1964; Jensen, 1978) che consentono di distinguere fra differenze nel tempo reale di elaborazione e nel tempo richiesto per la risposta motoria. Le differenze individuali nel tempo reale di elaborazione (esclusi cioè i tempi di reazione motoria) sono significativamente correlate con misure convenzionali d'intelligenza: p.es. è stata trovata una correlazione $r = -0,44$ con il subtest di Vocabolario della WISC-R (Hemmelgarn e Kehle, 1984; cfr. anche Jensen, 1987a).

E' stato anche osservato che "in sé, l'elemento critico non è la rapidità, ma piuttosto la selezione della rapidità: sapere quando funzionare rapidamente o lentamente ed esserne capaci, a seconda del compito o delle esigenze situazionali" (Sternberg, 1982). L'osservazione di Sternberg appare sostenuta dalla ripetuta osservazione di correlazioni significative tra variabilità intraindividuale nella rapidità e risultati ai test d'intelligenza (cfr. P.A. Vernon, 1986). Questa abilità nell'"impiego delle risorse" non viene peraltro valutata esplicitamente da nessun test. La qualità dell'esame cognitivo migliorerebbe se la rapidità venisse misurata differenzialmente, in rapporto a componenti diverse del processo intellettuale, come propone Sternberg (1982, p. 304) o, in alternativa, se si ottenesse una stima almeno sommaria della capacità dei soggetti di variare il proprio ritmo di lavoro in funzione delle esigenze.

Questa stima si può ottenere anche con strumenti già esistenti. Ad esempio, se sono stati somministrati test nelle cui istruzioni è esplicitamente detto che viene premiata la rapidità (p.es. Otis) insieme a test nelle cui istruzioni è esplicitamente detto che viene premiata l'accuratezza o che di fatto la esigono (p.es. test di ragionamento aritmetico in cui la lettura accurata del problema è fondamentale per la corretta soluzione), confrontando i risultati ottenuti dal soggetto nei due tipi di prove (a parità di contenuto) si può "stimare" la capacità del soggetto di adattare la rapidità a esigenze diverse. Resta comunque ignota la sua capacità di distinguere tra circostanze in cui è preferibile la rapidità alla ponderazione.

Le considerazioni appena esposte riguardo alla qualità di processi posti in atto trovano elementi a sostegno nei risultati delle ricerche sui rapporti tra intelligenza, rapidità e "personalità". Ad esempio, la dimensione **impulsività - riflessività** (Kagan, 1965) appare in alcuni studi significativamente correlata sia con la rapidità di reazione sia con i risultati dei test d'intelligenza (cfr. P.A. Vernon, 1986), ma in altri studi è la quantità di tempo impiegato nel "comprendere" e "codificare" un problema ad avere buona correlazione ($r = 0,62$) con i risultati ottenuti (cfr. Sternberg e Davidson, 1982).

Nell'insieme, i rapporti tra misure di rapidità di reazione e d'intelligenza variano in funzione di molti fattori, i più studiati dei quali finora sono:

- **Complessità del compito** (talvolta definita in riferimento alla qualità e quantità dei processi mentali implicati, talvolta definita operativamente in riferimento al tempo "di latenza", cioè alla quantità di tempo che passa prima che i soggetti rispondano; un esempio di esperimento al riguardo è in: Larson, Merritt, Williams, 1988; cfr. anche Tate, 1948); nelle ricerche compiute "sul campo", in cui la rapidità viene misurata contando quanti sono i quesiti affrontati, i fattori di "fluidità" vengono evidenziati prevalentemente quando i compiti sono semplici. In particolare, nelle ricerche dirette da Guilford (cfr. cap. X e XI) sono stati trovati fattori di *fluidità* di primo ordine, distinti fra loro per contenuto e per tipo di compito (p.es.: produzione rapida di molte parole che cominciano con la stessa lettera, calcolo aritmetico rapido, rapidità di percezione, aggiunta di molti dettagli decorativi a un disegno, ecc.) e di secondo ordine, distinti per tipo di contenuto (fluidità verbale, fluidità ideativa, ecc.). Altri

Autori hanno trovato un fattore generale di rapidità e fattori specifici di rapidità in varie attività (cfr. Holzinger, 1934; Lord, 1956; P.E. Vernon, 1953, p. 69 ss.).

- **Variabilità intraindividuale della rapidità:** da alcune ricerche sembra che i soggetti più abili siano più stabili della rapidità di risposta.

- **Appartenenza socioculturale:** da alcune ricerche sembra che gli effetti della rapidità contribuiscano a compromettere l'equivalenza delle misure nei confronti tra culture (p.es., tra studenti universitari tedeschi e statunitensi: Ellis, 1990; tra bambini cinesi e britannici: Lynn, Chan, Eysenck, 1991; Flynn, 1991).

- **Caratteristiche personali** quali l'estroversione (Schwegler, 1929; Notcutt, 1943; Brierley, 1961), la tipologia somatopsichica (i picnici sarebbero più rapidi degli astenici secondo Kretschmer: cfr. P.E. Vernon, 1953) e caratteriologica (i primari sarebbero più rapidi dei secondari: cfr. Calonghi, 1961).

- Alcune **psicopatologie** differenziano i soggetti congiuntamente nelle dimensioni rapidità e accuratezza: p.es. si è visto che le prestazioni di nevrotici introversi sono caratterizzate da bassa velocità e quelle di nevrotici estroversi da bassa accuratezza (Brierley, 1961), che gli isterici tendono a essere più rapidi mentre i distimici sono più accurati, che i pazienti maniaci sono più rapidi e meno accurati dei melanconici (cfr. P.E. Vernon, 1953).

I problemi interpretativi pratici si riferiscono prevalentemente ai risultati di test somministrati con limiti di tempo più o meno ristretti. Le non molte ricerche riferite a questo problema specifico evidenziano che:

(a) i punteggi dei test cognitivi somministrati con limiti di tempo sono misure fattorialmente complesse, in cui vanno distinte almeno due variabili: (1) il livello cognitivo o di competenza e (2) la velocità a cui il soggetto lavora (Davidson e Carroll, 1945);

(b) il tempo impiegato dai soggetti per affrontare tutti gli items del test una sola volta e il punteggio corrispondente al numero degli items risolti esattamente quando il test viene somministrato senza limiti di tempo sono indipendenti l'uno dall'altro (cfr. Baxter, 1941; Davidson e Carroll, 1945). Questa affermazione non è in contrasto con quanto affermato sulla correlazione positiva ma bassa fra rapidità di reazione e risultati ai test.

Nelle condizioni in cui gli psicologi operano abitualmente, non è possibile misurare

la rapidità dei soggetti indipendentemente dalla loro capacità di risolvere problemi. E' possibile però, **in pratica:**

(a) **confrontare la diversità di rendimento in prove più o meno sensibili alla rapidità** - inclusi i risultati di eventuali test di fluidità somministrati - ed, eventualmente, approfondire l'esame prendendo in considerazione il rapporto tra risposte giuste, errate e omesse (p.es.: si è visto che il numero delle omissioni varia a seconda della "primarietà" o "secondarietà" dei soggetti: Calonghi, 1961);

(b) prendere in considerazione i **fattori di personalità** che potrebbero essere connessi con la rapidità e con il rapporto rapidità - accuratezza (p.es. estroversione-introversione, impulsività-riflessività, ecc.);

(c) nel caso l'insieme di risultati sia incoerente (p. es.: un soggetto depresso che riesce meglio nei test "rapidi") vanno **ipotizzati errori nella somministrazione del test** (p.es.: sono stati concessi tempi non standard, qualche soggetto è riuscito a cominciare prima degli altri o ha continuato a lavorare dopo il segnale di stop) e va ripetuta la somministrazione, usando una forma parallela.

Se oltre alla "diagnosi" è previsto, come sarebbe auspicabile, anche un **intervento**, può essere utile riferirsi a un modello di apprendimento che prenda in esplicita considerazione il tempo impiegato dai soggetti nella soluzione dei problemi. Ad esempio, il modello di Carroll (Carroll, 1963) postula cinque classi di variabili come fonti di variazione del rendimento scolastico. Di queste, tre si esprimono in funzione del tempo impiegato in attività mentali, e quindi sono connesse con la rapidità: (1) *Attitudine*, nome dato alle variabili che determinano la quantità di tempo di cui uno studente ha bisogno per apprendere un compito dato, in condizioni ottimali di istruzione e di motivazione personale (l'attitudine è alta quando uno studente ha bisogno di relativamente poco tempo per apprendere, bassa quando gli occorre più tempo della media); (2) *opportunità di apprendere*, definita come la quantità di tempo concessa per apprendere, p.es. il programma di un corso scolastico annuale o di un "ciclo" didattico più breve; (3) *perseveranza*, intesa come la quantità di tempo che lo studente è disposto a spendere nell'apprendere un compito oppure un'unità didattica (si può interpretare come una definizione operativa della motivazione ad apprendere). Le rimanenti due si riferiscono ai risultati: (4) *qualità dell'istruzione*: se non è ottima, automaticamente aumenta il tempo necessario all'apprendimento; (5) *abilità a comprendere le spiegazioni*: quanto

più lo studente ne è sprovvisto, tanto maggiore è il tempo necessario per l'apprendimento. Il modello di Carroll ha ispirato il modello del *mastery learning* (Bloom, 1968) ed altri modelli di apprendimento che esplicitano componenti psicosociali ed emotive (cfr. Carroll, 1989). Il confronto tra risultati di test più o meno sensibili alla rapidità può dare informazioni sul primo gruppo di variabili ("Attitudini"): nel programmare l'intervento, in caso di basso livello attitudinale bisognerà fare un piano razionale per incrementare, a seconda delle circostanze, l'opportunità di apprendimento, la perseveranza, la qualità dell'istruzione o l'abilità a comprendere le spiegazioni.

9. Interpretare tenendo conto dei fattori motivazionali, dell'inganno, dell'esercizio previo

Uno dei problemi più rilevanti per la teoria psicometrica è la separazione tra varianza "vera" e "spuria" delle misure (cfr. c. VI). Nella varianza "spuria" o "rumore" di disturbo vengono inclusi gli effetti di qualsiasi variabile diversa da quella che si vuol misurare. Sotto l'aspetto applicativo, in fase di interpretazione dei risultati, alcune fonti di varianza "spuria" sono molto rilevanti in quanto inducono dubbi sull'autenticità delle risposte, cioè sulla volontà o la capacità del soggetto di dare informazioni realistiche sulle sue caratteristiche personali.

La verifica dell'autenticità delle risposte è un problema generale per chi si occupa di psicodiagnosi, indipendentemente dal fatto che le informazioni siano raccolte mediante test. Le ricerche compiute sulla falsificabilità delle risposte al colloquio psichiatrico e psicologico, sorprendentemente poco numerose, raggiungono conclusioni poco incoraggianti: "non vengono riferiti studi sulla capacità di professionisti della salute mentale di identificare l'inganno in base a colloqui" (Ziskin e Faust, 1988, p. 855). Le ricerche sulla falsificabilità delle risposte ai test sono molto più numerose, anche se lontane dall'essere esaustive, ed evidenziano che le possibilità di controllo esistono, ma sono di una certa complessità.

In particolare, per chi usa i test si pongono **problemi** quali:

- accertare se i risultati di **test cognitivi** "di livello", "attitudinali" o di profitto utilizzati nell'ambito della **selezione** del personale o della **valutazione scolastica** sono stati ottenuti senza l'aiuto di facilitazioni indebite (previa conoscenza del test, veri e propri esercizi guidati per imparare le risposte giuste, "copiature" dai vicini, fortuna nell'indovinare, ecc.) e senza trarre giovamento dall'abuso di risposte date a caso o seguendo uno schema fisso (p.es.: abcdabcd, abbcccdde...);
- accertare se i risultati di **test cognitivi usati per l'accertamento dell'integrità funzionale del s.n.c.** nell'ambito di questioni giudiziarie (p.es.: rimborsi dalle assicurazioni in seguito a incidenti) sono oggetto di distorsione volontaria dei risultati;
- accertare se i risultati di **test di personalità usati nell'ambito giudiziario o di lavoro** sono oggetto di distorsione volontaria in direzione positiva (sembrare migliori di quel che si è: per farsi assumere, per farsi dare l'affidamento dei figli, ecc.) o negativa (sembrare in condizioni peggiori di quelle reali: p. es. per ottenere una pena più mite, per attirare l'attenzione ed essere curati prima o meglio degli altri).
- accertare se i risultati di **test di personalità** usati in qualsiasi ambito sono scarsamente attendibili a causa di fattori quali l'imperfetta comprensione linguistica dei quesiti, il basso livello intellettuale, un elevato numero di risposte segnate a caso o seguendo uno schema (p.es.: VFF, VFVF...).

I risultati raggiunti variano a seconda del test e del metodo di controllo usato. Nell'insieme sfatano alcuni pregiudizi diffusi fra i non professionisti dei test: le risposte indovinate per puro caso incidono in modo trascurabile sul punteggio dei test cognitivi, se questi sono ben costruiti e adeguatamente lunghi; le risposte ai test proiettivi sono falsificabili quanto o più di quelle ai questionari, senza che l'esperienza del clinico possa dare garanzie contro le falsificazioni; i dati raccolti con l'osservazione sistematica sono soggetti a distorsioni (soprattutto: "effetto d'alone") pari per importanza e simili per natura a quelle (soprattutto: "desiderabilità sociale") che influiscono sui questionari autovalutativi.

a) Falsificazione delle risposte ai test cognitivi

Nella maggior parte delle situazioni applicative, i soggetti cercano di raggiungere *il più alto punteggio possibile* ai

test cognitivi. Se con le proprie forze non sono in grado di rispondere esattamente a un quesito, cercano di "arrangiarsi". All'inizio del secolo l'ipotesi prevalente era che la maggior parte dei soggetti, trovandosi in difficoltà, cercasse di indovinare a caso (cfr. metodi per la correzione dei punteggi basati su questa ipotesi, nel cap. II); questa ipotesi ha trovato numerose disconferme e attualmente non viene accreditata (cfr. Wilcox e Wilcox, 1988). In alternativa, sono state formulate altre ipotesi e proposti metodi di controllo con esse congruenti. Ad esempio, se sono state somministrate due forme diverse del test a soggetti seduti l'uno accanto all'altro, alcuni AA. propongono di correggere ogni protocollo con ambedue le griglie: una grossa differenza fra i due punteggi (soprattutto se quello ottenuto applicando la griglia sbagliata è più alto) è indicativa di imbroglio (Roberts, 1987). Altri modelli considerano distintamente diversi parametri: p.es. una stima dell'abilità dello studente, la difficoltà relativa degli items, lo schema delle risposte giuste e sbagliate (cfr. Frary, Tideman, Watts, 1977).

La maggior parte dei metodi di controllo recenti si basa però solo sul riferimento allo "schema di risposta". I sistemi di controllo più semplici escogitati per identificare i protocolli caratterizzati da schemi di risposta "aberranti" si basano sull'ipotesi che ogni soggetto normalmente dia più risposte giuste agli items "facili" e meno risposte giuste agli items "difficili": se ripetutamente si osserva il fenomeno opposto, il protocollo è considerato sospetto ed esaminato individualmente. In alcuni casi, la "strana" modalità di risposta può esser dovuta a errori meccanici di correzione (p.es.: griglia di correzione appoggiata in modo sbagliato o, se si usa il lettore ottico, foglio di risposta compilato con matite di diversa intensità o "preso male" dalla macchina perché sgualcito o strappato), in altri casi la "stranezza" è ragionevolmente imputabile a "variabili interferenti", quali la conoscenza previa del test, le copiatore, ecc.

Alcuni metodi per l'identificazione dei protocolli "aberranti" si basano sulla correlazione "prodotto dei momenti" calcolata tra lo "schema di risposta" del soggetto e uno schema-modello.

Ad esempio, è stato proposto un **Indice Personale Punto-Biseriale (PPBI)**: Tellegen, 1988) che assume come schema-modello i livelli di difficoltà degli items:

$$PPBI = \frac{\sum_{j=1}^n (y_{ij})(d_{.j})}{n} \quad [XX.1]$$

dove:

n = numero degli items

y_{ij} = punteggio del soggetto i nell'item j, standardizzato
in riferimento a tutti i punteggi del soggetto negli n items

$d_{.j}$ = livello di difficoltà dell'item j, standardizzato in riferimento a tutti gli items

Un valore elevato in PPBI denota coerenza: indica che il soggetto prevalentemente supera gli items facili e sbaglia quelli difficili. Il valore massimo e minimo di PPBI variano a seconda del punteggio totale del soggetto. Per eliminare questo inconveniente sono state proposte varie soluzioni, fra cui la più pratica ci sembra l'**Indice di Cautela Modificato** ("CI": Harnisch e Linn, 1981):

$$CI = \frac{\max(PPBI) - PPBI}{\max(PPBI) - \min(PPBI)} \quad [XX.2]$$

dove max (PPBI) e min (PPBI) sono i valori massimi e minimi ottenibili con un particolare punteggio totale.

CI varia fra 0 e 1 ed è negativamente correlato con PPBI: 1 esprime il massimo di incoerenza e quindi il massimo bisogno di cautela. Altri AA. propongono di usare un insieme di indici, da confrontare tra loro per un'interpretazione più accurata: p.es. suggeriscono di misurare sia la tendenza a dare molte risposte non-alfa dove ci si aspettano delle alfa (p. es. in risposta a domande facili), sia la tendenza opposta (Searls, Isaacson, Scott, 1990). Altri indici sono stati proposti in riferimento a precise teorie sugli items, quali l'IRT (cfr. Levine e Rubin, 1979). Fra questi, si possono menzionare un indice NCI, che è utile soprattutto in ambiente educativo perché, oltre a evidenziare protocolli "aberranti", riesce bene a evidenziare i diversi schemi di risposta di studenti che provengono da istituzioni scolastiche diverse, e un indice ICI che è più adatto quando si vogliono estrarre dei sottogruppi di soggetti rispondenti all'ipotesi di unidimensionalità del test (Tatsuoka e Tatsuoka, 1982; cfr. cap. I).

Sotto l'aspetto teorico, questi indici si basano sull'ipotesi che le differenze individuali riguardo alla facilità soggettiva degli items siano irrilevanti. Se invece si ipotizza che le varie capacità che connotano un dato stadio dello sviluppo intellettuale vengano raggiunte abitualmente in modo non simultaneo (ipotesi piagetiana del *décalage*) il modello teorico dev'essere diverso: può darsi che un soggetto risponda male a un item "facile" perché non ha ancora acquisito la capacità necessaria per risolvere quel tipo di problema, mentre è già in grado di risolvere, per effetto del *décalage*, un item più "difficile".

In concreto, il dilemma fra i due opposti modelli teorici viene risolto al momento della scelta del "**punteggio di eliminazione**" (*cut-off*) in base al quale si decide qual è il rischio d'errore più tollerabile. Spesso si preferisce mettere il "taglio" molto in alto (p.es.: 99° centile) per ridurre al minimo il rischio di eliminare "falsi" protocolli aberranti, anche se così aumenta il rischio di prendere per buoni protocolli non attendibili (cfr. Tellegen, 1988).

In alcuni casi i soggetti hanno interesse a dare risposte inferiori al proprio livello massimo di rendimento: tipicamente, quando vogliono documentare un deficit conseguente a un trauma, per ottenere un rimborso da una società assicuratrice.

Ricerche compiute ripetutamente negli USA hanno accertato che soggetti, anche sommariamente istruiti, riescono a dare **risposte apparentemente patologiche alle più accreditate batterie neuropsicologiche** (descritte nel cap. XIII), quali la Luria-Nebraska (cfr. Mensch e Woods, 1986) e la Halstead-Reitan (cfr. Heaton et alii, 1978). In quest'ultimo studio, i neuropsicologi a cui vennero fatti esaminare i protocolli ne classificarono esattamente una proporzione variabile fra il 50% e il 69% (cioè ottennero risultati poco migliori dell'assegnazione casuale). La proporzione di protocolli falsificati che vennero correttamente identificati andava dal 25% all'81%; la proporzione di casi reali di danno cerebrale correttamente identificata variava fra il 44% e l'81%. Con l'aiuto delle risposte date dai soggetti al MMPI venne però messa a punto una funzione discriminante che, applicata da un computer, consentiva di classificare correttamente tutti i soggetti. Un altro studio ha evidenziato che anche bambini di 9-12 anni, molto sommariamente istruiti, riescono a ingannare gli esperti (cfr. Faust, Hart, Guilmette, 1987).

Molte delle ricerche sulla falsificabilità dei test neuropsicologici impiegano anche una qualche forma della scala Wechsler, che risulta falsificabile quanto i test neuropsicologici (cfr. Heaton et alii, 1978; Faust, Hart, Guilmette, 1987).

Alcuni studi che apparentemente dimostrano la difficoltà di ingannare gli esperti falsificando i risultati di test neuropsicologici vengono criticati severamente sotto l'aspetto metodologico.

Non ci risulta che siano stati applicati in questi casi gli indici elaborati per identificare i protocolli "aberranti", due dei quali sono illustrati in riferimento alle formule [XX.1] e [XX.2]. Sulla carta, questi indici sono altrettanto appropriati a queste casistiche quanto a quelle, più comuni, in cui i soggetti cercano di ottenere risultati migliori rispetto al proprio livello abituale di rendimento.

b) Falsificazione delle risposte ai test di personalità

Nell'interpretare i risultati di un test di personalità ci si pone anzitutto il problema di capire in qual misura i punteggi sono distorti dalla presenza nel soggetto di caratteristiche o di strategie estranee a quel che il test si propone di misurare. I principali problemi oggetto di ricerca sono stati: (a) la **desiderabilità sociale**, ossia la tendenza dei soggetti a dare descrizioni di sé considerate positive nel sistema di valori del gruppo sociale di riferimento (p. es., in una selezione per posti di lavoro, i soggetti tendono ad autoattribuirsi le caratteristiche che pensano siano gradite ai dirigenti da cui l'assunzione dipende); alcuni AA. considerano la desiderabilità sociale affine o identica a costrutti che vanno sotto il nome di inganno (*faking*), simulazione, dissimulazione, gestione dell'impressione, distorsione motivazionale, ma altri la considerano nettamente distinta (cfr. Wiggins, 1973); (b) **stili di risposta** (includendo in questa categoria sia i *response styles* descritti da Jackson e Messick, 1958, sia i *response sets* descritti da Cronbach, 1946), ossia tendenze a dare ai quesiti risposte che, indipendentemente dal contenuto della domanda, producono tipi di risposta qualitativamente costanti: p.es. risposte acquiescenti (dire di sì troppo spesso), evasive, estreme, rapide o lente, coerenti o incoerenti (cfr. Frederiksen e Messick, 1959). In casi relativamente meno frequenti si può anche ipotizzare che alcuni soggetti abbiano dato **risposte a caso**, per

deliberato sabotaggio o perché costretti a rispondere senza una reale comprensione e accettazione del compito (p.es. quando i test vengono somministrati nell'ambito di un largo programma, con l'aiuto di somministratori poco preparati). Si parla di "stili di risposta" soltanto se: (a) ci sono differenze individuali attendibili nella caratteristica in questione; (b) si tratta di artefatti indotti dal metodo di misura; (c) sono caratteristiche almeno in parte indipendenti dal tratto che il test misura (Nunnally, 1978, p. 659). Alcuni AA. includono anche la desiderabilità sociale fra gli stili di risposta.

Il controllo dell'autenticità delle risposte ai test di personalità viene compiuto con approcci diversi, di cui i principali sono:

(1) **scale di controllo** composte di items caratterizzati da contenuti tali che un numero elevato di risposte "vero" o "falso" sia indicativo di distorsione volontaria, di eccessiva acquiescenza, di elevata difensività: ogni item è considerato un piccolo test della tendenza a dare risposte abnormi. Per quel che riguarda l'interpretazione, in aggiunta a quanto detto nei paragrafi seguenti, va tenuto presente che le scale di controllo inducono a considerare "abnormi" i punteggi alti nelle scale patologiche. Si faccia però attenzione a non trarre illazioni sull'opposto: la normalità, abitualmente, è indicata da punteggi medi, non da punteggi bassi. Il significato dei punteggi molto bassi, in scale unidirezionali come quelle del MMPI, non è chiaro, ma di per sé non esprime normalità (cfr. Wiggins, p. 427).

(2) impiego di **items sottili**, cioè caratterizzati da contenuti che non hanno connessione apparente con la variabile misurata, ma che empiricamente si è visto essere ben correlati con la variabile (la definizione operativa di "item sottile" pone peraltro problemi che qui non affrontiamo: cfr. Ward, 1986); questo approccio non dà sempre risultati brillanti e, per quel che appare da alcuni studi, sembra esiga revisioni e aggiornamenti periodici, come tutti gli indicatori fondati empiricamente (cfr. Worthington e Schlottman, 1986);

(3) impiego di **items a scelta forzata**, cioè presentazione di due o tre affermazioni tutte pari per desiderabilità sociale, ognuna riferita a un contenuto diverso (p.es.: mi piacerebbe essere un cantante famoso - mi piacerebbe essere un calciatore famoso - mi piacerebbe essere un disegnatore di moda famoso), tra cui il soggetto deve scegliere la preferita;

(4) **indici di coerenza interna** riferiti a risposte date ad items dai contenuti simili oppure contrapponendo risposte date ad items a contenuto "ovvio" e "sottile" attinenti a una stessa dimensione di personalità (v. esempi più oltre, nel paragrafo che tratta questo aspetto).

I questionari di personalità più importanti impiegano uno o più fra gli accorgimenti sopra descritti. Di solito nessun tipo di controllo è inserito nei questionari chiamati "lista di sintomi", che pure sono molto facili da falsificare in quanto si basano su items a contenuto "ovvio". Ad esempio, si è visto (Davis, 1979 in Ziskin e Faust, 1988) che anche soggetti privi di formazione psicologica erano capaci di falsificare sia in senso positivo sia in senso negativo le risposte al questionario di depressione di Beck (descritto nel cap. XIV insieme ad altri strutturalmente simili). Il questionario MMPI è probabilmente lo strumento su cui sono state raccolte il maggior numero di ricerche riguardo alla falsificabilità. In aggiunta alle scale standard di validità L, F, K (cfr. cap. XIV), sono stati messi a punto altri approcci di controllo fra i quali si possono menzionare: (a) un metodo basato sulla differenza tra il punteggio grezzo in F e in K (Gough, 1950a; Wilcox e Dawson, 1977; Gallucci, 1984); (b) una scala aggiuntiva di dissimulazione, Ds-r (Gough, 1950b); (c) una scala basata sul rapporto tra risposte ad items "sottili" e "ovvi" (Wiener-Harman Subtle-Obvious subscales: cfr. Ziskin e Faust, 1988, pp. 859-60). Queste scale, come quelle standard, consentono di identificare una larga proporzione dei soggetti che hanno cercato di apparire più patologici di quel che sono, ma non di "correggere" i risultati. Si vedano inoltre (cap. XIV) le scale di controllo introdotte nel MMPI-2.

L'influsso della desiderabilità sociale e degli stili di risposta in genere sulle risposte ai **test proiettivi** è stato pochissimo studiato, anche se noti studiosi di questi test ammettono che la desiderabilità sociale abbia "un effetto molto significativo su quanto viene di fatto trasmesso dal soggetto" (Exner, 1978), fondando le loro conclusioni pratiche sull'analisi dei processi di risposta tipicamente attivati dai soggetti (cfr. cap. XVIII; Exner, 1986).

Nel cap. XVIII sono stati esposti dati riguardo a singoli test, con particolare attenzione al Rorschach, che è lo strumento più studiato sotto questo aspetto (e che quindi, paradossalmente ed erroneamente, potrebbe sembrare il più facile da falsificare). Qui vengono riassunti problemi metodologici generali e dati alcuni esempi illustrativi.

Le prime ricerche sulla "falsificazione delle risposte" ai test proiettivi si collocano intorno al 1940 (Fosberg, 1938; 1941; 1943). Si caratterizzavano come disegni test-retest senza campioni di controllo, analizzati mediante statistiche poco potenti (soprattutto: chi quadro): vennero molto criticate sotto l'aspetto metodologico (Cronbach, 1949a). Il tentativo di migliorare il disegno test-retest aggiungendo un campione di controllo ha prodotto risultati poco interpretabili in alcuni studi (p.es.: Easton e Feigenbaum, 1967) ed è stato poco seguito. Attualmente vengono preferiti disegni basati su numerosi gruppi di controllo, che quasi sempre includono fra le ipotesi la possibilità di scoprire l'inganno.

Tutti i metodi incontrano difficoltà nel controllare le variabili "quantità delle risposte date", che appare facilmente manipolabile da chi tenta l'inganno (i protocolli falsificati, di solito, contengono poche risposte) e che influisce su tutte le altre variabili (cfr. Perry e Kinder, 1990). Peraltro, in base al solo indicatore "basso R" si può soltanto affermare che un protocollo non è valido, senza poter dire che è falsificato (cfr. Cohen, 1990).

L'esame da parte di "esperti" riesce in genere a identificare solo in parte i protocolli contraffatti, e per di più segnala come contraffatti una parte dei protocolli autentici. A questo riguardo la seguente ricerca statunitense, accuratamente progettata, può fornire un esempio illustrativo. Vennero confrontate le risposte date al Rorschach da: (a) sei pazienti psichiatrici ospitalizzati, che un gruppo di psichiatri e psicologi clinici aveva concordemente diagnosticato come "schizofrenici paranoidi", (b) un gruppo di sei studenti universitari, che non conoscevano né il test né la psichiatria a cui era stato detto "Fate in modo che i risultati del test mostrino che siete degli schizofrenici paranoidi, senza che si capisca che imbrogliate" (verranno chiamati "studenti non informati"), (c) un altro gruppo di sei studenti a cui era stata data la stessa istruzione ma era anche stato spiegato come si manifestava la schizofrenia paranoide ("studenti informati") e (d) un gruppo di controllo costituito da sei studenti a cui il Rorschach fu somministrato con le modalità standard (Albert, Fox, Kahn, 1980). Ogni protocollo fu valutato da sei-nove giudici, tutti esperti rorschachisti. Vennero classificati come psicotici il 72% degli studenti "informati", il 46% di quelli "non informati" e il 24% dei normali. Vennero indicati come falsificati il 4% dei protocolli di pazienti, il 9% di quelli degli studenti "informati", il 7% di quelli "non informati" e il 2% dei normali a cui il test era stato somministrato con le istruzioni standard. Rispetto ai questionari, l'identificazione e la correzione delle distorsioni nelle risposte ai test proiettivi pone maggiori problemi, perché è molto difficile predire in qual modo un soggetto risponderà quando cerca di sembrare "migliore" o "peggiore" (cfr. Carp e Shavzin, 1950).

Soprattutto nelle applicazioni di tipo giudiziario e di lavoro, va inoltre tenuto presente che nei test più comunemente usati (MMPI, Rorschach, ecc.) la falsificazione viene aiutata da molti sussidi: pubblicazioni apposite, corsi più o meno clandestini in cui vengono insegnate le "risposte giuste", suggerimenti da chi ha già sostenuto il test, ecc. I sussidi riguardano indifferentemente qualsiasi tipo di test (inclusi quelli di disegno, quelli basati su macchie, quelli simili al TAT) e in alcuni casi sono predisposti con la consulenza di persone che conoscono gli strumenti.

Potrebbero essere adattati e validati anche per i test proiettivi alcuni dei mezzi di controllo illustrati qui appresso, a cui viene dedicato più spazio per la rilevanza che hanno nell'interpretazione dei risultati.

c) Controlli basati su "indici di coerenza"

Gli indici di coerenza sono misure di variabilità intra-individuale basate sull'ipotesi che il soggetto "attendibile" dia risposte simili a domande simili (o identiche) tra loro per contenuto e che riceva punteggi simili nelle scale riferite a tratti correlati positivamente fra loro.

Uno dei primi indici di questo tipo è il **Test-Retest Index** (TR: Buechli e Ball, 1952; Greene, 1979) proposto per il MMPI. Si basa sui 16 items che vennero stampati nel MMPI due volte per facilitare la correzione meccanografica: si contano le incongruenze fra prima e seconda risposta data all'item. Questo indice ha un evidente limite: se il soggetto ha risposto sempre "Vero" o sempre "Falso" a tutti gli items risulta perfettamente coerente, anche se è ovvio che le sue risposte non sono attendibili.

Per compensare questo limite, insieme all'indice TR si può usare la **Carelessness Scale** (CS: "Scala di trascuratezza"; Greene, 1978), che utilizza anche coppie di items a cui, se si è coerenti, bisogna dare risposte

opposte. Anche le scale VRIN e TRIN di Tellegen (1988) verificano la coerenza in modo analogo. Una scala di controllo analoga a TR è stata inclusa nel PPS di Edwards, in aggiunta all'accorgimento di usare quesiti a scelta forzata.

Un **indice di accuratezza basato sulla correlazione fra items pari e dispari della stessa scala** è stato proposto da Cattell e collaboratori per individuare i protocolli a cui presumibilmente i soggetti hanno risposto a casaccio (Cattell, Eber, Tatsuoka, 1970, pp. 49-50).

Le **scale di controllo basate sulla variabilità intraindividuale** pongono molti problemi interpretativi: la variabilità individuale non è una caratteristica molto stabile e può essere in alcuni casi interpretata come indicativa di complessità dei costrutti misurati. Secondo alcuni Autori, potrebbe essere anche indicativa del fatto che la personalità di alcuni soggetti non si presta ad essere descritta con riferimento ad un sistema di tratti ("*untraitedness*": cfr. Tellegen, 1988).

d) Misure di desiderabilità sociale

Numerosi studi, fin dalla prima metà del secolo (P.E. Vernon, 1938, cit. in P.E. Vernon, 1963), si sono occupati del costrutto noto come desiderabilità sociale. Notevoli sono i contributi di A.L. Edwards, che studiò la desiderabilità sociale con particolare attenzione al MMPI e la definì in termini empirici, con riferimento a un criterio costituito dalle valutazioni di desiderabilità sociale attribuite a ogni item da un gruppo di giudici: "una risposta socialmente desiderabile è una risposta Vero a un'affermazione con un valore scalare socialmente desiderabile oppure una risposta Falso ad un'affermazione con un valore scalare indesiderabile, lasciando il concetto non definito per affermazioni con valori scalari esattamente neutri, cioè con valore 5 su una scala di desiderabilità sociale a 9 punti" (Edwards, 1957a; cfr. anche Edwards, 1970). Il criterio di Edwards è stato largamente accettato, nonostante alcuni AA. abbiano evidenziato che la correlazione tra desiderabilità "sociale" (definita dai giudici) e "personale", pur essendo elevata (Boe e Kogan, 1964) non è lineare ed ha quindi dei limiti applicativi (Goodstein e Heilbrun, 1959).

La tipica scala di desiderabilità sociale a cui Edwards si riferisce è la scala L del MMPI, composta di items socialmente desiderabili. La scala L e in genere le scale di desiderabilità sociale sono negativamente correlate con misure di ansia e positivamente correlate con la scala Pt e con il primo fattore estratto dal MMPI. Questo quadro correlazionale ha indotto alcuni Autori a interpretare i punteggi nelle scale patologiche del MMPI essenzialmente come misure della tendenza a descriversi in termini negativi. La correlazione però esprime una simmetria: se ci si basa solo sui dati correlazionali, si può affermare con pari diritto sia che le scale di desiderabilità sociale misurano la tendenza a descriversi in termini positivi, sia che le scale di desiderabilità sociale sono misure generali di adattamento. Sui soli dati correlazionali, di fatto, si possono esprimere una notevole varietà di ipotesi interpretative (cfr. Block, 1965). Per un'interpretazione realistica della desiderabilità sociale bisogna separare le due fonti di variazione (psicopatologia e desiderabilità sociale) e riferirsi a criteri comportamentali esterni ai questionari. Un tentativo di separare le due variabili è stato compiuto da Block, mettendo insieme una scala (scala ER-S, da Ego-resiliency - Subtle) composta di items neutri sotto l'aspetto della desiderabilità sociale e ben correlata positivamente con la scala di desiderabilità sociale di Edwards e negativamente con la scala Pt del MMPI (Block, 1965). Dato che la scala ER-S era poco sensibile alle distorsioni imputabili alla desiderabilità sociale e in particolare ai tentativi espliciti di imbroglio in direzione positiva pur rimanendo altamente correlata con le misure di desiderabilità sociale e di ansia, l'A. concluse che era corretto interpretare le misure di desiderabilità sociale come misure di ansia, e non viceversa. Conclusioni simili sono state tratte ragionando sulle differenze significative osservate, nella costruzione del MMPI, fra campioni clinicamente differenziati (cfr. Wiggins, 1973).

Attualmente, è comunemente ammesso che la desiderabilità sociale spiega una parte considerevole della varianza dei questionari autodescrittivi (secondo alcuni AA. satura il primo fattore di tutti i questionari di questo tipo: cfr. Edwards, 1964), che però continuano ad esser considerati utili nella misura in cui la varianza che resta dopo aver eliminato l'influsso della desiderabilità sociale è tale da produrre altri fattori sufficientemente forti.

Per di più, lo stesso punteggio in "desiderabilità sociale" è considerato non tanto un mezzo per verificare se i soggetti "imbrogliano" quanto "un'importante variabile-predittore, per suo proprio diritto" (Lanyon e Goodstein,

1982, p. 163). Per esempio, la *Marlow-Crowne Social Desirability Scale (MCS*: Crowne e Marlow, 1960) è un discreto predittore di costrutti quali la difensività, la vulnerabilità dell'auto-stima, la dipendenza dall'approvazione, l'"apertura" nel comunicare con i propri pari e con lo psicologo all'interno del colloquio clinico, con la terminazione prematura della psicoterapia (cfr. Evans, 1982).

La desiderabilità sociale peraltro è una variabile composita: "non esiste una singola 'desiderabilità sociale, ma **numeroso distinte desiderabilità a seconda della situazione di testing**" (Barton, 1986). Ad esempio, Cattell e coll. indicarono due distinte desiderabilità sociali: (a) in direzione di valori morali e religiosi comuni agli appartenenti ad una cultura; (b) in direzione di quel che viene apprezzato dal gruppo di appartenenza, soprattutto nel caso di gang giovanili (Cattell, Pierson, Finkbeiner, 1976). Tutte e due queste componenti sono presumibilmente presenti quando si chiede ai soggetti, a scopo sperimentale, di rispondere al questionario cercando di dare una "buona" o un "cattiva" impressione. Un tipo di desiderabilità sociale diverso si aggiunge quando i questionari vengono usati a scopo di selezione per un lavoro: in quei casi i soggetti prendono come riferimento lo stereotipo dell'impiegato "di successo" in quel settore, e possono dare risposte distorte in direzione diversa a seconda che l'assunzione sia come funzionario di banca, come venditore, come tecnico, come psicologo.

Tenendo conto di un insieme di studi prodotti negli ultimi quarant'anni, si può dire che **le componenti più importanti di un punteggio in "desiderabilità sociale" sono**: (1) livello di adattamento generale del soggetto: chi è ben adattato tende a dare una buona immagine di sé agli altri, soprattutto in situazioni quali la potenziale assunzione a un posto di lavoro; di fatto, i pazienti affetti da nevrosi hanno la tendenza ad autodescrivere in termini sfavorevoli (P.E. Vernon, 1963, p. 202), ma è anche noto che la tendenza di molti pazienti psichiatrici a descriversi in termini sfavorevoli si accompagna alla tendenza a produrre comportamenti socialmente indesiderabili (cfr. Heilbrun, 1964) ed è stato constatato che prigionieri "non violenti" ottengono punteggi medi più alti di desiderabilità sociale (scala di Marlowe-Crowne) rispetto a prigionieri "violenti" (Selby, 1984); (2) conoscenza che gli individui hanno delle proprie caratteristiche personali (la tendenza ad autoattribuirsi caratteristiche positive, anche quando non c'è inganno deliberato, corrisponde all'"effetto di alone" che si osserva nelle eterovalutazioni: cfr. P.E. Vernon, 1963, p. 202); (3) franchezza dei soggetti (Nunnally, 1978, p. 558); (4) capacità di autoanalizzarsi (P.E. Vernon, 1963, p. 202).

Secondo Cattell la misura della desiderabilità sociale dei questionari dovrebbe includere: (1) una stima dell'intensità del desiderio che un individuo ha di distorcere un dato tratto; (2) un riferimento alla situazione di testing a cui è connesso quel desiderio; (3) una stima della "vulnerabilità" di quel test alla distorsione, per ragioni connesse con le modalità di costruzione; (4) una stima della capacità dell'individuo nel distorcere il test (Cattell, 1986c).

Questi principi interpretativi si possono utilizzare nell'interpretazione di tutte le scale riferite al costrutto "desiderabilità sociale", qualunque sia la loro denominazione specifica: p.es. la scala L del MMPI, la scala "Buona impressione" del CPI, la scala di distorsione motivazionale presente nelle forme C e D del 16 PF. In qualche misura e a titolo di ipotesi queste conclusioni si possono applicare all'interpretazione dei risultati di scale i cui contenuti vanno in direzione opposta, come ad esempio la scala F del MMPI, composta di items socialmente indesiderabili e assai meno studiata.

Sono anche state formulate varie **proposte per eliminare le distorsioni** indotte dalla desiderabilità sociale o per utilizzare costruttivamente il punteggio di desiderabilità sociale, senza però arrivare a soluzioni soddisfacenti. Le principali soluzioni proposte sono:

(a) scrivere **items il più possibile neutri quanto a desiderabilità sociale** (p.es. con valore intermedio nelle valutazioni di desiderabilità sociale date da giudici), usando anche accorgimenti linguistici tali da indurre una minor proporzione di risposte conformiste (p.es. si è osservato che ci sono più soggetti che rispondono "No" a un item sgradevole rispetto ai soggetti che rispondono "Sì" se lo stesso contenuto viene proposto in forma di comportamento socialmente accettabile: P.E. Vernon, 1963, p. 202); questa soluzione è particolarmente difficile da mettere in atto con i questionari che misurano variabili psicopatologiche o altre caratteristiche in sé poco desiderabili;

(b) **escludere il punteggio di desiderabilità sociale con mezzi statistici** (correlazione parziale): così facendo però si riduce l'attendibilità di tutte le scale, si elimina gran parte dell'informazione attinente ai fattori "di secondo ordine" ansietà (Barton, 1986; Castelli-Sawicki, Wallbrown, Blixt, 1983) e estroversione-introversione (Cattell, Eber, Tatsuka, 1970, p. 52) e si rendono

meno confrontabili tra loro scale omonime di questionari diversi, dato che la desiderabilità sociale è una caratteristica complessa, interpretata in vario modo a seconda degli Autori (Nunnally, 1978, p. 560);

(c) formulare gli items con la modalità della "**scelta forzata**", che però oltre all'inconveniente di una notevole difficoltà di costruzione comporta effetti collaterali indesiderabili, quali la produzione di punteggi ipsativi (cfr. cap. III);

(d) introdurre una **correzione basata sul punteggio riportato dal soggetto in una scala di controllo** (p.es.: K al MMPI, MD al 16 P.F.): i risultati ottenuti con questo approccio hanno rivelato che è molto difficile fissare in modo valido sia i "punteggi di eliminazione" (cut-off) sia le frazioni di punteggio da sottrarre (cfr. il metodo introdotto nella più recente versione del CPI: Lanning, 1989): fra l'altro, il livello ottimale di correzione dipende dalle caratteristiche del campione (in alcuni studi appaiono anche differenze tra sessi: cfr. Castelli-Sawicki, Wallbrown, Blixt, 1983) e non può essere stabilito una volta per tutte (cfr. Argentero, 1985 per il 16 PF, Lanning, 1989 per la completa revisione delle scale di controllo del CPI nella più recente edizione statunitense e quanto detto nel cap. XIV in riferimento alla scala K del MMPI);

(e) introdurre una **correzione basata sui risultati ottenuti somministrando il questionario in situazioni diverse** (p.es.: inganno deliberato/onestà; quando si cerca di avere un lavoro, quando si cerca di ottenere un appuntamento con un ragazzo o una ragazza, ecc.); la correzione può essere basata semplicemente sulla differenza tra i punteggi medi ottenuti nelle due condizioni confrontate (onestà/inganno, tipicamente: cfr. Gordon e Gross, 1978), sulla varianza di queste differenze (Gordon e Gross, 1978; Lautenschlager, 1986), sulla correlazione tra le due serie di punteggi, confrontata con indici test-retest (Gordon e Gross, 1978; Lautenschlager, 1986), sulle analisi fattoriali dei punteggi ottenuti in condizioni diverse (Cattell, Eber, Tatsuoka, 1970, pp. 53-55; Cattell, 1986). Queste soluzioni sono laboriose, sia perché esigono la raccolta di dati empirici riferiti alle più comuni situazioni che stimolano la distorsione, sia perché la direzione e l'entità della distorsione sembrano essere diverse in funzione di caratteristiche dei soggetti, quali ad esempio il sesso, il che comporta la raccolta di dati differenziati per campioni variamente caratterizzati.

[Aggiunta 2021. Calcolare norme statistiche differenziate su campioni a cui il test è stato somministrato in situazioni diverse e standardizzare i punteggi in riferimento alle norme corrispondenti alla specifica situazione. Ad esempio, il test Taleia 400-A (Boncori e Di Marco, 2007) consente di standardizzare i punteggi in riferimento a norme calcolate per situazioni di diagnosi clinica, di orientamento, di selezione del personale, selezionabili direttamente dal software.]

Tutte le soluzioni escogitate si basano su dati empiricamente raccolti. Sono quindi strettamente riferite alle caratteristiche della popolazione a cui appartengono i campioni su cui sono state validate: se "esportate" da una cultura all'altra (p.es. dagli USA all'Italia) esigono controvalidazione, se elaborate più di dieci anni fa esigono revisione per un necessario adeguamento ai mutati oggetti della desiderabilità sociale. Anche i punteggi-limite devono essere oggetto di verifica e di revisione periodica.

e) Stili di risposta

Le ricerche sugli stili di risposta che possono indurre i soggetti a produrre autodescrizioni involontariamente distorte hanno avuto un notevole sviluppo fra il 1950 e il 1970 (cfr. Rorer, 1965; Damarin e Messick, 1965; McGee, 1967). Allo stato attuale, si può dire che la controversia sugli stili di risposta è stata solo un aspetto specifico della problematica generale riguardante l'interpretazione dei risultati dei test (cfr. Wiggins, 1973, p. 416) e che la maggior parte delle preoccupazioni sollevate sulle distorsioni originate dagli stili di risposta si sono rivelate falsi allarmi (Nunnally, 1978, p. 559). Diamo quindi solo alcuni cenni sugli "stili di risposta" che hanno stimolato una maggior produzione di ricerche.

L'**acquiescenza** è la tendenza a rispondere "sì" ai quesiti indipendentemente dal loro contenuto. Fu studiata originariamente in rapporto alla scala F per la misurazione degli atteggiamenti autoritari. Attualmente, dopo numerose ricerche, molte delle quali oggetto di critiche metodologiche serie, si ritiene che la tendenza ad acconsentire sia di modesta importanza, sia come variabile di personalità in sé sia come fonte di invalidità nelle risposte ai questionari che misurano altre variabili.

La costruzione di misure di acquiescenza non è semplice, in quanto richiede un accurato controllo di altre variabili (p. es. della desiderabilità sociale), ed esige una provata eterogeneità nei contenuti. Anche quando tali esigenze vengono sostanzialmente rispettate, come nella scala Dy-3 elaborata da Jackson e Messick per il MMPI (cfr. Jackson e Messick, 1962), i risultati sono tutt'altro che univoci (cfr. Block, 1965 e la polemica riassunta in Wiggins, 1973). La soluzione abitualmente consigliata per superare il problema consiste nel costruire i test in modo che le risposte indicative siano costituite per metà da Sì e per metà da No (Nunnally, 1978, p. 669), ma anche in questo caso i risultati possono essere di dubbio valore (cfr. Knowles, 1963). Introdurre correzioni per escludere l'influsso dell'acquiescenza, che è correlata con l'ansietà, comporterebbe distorsioni maggiori (Cattell, Eber, Tatsuoka, 1970, pp. 52-53).

Tuttavia, quando s'interpretano i risultati di questionari sintomatici (cfr. cap. XIV) per i quali la risposta Sì è sempre indicativa di patologia, è bene porsi il problema dell'acquiescenza, soprattutto se un soggetto ha un elevato livello di ansia, e cercare un riscontro esterno per gli eventuali punteggi patologici che risultano dal questionario sintomatico.

La **tendenza a dare risposte estreme** si manifesta soprattutto nella misurazione degli atteggiamenti, quando la risposta viene registrata lungo una scala che esprime intensità variabili. Anche questa tendenza incide limitatamente sulla valutazione della maggior parte delle persone. Se si hanno elementi per ritenere preferibile apportare una correzione, la soluzione più semplice consiste nel ridurre i punteggi di ogni item a dicotomie (Nunnally, 1978, p. 670).

La *tendenza a dare risposte devianti* è stata studiata e posta al centro dell'attenzione nel quadro tipicamente empirista che fonda l'identificazione degli indicatori sul metodo dei gruppi a contrasto (cfr. Wiggins, 1973, pp. 416-420). L'assunto principale su cui si basa l'ipotesi è che le persone che tendono ad avere reazioni devianti "significative per l'anormalità" hanno reazioni devianti in genere, anche in comportamenti apparentemente irrilevanti (cfr. distinzione tra items "ovvi" e "sottili"). Conseguentemente, se si mettono a confronto le risposte date a un qualsiasi questionario o a un qualsiasi compito (p.es. scelta o produzione di disegni) da un gruppo "deviante" e da un gruppo "normale", è facile identificare una notevole quantità di risposte tipiche del gruppo deviante anche "in aree non critiche del comportamento" (Berg, 1955; Berg, 1967). Nella sua formulazione più radicale, questa ipotesi implica la non rilevanza dei contenuti e del costrutto teorico nella diagnosi psicologica.

L'ipotesi della deviazione generalizzata è stata sottoposta a varie critiche teoriche (cfr. Wiggins, 1973). Sotto l'aspetto pratico, è stato dimostrato che è facile costruire scale con il metodo sopra delineato e provarne la validità in una prima ricerca, ma che la controvalidazione su campioni diversi dai primi evidenzia fatti contrastanti con l'ipotesi: la validità delle scale basate sui contenuti resta, mentre la validità delle scale basate su items solo empiricamente discriminativi scompare (cfr. Goldberg e Slovic, 1967).

(C) IL "RAPPORTO PSICOLOGICO" E LA COMUNICAZIONE DEI RISULTATI

I risultati dei test vengono di solito esposti in forma numerica e commentati con alcune frasi che mirano a evidenziare una struttura d'insieme, che unifichi in termini psicologici le varie informazioni disponibili. Questo commento viene chiamato "rapporto psicologico" e può essere scritto per comunicare i risultati dei test ad altri professionisti (p.es. a uno psichiatra, a un collegio docente, a un datore di lavoro) oppure all'interessato stesso.

La stesura del "rapporto psicologico" è affidata per lo più alla competenza psicologica generale acquisita da chi deve redigerlo. Alcune informazioni e norme metodologiche possono aiutare a produrre un rapporto che non vanifichi la validità faticosamente assicurata ai test.

10. La diagnosi psicologica e il rapporto psicologico sui risultati dei test

I test, come si è detto più volte nel corso di questo Manuale, sono strumenti per la raccolta di informazioni. Un insieme di informazioni però, indipendentemente dalla qualità delle singole informazioni raccolte e dal loro numero, non è una diagnosi. "La diagnosi è l'organizzazione critica di dati osservati ed evocati allo scopo di prendere una decisione" (Saraceni e Montesarchio, 1988). Per una diagnosi psicologica valida si ritiene comunemente che i dati siano raccolti con una varietà di approcci metodologici (quelli "classici" sono il colloquio, l'osservazione sistematica e i test), così da compensare le distorsioni tipiche di ciascuno.

Si tenga presente che la correlazione tra valutazioni ottenute con approcci diversi, anche se statisticamente significativa, di solito non è elevata. Ad esempio, in uno studio sulla valutazione della psicopatia (Hare, 1985), compiuto su circa 300 detenuti di un penitenziario statunitense, si trovò il quadro correlazionale riprodotto nella tab. XX.4.

TAB. XX.4

Correlazioni tra procedimenti diversi per la valutazione della psicopatia (sopra la diagonale) e (sotto la diagonale) numero di soggetti su cui si basa ciascuna correlazione (da Hare, 1985)

Variabile	1	2	3	4	5	6	7	8	9
1. Valutazione		.80*	.57*	.28*	.21	.22	-.29*	.27	.36*
2. Lista di controllo	229		.67*	.38*	.26*	.27*	-.32*	.35*	.33*
3. DSM-III	229	245		.35*	.29*	.21	-.37*	.33*	.44*
4. SRP*	184	226	199		.26	.36*	-.53*	.40*	.49*
5. Pd	105	138	122	120		.14	-.34*	-	-
6. Ma	105	138	122	120	138		-.31*	-	.27
7. So	175	219	193	200	118	118		-.42*	-
8. Pd + Ma	105	138	122	120	138	-	118		-
9. Pd - So	105	118	102	109	-	118	- -	-	-

* *Self-Report Psychopathy Scale* di Hare

I coefficienti di correlazione relativamente più elevati sono fra strumenti simili: Valutazione globale, lista di controllo e DSM-III correlano mediamente $r = 0,60$; autovalutazione della psicopatia e valutazioni della psicopatia desunte da scale di questionari autovalutativi hanno coefficienti di correlazione mediani intorno a $r = 0,35$ (in valore assoluto); metodi osservativi e autodescrittivi $r = 0,27$. I coefficienti di correlazione tra scale dei questionari sono, in questo caso, particolarmente bassi perché la definizione del costrutto "psicopatia" è cambiata nel tempo e non corrisponde esattamente né a quella su cui si basano le scale omonime (nel caso: la scala Pd del MMPI) né a quella su cui si basano scale che si riferiscono a costrutti affini ma non identici (scala So del CPI, Scala Ma del MMPI). Coefficienti di correlazione di questo livello indicano chiaramente la necessità di fondare le diagnosi su una pluralità di metodi, dato che nessuno dei singoli approcci consente di cogliere una porzione sufficientemente ampia della realtà.

Va anche tenuta presente l'importanza di raccogliere informazioni da una certa varietà di fonti. Per esempio, problemi emotivi e comportamentali di bambini vengono valutati in modo abbastanza diverso da insegnanti e genitori. I coefficienti di correlazioni più elevati si osservano per problemi quali l'iperattività (intorno a $r = 0,45$) e l'ansietà (intorno a $r = 0,30$); per la maggior parte dei problemi specifici (p.es.: ha delle ossessioni, è crudele con gli altri, manca di senso di colpa, dice bugie, si affatica facilmente, ecc.) i coefficienti di correlazione sono vicini allo zero, soprattutto per i bambini al di sotto dei sei anni (cfr. Mealor e Richmond, 1980; Verhulst e Akkerhuis, 1989). Lo scarso accordo non stupisce, se si pensa alla diversità delle situazioni in cui il ragazzo viene osservato e alla differenza tra "sé sociali" posti in atto nei due diversi contesti. Informazioni attinte da fonti diverse non solo danno allo psicologo un'immagine più completa, ma in caso siano discordanti rivelano qualcosa in più della loro somma. In base a queste considerazioni, se il "rapporto psicologico" sintetizza soltanto le informazioni raccolte con i test, non è una diagnosi, anche se ne ha tutte le caratteristiche formali.

11. Preparazione del rapporto psicologico

L'allestimento del rapporto psicologico (d'ora in poi: RP) esige una preparazione materiale dei dati, in forma ordinata e facile da commentare, e una preparazione mentale dello psicologo: riferita ai dati, al soggetto, a se stesso.

a) Preparazione dei dati su cui si basa il rapporto

La preparazione materiale dei dati, che può essere affidata a personale tecnico e a un programma computerizzato, consiste nel disporre tutti i dati raccolti in un **prospetto di facile ispezione**. Questo prospetto è più facile da realizzare se tutti i test esprimono i risultati in punteggi standardizzati (cfr. fig. XX.13). Nel caso i punteggi dei vari test somministrati alla stessa persona non fossero espressi nello stesso tipo di punteggio standardizzato, vanno trasformati in modo da avere uniformità (per i metodi di trasformazione, cfr. cap. III). Ovviamente, dovranno essere tutti trasformati nella scala che ha il minor numero di livelli: p.es., se alcuni punteggi sono *stanine* e altri sono punti T, vanno tutti trasformati in *stanine*.

La forma grafica in cui spesso sono espressi i risultati è solo apparentemente più chiara della forma numerica. Di fatto, si è visto che quando i dati sono presentati in forma solo grafica, l'accordo interpretativo non è soddisfacente (Gibson e Ottenbacher, 1988).

Anche i risultati espressi in forma non numerica (p.es. Rorschach, TAT) dovranno essere sintetizzati in un prospetto che consenta a chi interpreta di avere davanti a sé, rapidamente ispezionabili, tutte le informazioni raccolte.

FIG. XX.13

Prospetto che riporta in forma grafica e numerica i risultati di un insieme di test

SCUOLA SUPERIORE "DANTE ALIGHIERI"		N° PROT. 124								
TEST		1	2	3	4	5	6	7	8	9
* * OTIS ('V:ED')	4	####								
D48 ('G')	7	#####	##							
* * PROFILO DELLE ABILITA' (GATB)										
- PERCEZIONE IN LAVORI UFFICIO	2	##								
- CALCOLO ARITMETICO	5	#####								
- RAPPORTI SPAZIALI	6	#####	#							
- VOCABOLARIO	3	###								
- ABBINAMENTO DI OGGETTI	5	#####								
- RAGIONAMENTO ARITMETICO	6	#####	#							
- ABBINAMENTO DI FORME	5	#####								
* * QUESTIONARIO R										
- A: SIZOTIMIA/AFFETTOTIMIA	3	###								
- B: INTELL. CRISTALLIZZATA	6	#####	#							
- C: FORZA DELL'IO	5	#####								
- E: SOMMISSIONE/DOMINANZA	2	##								
- F: MALUMORE/BUONUMORE	4	####								
- G: FORZA DEL SUPER IO	1	#								
- H: THRECTIA/PARMIA	2	##								
- I: HARRIA/PREMSIA	4	####								
- L: ALAXIA/PROTENSION	1	#								
- M: PRAXERNIA/AUTIA	4	####								
- N: INGENUITA'/SCALTREZZA	2	##								
- O: SENTIM. DI ADEGUATEZZA/COLPA	7	#####	##							
- Q1: CONSERVATOR/RADICALISMO	7	#####	##							
- Q2: DIPEND. DAL GRUPPO/AUTOSUFF.	6	#####	#							
- Q3: INTEGRAZIONE CONCETTO DI SE'	2	##								
- Q4: TENSIONE ERGICA	4	####								
- AN: ANSIETA'	6	#####	#							
* * P.N.P.										
- L: DESIDERABILITA' SOCIALE	7	#####	#							
- N1	4	####								
- N2	6	#####	#							
- P	4	####								

b) Impostare la "traccia" del rapporto

Quando finalmente l'insieme dei dati è raccolto, lo psicologo deve affrontare il problema tipico dei temi scolastici: "Come comincio?".

Una norma pratica può essere quella riassunta da Holt in: "cerca di trovare qual è la vera domanda a cui il cliente

vuol avere risposta" (Holt, 1986). In altri termini, il commento ai dati può e deve esser diverso a seconda della finalità per cui i test sono stati somministrati (v. anche, su questo punto, Saraceni e Montesarchio, 1988).

Il secondo punto della preparazione mentale dello psicologo consiste nel chiarire a se stesso chi è **il destinatario del rapporto** e qual è il ruolo che lo psicologo assume nei suoi confronti. L'ansietà connessa con le preoccupazioni (giustificate) su "chi altro" leggerà il rapporto può essere meglio controllata se si decide di fare questo controllo in un secondo momento, nell'ambito della revisione del rapporto già scritto. Al momento della prima stesura è bene preoccuparsi soltanto di scrivere la verità, di scriverla in modo da dare risposta al quesito per il quale i test sono stati somministrati, e di scriverla in modo che sia correttamente e costruttivamente recepita dall'interlocutore principale.

A seconda dello scopo del RP e del suo destinatario si decideranno i "paragrafi" in cui articolare il contenuto. Alcuni AA. suggeriscono schemi di massima, più o meno complessi, che si riferiscono alla psicodiagnosi nel suo insieme ma includono anche una parte riguardante i test. Nei rapporti su soggetti che presentano problemi clinici, inclusi la maggior parte di quelli presentati nell'ambito di perizie giudiziarie cfr. Kellerman e Burry, 1981; Saraceni e Montesarchio, 1988, pp. 117-119; Sundberg, 1977, p. 262), sono sempre considerate aree quali:

- intelligenza
- strutture della personalità
- dinamiche della personalità (difese, controlli, ecc.).

A seconda della problematica e dell'orientamento teorico del clinico viene dato maggiore o minore spazio ai problemi riguardanti il quadro psicopatologico, le caratteristiche fisiche e fisiologiche, le relazioni interpersonali e familiari, l'ambiente.

Un rapporto psicologico completo include di solito numerose informazioni.

[2021. (1) **Dati anagrafici e anamnestici**, che attualmente è possibile gestire con software che creano banche dati funzionali per gli psicologi, oppure inserire in una propria banca dati che facilmente si può organizzare con software generici come Excel, Access, FileMaker e altri. Questi dati non fanno parte del report, ma possono essere utili per interpretare bene i dati raccolti e produrre un report accurato e valido, che tenga conto dell'influsso che hanno alcuni di questi dati (età, sesso, livello culturale ecc.) sui risultati degli strumenti psicodiagnostici, com'è documentato nelle pagine precedenti. (2) Trascrizione (o registrazione audio allegata) e sintesi del **colloquio introduttivo**. Si annotano qui anche informazioni sul comportamento tenuto durante il colloquio dal soggetto e da eventuali accompagnatori, sulle interazioni del soggetto con l'intervistatore e con eventuali altre persone e sullo stile comunicativo del soggetto, sul modo in cui affronta "domande difficili" e su come parla (o non parla) dei suoi problemi. La sintesi del colloquio si conclude con ipotesi diagnostiche e indicazioni sugli strumenti psicodiagnostici da utilizzare. Fra gli strumenti ci possono essere altri colloqui sia con il soggetto sia con altre persone che possono dare informazioni aggiuntive e/o di controllo. (3) Funzionamento cognitivo. Secondo l'età e i problemi del soggetto si sceglierà se verificare il livello cognitivo generale (QI), la contrapposizione tra Intelligenza culturale (Gc) e Intelligenza fluida (Gf), diverse abilità più o meno specifiche: ragionamento verbale, ragionamento spaziale, rapidità percettiva, competenze linguistiche, aritmetiche, informatiche e altre. Ecco un esempio di schema per l'esame psicologico di bambini, basato su esperienze compiute con migliaia di casi (Blau, 1991).

[2021. Allo schema di Blau sono state fatte numerose aggiunte, per renderlo una base utilizzabile anche con gli adulti e inserire informazioni raccomandate da testi autorevoli, come il DSM, e da manuali di psicologia clinica. È stata evidenziata l'articolazione dello schema in tre settori:

- una prima parte che contiene dati anagrafici e anamnestici,
- una seconda parte con i risultati di test, osservazioni dirette, colloqui e altri esami,
- la terza parte, "Sintesi", è il vero e proprio Rapporto.

Nella prima parte, anagrafica e anamnestica, le informazioni contenute nel primo riquadro vanno annotate qualunque sia la situazione specifica per cui viene posta in atto la valutazione.

Lo schema di Blau, presente nel testo originale, si riferiva all'esame di bambini.

Per l'esame di adolescenti e adulti, le voci aggiunte all'anagrafica sono prese dal foglio di risposta del questionario TALEIA-

400 A (Boncori e Di Marco, 2007).

La classificazione delle professioni in livelli è desunta dall'ISTAT. Nel sito istituzionale dell'ISTAT si può trovare l'elenco particolareggiato di tutte le professioni incluse in ciascuna delle nove categorie.

Tutte le parti aggiunte nel 2021 sono in corsivo.]

RAPPORTO PSICOLOGICO – I PARTE: DATI ANAGRAFICI E ANAMNESTICI

Dati generali:	Date degli esami.....
Nome.....	Genere.....
Indirizzo.....	Tel.....
Data di nascita.....	Età.....
Destinatario del Report	
Motivazione dichiarata per ricevere la valutazione	

Stato civile:	Convivente	Coniugata/o	Separata/o	Divorziata/o	Vedova/o	Celibe/Nubile					
Zona di nascita:	Italia: Nord	Centro	Sud	Isole	Estero: Europa	Medio Or	Nord America	Sud Am	Africa	Asia	Altro:..
Zona di residenza abituale:	Italia: Nord	Centro	Sud	Isole	Estero: Europa	Medio Or	Nord America	Sud Am		Asia	Altro:..
Titolo di studio:	Elementari	Medie	Sec. 2° grado/Laurea 3nn	Laurea >3nnale e oltre	Studi dopo laurea						
Lingua "materna"	Italiano	Francese	Tedesco	Ladino	Romeno	Altre					

Familiari conviventi.....

Familiari di 1° grado non conviventi

Livello socio-economico

Tipo studi scuola second. 2° grado	Scien- tifico	Linguisti- co	Arte	Scienze Umane	ITC	ITIS	Musica	Albergo	Grafica, Est.	Form.Pr. Inf. e tecn.	Operai Amb	Altro
---	------------------	------------------	------	------------------	-----	------	--------	---------	------------------	--------------------------	---------------	-------

Tipo studi universitari:

Agr.Alim. Zool	Arch.Urb. Ing.ed	Biologia	Chimica	DAMS
Econ. Ammin	Giurispr. Sc.Pol	Ing.elett. telem.inf.	Ing.gest. medica	Ing.ind. mecc
Lettere, Beni cult	Altro	Lingue	Matem. Fisica	Medicina. Odont.Vet
Navigazione	Psicologia	Sc.Educ. Sc.Form	Sc.fisiche e motorie	Sc.terr.amb. geogr
Sociol. Sc.Comun	Statistica	Turismo		

Professione:

I. Dirigente / Imprenditore / Doc Univ /Magistr	II. Professionista / artista	III. Tecn / Ins/ Uffic. Pol Forze Arm
IV Impiegati	V.Commerc.Istruttore.ServPers	VI, Artigiano. Oper.spec.Agricoltore
VII.Operai semiqualeficati	VIII. Profess non qualif	IX. Forze armate

Anni totali di scolarità da 6 anni ad oggi: ____

RAPPORTO PSICOLOGICO – II PARTE: DATI RACCOLTI CON IL COLLOQUIO INIZIALE E GLI ESAMI STRUMENTALI

Motivo per cui è venuto (interpretazione dello psicologo)

Fattori di sfondo (storia personale)

Eventi significativi negli ultimi tre mesi

Crescita e fattori evolutivi

Procedure d'esame (inclusi i test, in ordine di somministrazione)

1.....

2.....

3.....

...

Reazione alla valutazione psicologica (incluso il comportamento del Sogg. durante i test e i colloqui)

.....

Risultati della valutazione (2021: strumento per strumento, con singoli report allegati)

[2021. Premessa, dal DSM-5. In base agli accertamenti allegati si è in grado di escludere che le alterazioni psicologiche rilevate siano::
 effetti fisiologici di una o più sostanza (per es. una droga, un farmaco) ☐
 una condizione medica (per es. ipertiroidismo, disturbi cardiopolmonari) ☐

Fattori intellettivi generali (QI e simili) Gc Gf Ragionamento spaziale Problem solving Altro:
 Fattori neuropsicologici
 Fattori di profitto (test oggettivi somministrati dallo psicologo distinti dalle valutazioni dell'insegnante)

Fattori di personalità (riferiti al modello teorico di Blau – nota di LB), schematizzati in:

Attività interpersonali / Prime identificazioni

Struttura dell'ansietà / Difese e esternalizzazioni

[2021. Se il riferimento è a un modello di personalità diverso da quello di Blau, si usano indicatori diversi, attinenti al modello teorico usato.

Esempi:

Modello	Area 1	Area 1	Area 1	Area 1	Area 1	Controllo 1	Controllo 2	Controllo 3
Cattell	Estroversione	Ansietà	Durezza	Indipendenza	Autocontrollo	L	F	
Big Five	Energia	Amicalità	Coscienziosità	Stabilità emotiva	Apertura mentale	L		
DSM	I asse	II asse A	II asse B	II asse C		L	F	Inc

RAPPORTO PSICOLOGICO – III PARTE: SINTESI E RACCOMANDAZIONI

Sintesi

[2021. In un report clinico le informazioni sui metodi usati e sui risultati ottenuti contribuiscono a incrementare la fiducia che il cliente o il paziente deve avere nel clinico. La sintesi invece riguarda l'avvio di un "contratto" e implica la soluzione di problemi in collaborazione (Korchin, 1977, pp. 288-290), anche se l'intervento non verrà svolto da chi scrive il report. Deve quindi contenere conferma o disconferma o messa a punto dei problemi segnalati, ma anche segnalazione dei "punti di forza" della persona e delle linee d'azione che sembrano più promettenti per la soluzione dei problemi.]

Raccomandazioni [2021. Nel caso di bambini o di adulti non capaci di autogestirsi] separatamente per ciascun membro della famiglia, più raccomandazioni generali, che includono consigli per eventuali esami medici, attività del tempo libero, ecc.

Nei rapporti finalizzati all'orientamento scolastico e professionale «diagnostico» (cfr. Boncori, 1985b; Viglietti, 1983; Polacek, 1983; Polacek et alii 1985, 1987) vengono abitualmente esaminati i punti seguenti:

- quadro degli interessi e dei valori
- intelligenza e attitudini
- possibilità concrete in rapporto all'ambiente (punto particolarmente importante nell'orientamento dei portatori di handicap: cfr. Salmon, 1970),

e, a seconda dell'impostazione teorica seguita:

- problematiche d'interazione familiare e con i pari
- strutture della personalità
- dinamiche della personalità.

I rapporti di orientamento «formativo»,

[2021. I test di orientamento «formativo» - Focus (Boncori e Di Marco, 2012, 2019; 1, 2, 3... Futuro! (Soresi e Nota, 2013); ProSpera (Nota e Soresi, 2013) sono ormai tutti dotati di un software dedicato, che produce automaticamente i report. Possono articolarsi su diverse voci, che sono correlate con il modello formativo proposto e sono specificate anche nella descrizione del test presente nel catalogo dell'Editore].

Un rapporto psicologico può variare in **lunghezza** dalle dieci righe a una pagina dattiloscritta. [2021.. I report automatici sono diversi per lunghezza e non sempre i report più lunghi sono quelli che danno maggiori informazioni. I report sull'orientamento sono di solito più lunghi dei report clinici, perché progettati per essere consegnati anche senza colloquio. Il

Manuale dovrebbe offrire dati sulla loro comprensibilità per i soggetti e sulla loro validità.]

Identificare le aree di contenuto in cui si articolerà il rapporto è molto impegnativo. Basti pensare a quel che è avvenuto per la diagnosi psichiatrica: probabilmente, una delle ragioni che hanno consentito il superamento dell'estrema impopolarità della diagnosi (fra il 1940 e il 1960) è stato il quasi completo rinnovamento dei "sistemi diagnostici", divenuti molto più complessi e articolati (cfr. Robins e Helzer, 1986).

c) Il "tono" del rapporto

Chi scrive il RP deve porsi anche il problema della sua accettazione da parte di chi lo legge.

Numerosi studi hanno evidenziato la tendenza delle persone ad accettare come vere descrizioni della propria personalità che non hanno nessun rapporto con la realtà (sono scritte senza sapere a chi verranno attribuite), ma sono di tono positivo e generiche (cfr. panoramiche in Dickson e Kelly, 1985; Furnham e Schofield, 1986; Furnham e Varian, 1988; Snyder, Shenkel e Lowery, 1977). Questo fenomeno è stato chiamato "effetto Barnum" (cfr. Johnson et alii, 1985), dal nome dell'imbonitore di circo che faceva colpo sugli spettatori consegnando loro "profili di personalità" preconfezionati e sorprendentemente "giusti" (c'è tuttora chi si presenta in televisione a fare lo stesso "gioco", anche senza dire che è un gioco...).

Dall'insieme di questi studi (cfr. Ruzzene e Nollar, 1986) si possono enucleare alcuni **criteri di accettabilità**: (a) se le caratteristiche descritte nel rapporto sono positive, è più probabile che le persone ritengano la descrizione accurata (cfr. Snyder e Shenkel, 1976); (b) l'accettazione del RP è correlata con caratteristiche cognitive, motivazionali e di personalità: p.es. i nevrotici (nel senso dato a questo termine da Eysenck) sono più disponibili ad accettare come vere informazioni inventate, anche di tono negativo; (c) hanno più probabilità di esser considerate vere le affermazioni di tipo generale che quelle specifiche.

Anche l'**ordine in cui sono disposte informazioni "buone" e "cattive"** può influire sull'interpretazione che ne dà il lettore. Risultati di esperimenti condotti con professionisti e con studenti australiani evidenziano che le informazioni "cattive", se presentate prima di quelle "buone", influenzano negativamente l'insieme. Questo effetto è più forte tra i professionisti che fra gli studenti (cfr. Pain e Sharpley, 1989).

Sotto l'aspetto pratico, l'insieme di questi dati consiglia di non porre mai all'inizio di un RP informazioni "cattive" e di cercar di dare all'insieme un tono positivo, presentando i dati specifici negativi in un quadro generale da cui si evincano anche gli elementi "positivi" e "sani" della personalità del soggetto. Di fatto, nella lunga pratica di un servizio di orientamento universitario, abbiamo ripetutamente osservato che la presenza di informazioni non gratificanti all'inizio o al centro del RP (in particolare: informazioni riguardanti il rendimento ai test cognitivi) in molti casi intralcia notevolmente la comunicazione con il soggetto, che può irrigidirsi e distorcere le informazioni successive.

Quando il RP è rivolto a un altro professionista, bisogna fare attenzione al termine con cui ci si riferisce alla persona esaminata. In un esperimento compiuto negli USA, lo stesso colloquio videoregistrato venne mostrato a 40 psicologi clinici di vario orientamento teorico, dicendo ad alcuni che la persona intervistata si era presentata per essere assunta in un posto di lavoro e ad altri che si trattava di un paziente. Si trovò che le valutazioni date dai clinici sull'adattamento del soggetto risentivano significativamente dell'etichetta che aveva preceduto la visione del colloquio (Langer e Abelson, 1974). In altri esperimenti si è trovato che il RP veniva interpretato in modo diverso in funzione della classe sociale del soggetto, desunta dalla conoscenza della sua professione (cfr. Routh e King, 1972). In un esperimento venne detto ad alcuni infermieri psichiatrici che l'intervistata era vittima di uno stupro e ad altri che aveva un disordine della personalità, chiedendo poi di valutarne nove caratteristiche personali in riferimento a una scala Likert. Furono trovate differenze statisticamente significative nell'attribuzione di socialità, potenza, piacevolezza (Love e Jackson, 1988). È corretto quindi scegliere un termine che non induca il lettore del rapporto a interpretazioni distorte.

d) La "lettura" dei dati

L'abilità umana nel gestire e organizzare una grande quantità di informazioni è limitata. Si è accennato nel cap. I ad alcune strategie poste in atto nella "**conoscenza interpersonale ingenua**"; per quel che riguarda in particolare i processi attivati quando si devono elaborare informazioni complesse, si possono consultare alcuni recenti studi di psicologia cognitiva (p.es.: Newell e Simon, 1972; Simon, 1981; Tversky e Kahneman, 1974). In ogni caso, si tratta di strategie semplificative, che non operano su basi razionali e che quindi vanno sottoposte a controllo. In una ricerca sul modo di operare prevalente in un gruppo di consiglieri di orientamento francesi, ad esempio, si è visto che i risultati dei test vengono commentati all'utente soltanto in alcuni casi, variabili a seconda dell'operatore e delle caratteristiche d'insieme del soggetto, anche se è costante la tendenza a ignorare i risultati di livello medio (Mullet, 1983). Al di là delle casistiche particolari, spesso il processo si articola in un primo momento di ispezione dei dati, in cui viene colto qualche elemento soggettivamente considerato rilevante e in un secondo momento in cui vengono selezionate tutte e soltanto le informazioni che possono servire a sostenere la prima intuizione.

Per contrastare l'effetto di queste tendenze irrazionali bisogna cercar di **definire razionalmente i rapporti tra le variabili elencate nel "profilo" e i rapporti tra variabili e criterio**, che nel caso concreto può essere il quesito a cui il soggetto (o comunque il richiedente) aspetta risposta.

La definizione dei rapporti tra le variabili richiede una buona conoscenza dei costrutti psicologici misurati, conoscenza che non è desumibile soltanto dalla parola-etichetta. Si veda a questo proposito quanto detto in questa parte del Manuale riguardo alla scelta dei test e all'interpretazione dei risultati.

La definizione razionale dei rapporti tra variabili e criterio esige una buona conoscenza degli studi sulla validità predittiva dei test. Ad esempio, nella stesura di un rapporto da usare per l'orientamento scolastico in terza media, va dato più valore predittivo ai risultati scolastici correnti che non ai risultati di test d'intelligenza generale "culture fair", mentre nella stesura di un rapporto da usare per l'orientamento universitario di uno studente che, provenendo dal liceo classico, vorrebbe iscriversi alla facoltà di Ingegneria, i risultati di un test "culture fair" hanno un ruolo predittivo più importante.

Quando alcuni dati sembrano in conflitto tra loro, si deve anzitutto accertare se la discordanza è reale o apparente. Questo comporta da un lato un riesame dei costrutti, almeno sui Manuali dei test, e dall'altro una più attenta considerazione dei punteggi: si è tenuto conto dell'"errore standard" di ciascuna delle due misure? (se l'errore standard è ampio, molte differenze sono solo apparenti); si è tenuto conto che in alcuni test sono interpretabili solo i punteggi alti? (p.es.: nel MMPI); si è tenuto conto che in alcune situazioni ha più senso interpretare i punteggi bassi che quelli alti? (p.es.: un ragazzo con punteggio molto basso ai test di abilità spaziale non va incoraggiato a studiare architettura, ma non è affatto detto che un ragazzo con punteggi alti in abilità spaziale riesca un buon architetto). Se l'incongruenza fra due punteggi appare non casuale, va controllata l'esattezza dei punteggi discordanti, ripercorrendo se necessario tutti i "passaggi" dell'attribuzione: verbale di somministrazione, correzione, trascrizione del punteggio grezzo, trasformazione in punto standard, trascrizione del punto standard. Se la discordanza permane ed è lieve, se ne può cercare la ragione nei rapporti con altre misure presenti nel profilo. Se si tratta invece di una grossa discordanza, è necessario ricorrere a informazioni desunte con mezzi diversi dai test e, in alcuni casi, ripetere in tutto o in parte l'esame testologico.

In ogni caso, a conclusione del rapporto è bene segnalarne i limiti che potrebbero non essere ovvii per il lettore: p.es., il fatto che i test sono un "campione" di comportamento, che per effetto del caso potrebbe essere stato raccolto in un momento poco adatto a far esprimere al soggetto le sue "reali" capacità e caratteristiche personali.

12. La comunicazione dei risultati al soggetto

a) Il linguaggio del rapporto psicologico

Il RP dovrebbe essere scritto con la massima chiarezza e la massima sintesi.

Per quel che riguarda la **terminologia psicologica e psichiatrica**, si è già detto che i termini tecnici usati per designare le variabili psicologiche e le categorie nosografiche sono lontani dall'essere standardizzati (cfr. introduzione alla parte IV) e sono interpretati in modo molto vario anche dai professionisti. Ad esempio, sottoponendo ad analisi delle componenti principali le opinioni sulla schizofrenia espresse da 183 psicologi clinici,

psichiatri e assistenti sociali psichiatrici, sono stati evidenziati otto fattori, che rappresentano punti di vista molto differenziati riguardo alla fenomenologia, all'etiologia e alla prognosi della schizofrenia (Fitzgibbons e Shearn, 1972). Per rispettare una necessaria esigenza di chiarezza, nei rapporti scritti per i professionisti va citata chiaramente (inclusi riferimenti bibliografici) la nosografia o la teoria della personalità a cui ci si riferisce: p.es. DSM-III o DSM-III-R, fattori definiti da Eysenck, fattori definiti da Cattell, teoria psicodinamica di McClelland, di Bellak, ecc.. Quando i RP sono scritti per gl'interessati o le loro famiglie vanno ideate definizioni standard, nella lingua di tutti i giorni, che devono essere utilizzate senza variazioni: può darsi che lo stile ne soffra, ma è la chiarezza che non può essere sacrificata.

La **sintassi** dev'essere essenziale: frasi brevi e dalla costruzione diretta (soggetto - verbo - complementi), con poche proposizioni subordinate.

Le affermazioni devono avere un **significato chiaro e univoco e fare riferimento ai fatti**: anche nei rapporti rivolti ad altri professionisti, il riferimento a fatti concreti è preferibile all'uso di termini astratti o tecnici, che allo stato attuale sono lontani dall'essere univoci. Ad esempio, se un soggetto ottiene risultati scarsi a test che richiedono rapidità (p.es.: fluidità verbale, abilità numerica) e mostra anche segni di forte Super Io e di "cattivo umore", l'insieme di queste informazioni dettagliate può essere più utile per l'interlocutore che non l'etichetta di "depresso". Questo accorgimento è importante anche per l'area cognitiva, specialmente quando il risultato dei test di livello non è brillante.

Il lettore deve anche comprendere la **razionalità del processo** attraverso il quale si arriva a certe conclusioni, per evitare la tentazione di attribuire allo psicologo il ruolo di mago o di prestigiatore e ai test il ruolo di sfera di cristallo o di tarocchi.

Infine, va sempre tenuto presente che il rapporto psicologico è un "testo": come tale, la sua stesura deve seguire le stesse leggi a cui obbedisce un qualsiasi narratore quando scrive una storia, vera o falsa che sia. Le informazioni non posso essere elencate a caso, ma devono avere collegamenti tra loro, seguire una "trama", comporsi in un "intreccio". **Il testo deve avere una sua unitarietà**, "tenersi insieme" (cfr. Sarbin, 1986).

b) Le modalità della comunicazione

La centralità dei soggetti come persone, diventata sempre più un punto irrinunciabile della prassi psicologica contemporanea, va tenuta presente anche nelle modalità di comunicazione dei risultati.

Attualmente è sempre più raro che i risultati dei test vengano comunicati secondo il modello "scolastico", in cui i risultati sono di solito comprensibili ma indiscutibili, oppure secondo il modello "medico", che utilizza anche una terminologia tecnica (si pensi ai risultati delle analisi biologiche) tale da impedire alla grande maggioranza dei pazienti di prendere parte attiva alla comunicazione.

Per effetto anche di movimenti come lo statunitense *truth in testing* ("la verità nei test") si è andato progressivamente affermando **un modello di comunicazione "riferito alla persona"**, articolato nel modo che segue (cfr. Bradley e Snowman, 1981):

- parlare con chi ha sostenuto il test per stabilire se l'esperienza è stata valida;
- consentire al soggetto (se questi lo chiede) di rivedere sia gli stimoli del test sia le risposte date;
- identificare punti forti e punti deboli che sono stati evidenziati nel rendimento ai test e discutere sul modo di trattarli;
- considerare in qual modo le informazioni acquisite possano essere utilizzate nel prendere decisioni;
- considerare il rendimento in riferimento a criteri.

L'operatore che comunica i risultati al soggetto deve conoscere bene i test che interpreta: deve comprendere bene le variabili misurate ed essere anche in grado di spiegare al soggetto l'attinenza fra le risposte a singoli items e i risultati del test.

I risultati del test vanno comunicati agl'interessati con linguaggio non tecnico.

Il soggetto dovrebbe uscire dalla seduta di comunicazione dei risultati con una migliore conoscenza di sé.

13. Rapporti psicologici generati dal computer

La diffusione dei personal computers e la progressiva semplificazione del loro uso ha modificato profondamente il trattamento dei test anche da parte dello psicologo applicato. E' ormai ovvio che la correzione dei questionari di personalità, di solito molto laboriosa, vada fatta con il computer, così com'è ovvio usare il computer per standardizzare i punteggi e costituire una propria "banca dati" su cui basare delle "norme locali". Per ora in Italia è ancora poco sviluppato l'uso del computer per decidere sulla "similarità dei profili" (p.es.: in campo di orientamento, assumendo come criterio il profilo di personalità di professionisti soddisfatti; in campo clinico, assumendo come criterio i profili di soggetti con patologie chiaramente diagnosticate; cfr. Cattell, Eber, Tatsuoka, 1970) ed, eventualmente, sulla loro "aberranza" (metodi statistici sostitutivi delle "scale di controllo": cfr. paragrafo 9), ma probabilmente solo perché mancano dati italiani di riferimento.

Comincia invece a diffondersi anche da noi l'uso dei "rapporti psicologici automatizzati", a cui si è accennato nella IV parte in riferimento a singoli test. Qui vengono dati un quadro generale del problema e qualche indicazione pratica per decidere sulla validità dei sistemi e per utilizzare validamente quanto è disponibile.

a) Tipologia dei programmi che originano RP

[2021. Questo paragrafo ha un interesse prevalentemente storico. La potenza dei computer e dei linguaggi di programmazione attuale rende comuni report automatici molto simili al linguaggio naturale. Il principale controllo critico va fatto sulla rispondenza tra affermazioni interpretative e dati.]

I primi rapporti psicologici scritti dal calcolatore fra il 1960 e il 1970 erano molto scarni, elaborati con poche regole, attingendo a "banche di affermazioni" limitate (cfr. Klett e Pumroy, 1971). Riguardavano prevalentemente il MMPI (alcuni di questi sistemi sono descritti nel cap. XIV) e si riferivano all'ambiente medico, probabilmente perché l'esigenza di esaminare grandi numeri di soggetti si poneva soprattutto negli ospedali (rassegne sui primi lavori in campo di automatizzazione dei RP si trovano in: Hedlund e Hickman, 1977; Sidowski, Johnson, Williams, 1980). Attualmente, sono disponibili programmi che scrivono un RP automatizzato per tutti i principali questionari autovalutativi (MMPI, CPI, 16 PF e altri: cfr. Buros, 1978; Conoley e Kramer, 1989; Lanyon e Goodstein, 1982, p. 226; J.V. Mitchell, 1985), per "batterie" di uso frequente, anche localmente (cfr. French, 1986) e per alcuni test proiettivi (Rorschach, Holtzman: cfr. Klett e Pumroy, 1971), con riferimento a una notevole varietà di problematiche: p.es. per il 16 PF (edizione statunitense) sono stati predisposti distinti programmi per scrivere un "Profilo di sviluppo della carriera personale", un "Rapporto per il counseling matrimoniale", una "Rassegna del profilo dei dirigenti", un "Piano individualizzato per la gestione dello stress", ecc. (Krug, 1986). I risultati di un importante questionario psicopatologico, il Millon Clinical Multiaxial Inventory, sono ottenibili solo da un programma computerizzato (cfr. cap. XIV). Negli USA cominciano ad essere pubblicati programmi (utilizzati su personal computers sia del tipo IBM sia del tipo Apple) che analizzano i punteggi di test cognitivi come il WISC-R e stampano un RP interpretativo. Alcuni di questi programmi sono criticati abbastanza severamente (p.es.: Vance, Paparella, Piazza, 1983; cfr. Conoley e Kramer, 1989), altri invece sono più apprezzati dai recensori (p.es.: Nicholson, 1986; cfr. Sibley, 1989).

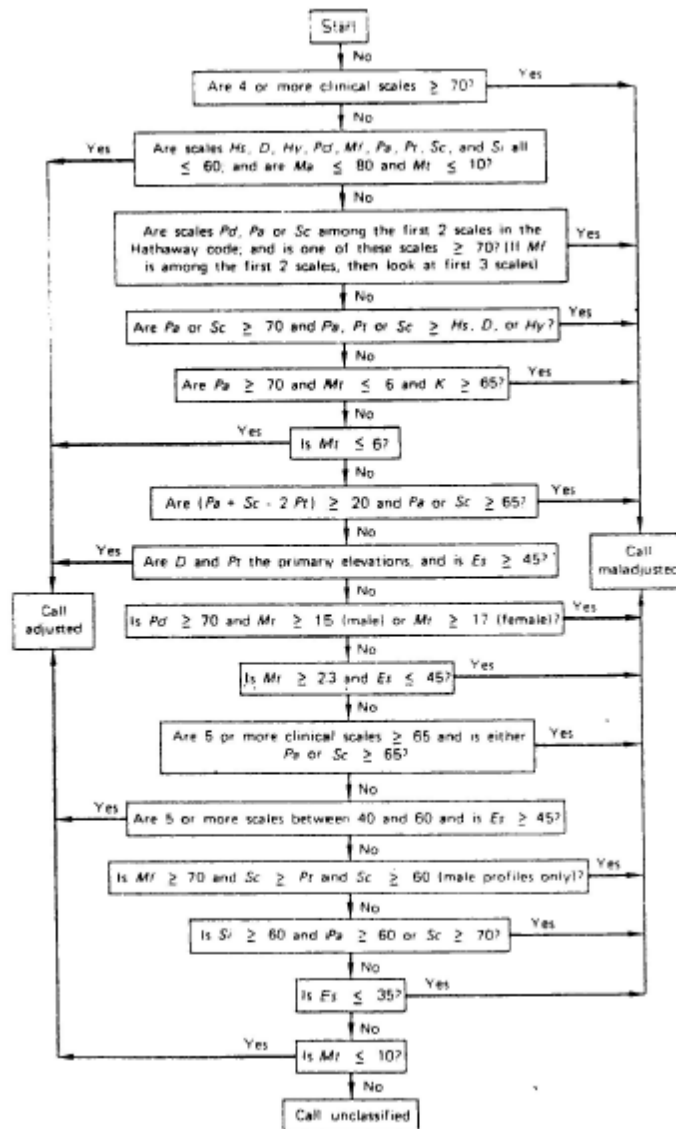
In riferimento ai *test neuropsicologici* è tuttora controversa la possibilità di ottenere rapporti automatizzati che siano qualitativamente migliori dei RP compilati dai clinici, finché non sia meglio chiarito l'influsso di alcuni fattori che incidono significativamente sui risultati di questi test (cfr. Adams e Heaton, 1985; Kleinmuntz, 1987; Heaton e Adams, 1987). Un caso a sé è costituito dallo sviluppo dei sistemi automatizzati per l'orientamento, che in gran parte vengono gestiti dai soggetti stessi (cfr. L. Boncori, 1987). Indipendentemente dall'uso dei test, sono stati predisposti numerosi programmi automatizzati per la diagnosi medica e psichiatrica (cfr. Quadrio e Rizzardi Cremaschi, 1980; Lanyon e Goodstein, 1982).

I programmi esistenti possono essere classificati in (Roid, 1986): (a) **descrittivi** (viene menzionata la scala e data una sommaria classificazione dei punteggi, p.es. con frasi quali "sopra la media", "possibili problemi organici", ecc.); (b) **modellati sull'esperienza clinica**; (c) **clinico-attuariali**. Solo di alcuni programmi per la stesura di RP sono noti, almeno nelle linee generali, gli elementi essenziali che consentono di avere indicazioni almeno sulla loro "validità di contenuto", e cioè: (1) le fonti e la metodologia con cui è stata raccolta la "banca dati" delle affermazioni interpretative (cioè le singole frasi di commento, che vengono selezionate e assemblate dal programma); (2) i principi seguiti nel predisporre le "regole" che costituiscono la struttura del programma; (3) i criteri seguiti per rendere più leggibile la redazione del RP ed eliminarne ridondanze e incongruenze.

I sistemi più accreditati si basano su: a) banche dati messe insieme raccogliendo valutazioni date da più clinici esperti, in riferimento a una casistica molto ampia; b) una simulazione del processo interpretativo seguito da un buon clinico (cfr. fig. XX.14). Maggiori garanzie di validità si avrebbero se i programmi fossero redatti tenendo

FIG. XX.14

Stralcio di regole decisionali assegnate a un programma per l'interpretazione computerizzata del MMPI (da Kleinmuntz, 1967, p. 365)



presenti "regole" già validate nell'ambito delle ricerche attuariali sulla validità rispetto a criteri, ma questo approccio non sembra di fatto seguito (cfr. Lanyon e Goodstein, 1982, p. 225).

La sezione del diagramma di flusso riprodotta nella fig. XX.14 riguarda l'identificazione del disadattamento in studenti universitari. I numerosi confronti elencati nell'analisi sono stati individuati analizzando il lavoro di buoni clinici (p.es. con il metodo della "riflessione parlata"): è intuitivo che il computer, certo incapace di "idea-re" quali sono i confronti da fare e le conclusioni da prendere in presenza di certe condizioni, una volta "istruito" farà tutti i confronti con una sistematicità impossibile a qualsiasi operatore umano e trarrà decisioni secondo le regole razionalmente stabilite, senza farsi influenzare da "fattori di alone" o da pregiudizi. Questa ipotesi razionale è stata anche comprovata sperimentalmente. In un esperimento statunitense venne chiesto a dieci clinici esperti di valutare 126 protocolli di test, riguardanti studenti universitari. I clinici dovevano disporre i protocolli in graduatoria, dal ragazzo più "adattato" a quello più "disadattato". Come criterio esterno di disadattamento vennero prese le valutazioni date da compagni di scuola e l'eventuale condizione di essere in cura

psicoterapica. Le valutazioni del clinico che meglio prediceva i criteri esterni vennero prese come modello per costruire un programma per la stesura di un RP. Tutti e 126 i protocolli vennero nuovamente valutati dal computer e si osservò che queste valutazioni erano meglio correlate con il criterio che non le valutazioni del clinico in base alle quali era stato costruito il programma (Kleinmuntz, 1963), ovviamente perché il computer aveva applicato le stesse regole, ma con maggior sistematicità.

b) Alcuni programmi italiani che originano RP

[Nota dell'Autrice – 2021. *Le informazioni che seguono, scritte in corpo più piccolo, si riferiscono a prima del 2000 e i software menzionati non sono più accessibili, anche perché programmati su sistemi operativi non più in uso. A uno sguardo rapido su internet, senza accedere a Banche Dati specifiche, l'A. ha trovato centinaia di prodotti che mirano a organizzare in cartelle le informazioni raccolte dallo psicologo, ma nessuno che si proponga di costruire un report con la sintesi dei risultati di più strumenti. Si sono invece moltiplicati i software per lo scoring di singoli test che producono anche un report per quel test o una scheda d'intervista individualizzata che aiuta a formulare una diagnosi (Taleia). L'A. è disponibile a dare una valutazione critica di strumenti che le vengono segnalati.]*

Nella parte IV sono stati menzionati alcuni programmi italiani che scrivono RP in commento a singoli test (si vedano, in particolare, i capp. XIV, XV, XVIII). Qui accenniamo a programmi che commentano i risultati di un insieme di test.

Il **package GPQ-88, "Gestione prove e questionari"** (Dama, De Iorio, Poggiali, 1988) viene descritto come un "sistema modulare aperto per il trattamento e l'esame personalizzato di prove e questionari. Opera su un elaboratore MS-DOS che abbia almeno 256 Kb di memoria, due drives da 360 Kb oppure disco rigido da 10 Mb. Si compone dei quattro "moduli" qui appresso descritti.

Modulo Approntamento/revisione prove: consente di allestire una nuova prova, con un massimo di 600 quesiti e 34 dimensioni e un massimo di 5 alternative di risposta per ogni domanda;

Modulo Allestimento/revisione archivi: per un numero di questionari limitato solo dalla memoria di massa del computer, si possono costituire singoli archivi, ciascuno dei quali contiene il testo delle domande, la denominazione delle scale, media e deviazione standard per ogni scala (calcolate su un campione normativo), eventuali "ponderazioni" per cui moltiplicare il punteggio di alcune scale; si possono standardizzare i punteggi trasformandoli in punti zeta, T, C, stanine o altro (da definire con una formula); si può stampare un profilo grafico dei punteggi; si possono immettere criteri di valutazione dei punteggi (p.es. per escludere chi, nel MMPI, ha un punteggio L superiore a un livello dato).

Modulo Immissione, archiviazione e riesame questionari: immette le risposte a ciascun quesito; questo può esser fatto nel contesto di una somministrazione computerizzata (il soggetto legge il quesito sul monitor e immette direttamente la propria risposta) oppure, quando i soggetti abbiano compilato fogli di risposta, digitando le loro risposte; le risposte sono archiviate in un file che contiene anche i dati anagrafici essenziali del soggetto, protetto con un'apposita password per assicurare la riservatezza dei dati. Al termine dell'immissione si può ottenere immediatamente il punteggio grezzo nelle singole scale. E' possibile ottenere la visualizzazione o la stampa delle risposte date da ogni soggetto all'intero questionario o a parte di esso (singole scale, "items critici").

Modulo Scoring e valutazione dei questionari: consente la correzione automatica di un questionario; visualizza o stampa il punteggio grezzo, "corretto", standardizzato; disegna un grafico dei punteggi. Stampa un breve commento ai dati, caratterizzato (rispetto ad altri sistemi) dal fatto di essere preceduto dalla stampa della formula che risulta soddisfatta per quel criterio, con i valori di soglia e il punteggio ottenuto dal soggetto. Il commento è rivolto a professionisti.

Il **programma ORUV** (Boncori e Solfaroli Camillocci, 1985) genera un commento ai test e ai questionari della batteria usata dal servizio di orientamento universitario del Dipartimento di Psicologia dell'Università di Roma "La Sapienza" (cfr. cap. XI), che include misure di abilità generale e specifica (Otis Quick, D 48, GATB), questionari di personalità (una forma alternativa del 16 PF di Cattell, il PNP di Pichot), un questionario "ad hoc", prevalentemente anamnestico, che esplora interessi e aspirazioni del soggetto e chiede informazioni sulle caratteristiche familiari e ambientali (L. Boncori, 1985b). Il commento ai dati, rivolto ai soggetti, occupa mediamente mezza pagina dattiloscritta ed è finalizzato a interpretare le informazioni raccolte su 72 variabili) in funzione degli interessi accademici manifestati dallo studente. Il programma si basa sulla descrizione analitica del procedimento seguito dagli psicologi che collaborano al servizio. Data la destinazione del messaggio, se i punteggi evidenziano problemi personali di tipo emotivo si consiglia un colloquio con lo psicologo, senza esplicitare le problematiche. Questo RP viene sempre consegnato nell'ambito di un colloquio psicologico semi-strutturato, simile a quello descritto nel paragrafo precedente.

Il programma non verrà commercializzato finché la sua validazione non abbia raggiunto un livello considerato sufficientemente sicuro.

c) Validità dei programmi interpretativi automatizzati

È ancora vero oggi quel che è stato scritto nel 1982 [confermo al 2021: anzi, la situazione è peggiorata] riguardo ai controlli sulla validità dei programmi interpretativi automatizzati: "per la maggior parte dei programmi non ci sono stati tentativi seri di dimostrarne la validità" (Lanyon e Goodstein, 1982, p. 228). Su questa situazione influisce il costo elevato dell'allestimento iniziale, che induce gli Autori a commercializzare prematuramente il loro prodotto per ammortizzare le spese. Il tentativo dell'American Psychological Association di regolamentare gli aspetti etici dell'uso dei RP automatizzati (APA, 1967) è attualmente considerato inadeguato (cfr. Lanyon e Goodstein, 1982, p. 229). Recentemente sono apparsi contributi per la definizione di metodologie appropriate alla valutazione della validità di programmi interpretativi computerizzati (cfr. McCullough, 1991).

Come qualsiasi altro strumento diagnostico, i RP automatizzati dovrebbero essere validati almeno rispetto al contenuto e rispetto a criteri diagnostici o predittivi (per i rapporti finalizzati all'orientamento, cfr. L. Boncori, 1989).

La **validità del contenuto** dovrebbe essere verificata almeno sotto gli aspetti seguenti: [il peggioramento è dovuto al fatto che non solo non vengono pubblicate le dimostrazioni, ma neanche descritti i procedimenti e i criteri. Quindi, i software migliori sono quelli in cui c'è almeno una descrizione che aiuta a capire i limiti del prodotto.]

a) documentare quali sono **le fonti delle affermazioni** contenute nella "banca dati" e raccogliere valutazioni di un gruppo di esperti sull'adeguatezza e la completezza di questo corpus;

b) ottenere valutazioni di un gruppo di esperti sull'**adeguatezza dei criteri** seguiti nell'assegnare le regole di selezione delle affermazioni;

c) dimostrare che **le affermazioni contenute nella banca dati sono adatte a descrivere uno o più tipi di popolazione**, precisamente caratterizzati;

d) dimostrare che le **regole decisionali** incluse nel programma sono adatte a descrivere uno o più tipi di popolazione, precisamente caratterizzati;

e) dimostrare che i RP prodotti **vengono compresi pienamente e senza fraintendimenti** da un campione della popolazione a cui sono rivolti.

Per quel che riguarda la dimostrazione della **validità rispetto a criteri** (cfr. Lanyon e Goodstein, 1982, p. 228-231; Roid, 1986) vanno presi in considerazione almeno i seguenti problemi:

a) il **criterio diagnostico** non può essere costituito chiedendo semplicemente ad uno o più esperti di valutare se sono adeguate le affermazioni contenute in un qualsiasi gruppo di RP automatizzati "reali"; per evitare distorsioni di giudizio analoghe a quelle introdotte dall'"effetto Barnum", bisogna che ogni RP contenga sia una proporzione di affermazioni "reali", sia una proporzione di affermazioni "sbagliate" o inutili, per poter valutare in qual proporzione gli esperti identificano correttamente i diversi tipi di affermazione; comunque, vanno posti in atto accorgimenti che consentano di controllare le sempre esistenti "distorsioni personali" dei giudici;

b) le **caratteristiche dei soggetti** ai quali si riferiscono i RP automatizzati devono essere accuratamente descritte e si deve evitare la generalizzazione delle misure di validità da un tipo di popolazione all'altro;

c) la validazione deve mirare a **utilizzazioni specifiche**, ben precisate (p.es.: diagnosi psicopatologica, orientamento, counseling matrimoniale, ecc.);

d) le **regole che si riferiscono a punteggi di eliminazione ("cut-off")** devono essere validate in riferimento alla proporzione in cui un dato fenomeno è presente nel tipo di popolazione su cui si valida il RP (p.es.: se il programma si riferisce a una selezione professionale, prima di decidere quale proporzione di soggetti va dichiarata "inadatta a svolgere il compito X" bisogna raccogliere dati empirici che dicano qual è la percentuale di soggetti o qual è il livello di punteggio a cui, in quella data popolazione, è stata osservata incapacità di eseguire il compito).

Queste esigenze minime non sono impossibili da soddisfare. E' quindi da augurarsi di avere presto sul mercato

programmi che, uscendo dalla fase sperimentale, diano all'utente almeno una parte delle garanzie richieste.

SOMMARIO

In questo ultimo capitolo si è cercato di raccogliere alcune indicazioni utili per chi deve interpretare i risultati di un insieme di test. Per molti aspetti ci si è limitati a riassumere, in forma di suggerimenti pratici, le conclusioni di analisi critiche presentate nelle precedenti parti del Manuale, integrandole con suggerimenti tratti dall'esperienza professionale in vari settori applicativi. Una sintesi limitata alla prevenzione di alcuni gravi abusi interpretativi è esposta nella tab. XX.5.

TAB. XX.5

Pro-memoria per l'interpretazione dei test: evitare i principali abusi

I. Punteggi:

- **non utilizzare solo i punteggi dei test** per prendere decisioni la cui validità e attendibilità nessun test, da solo, può garantire (p.es.: promozioni scolastiche, assunzioni, diagnosi cliniche: cfr. Koppitz, 1982);
- **non utilizzare le medie dei punteggi dei test** per segnalare scuole "buone" e "cattive", insegnanti "buoni" e "cattivi", terapeuti efficaci e inefficaci, ignorando il livello iniziale degli allievi o dei pazienti (cfr. Gardner, 1982).

II. Attendibilità dei punteggi:

- **non codificare i risultati del test** senza tener conto dei problemi di oggettività segnalati dalle ricerche: p.es., per i test "a risposta aperta" dovrebbero sempre essere usate griglie precise di riferimento e per molti di essi va data una siglatura indipendente da almeno due operatori;
- quando i punteggi vengono usati per classificare i soggetti, far sempre riferimento all'**errore standard della misura**, che di solito segnala margini d'errore abbastanza ampi e riduce il rischio di "etichettamenti";
- **non considerare qualsiasi punteggio misura di una caratteristica stabile della personalità**: è noto che alcuni test (inclusi quelli proiettivi) sono sensibili a stati transitori e ad effetti situazionali (cfr. Anzieu, 1967; Ziskin e Faust, 1988).

III. Valutare criticamente i dati contenuti nel Manuale:

- **evitare interpretazioni che non si basano su verifiche preliminari della validità del test**, anche se queste vengono proposte nel Manuale (cfr. Anzieu, 1965, 1967);
- **evitare interpretazioni basate su verifiche della validità inappropriate**, perché compiute decenni prima, o su campioni numericamente inadeguati, atipici (p.es. solo maschi, solo studenti, solo utenti di strutture pubbliche...), distanti geograficamente o culturalmente (cfr. Gardner, 1982);
- **distinguere se l'interpretazione proposta nel Manuale è stata validata nel suo insieme oppure soltanto sotto alcuni aspetti**: nell'interpretazione, limitarsi a considerare i soli aspetti la cui validità è stata provata;
- **utilizzare le norme statistiche contenute nel Manuale solo dopo aver verificato che sono adeguate alla popolazione su cui si opera**: p.es. che sono calcolate su soggetti italiani, residenti in un tipo di zona (città, paese, campagna...; Nord, Centro, Sud, Isole ...) affine a quello in cui risiedono i propri soggetti, caratterizzati per sesso e status socioculturale in modo analogo ai propri soggetti (il che include la necessità che siano recenti: dieci anni di anzianità sono già rischiosi per le norme); il campione su cui sono calcolate le norme dovrebbe essere estratto a caso (o con un procedimento chiaramente descritto, assimilabile al caso) e non sia inferiore a $N = 300$.

IV. Salvaguardare la validità delle interpretazioni con la professionalità dell'interprete:

- **non interpretare il significato psicologico della variabile sommariamente, in base a stereotipi** (p.es., un

consigliere improvvisato potrebbe affermare: "il test d'interesse dice in quale settore lo studente può riuscire meglio", mentre un alto livello di interesse – da solo - non può essere considerato come buon predittore della riuscita); il significato del costrutto misurato da un test va individuato in riferimento alla specifica teoria o sistema nosografico;

- **nell'interpretazione va precisata la nosologia di riferimento** (2021: *ICD 10, ICD 11, DSM IV, DSM-5...*) e vanno evitate le categorie nosologiche imprecise (Rossel, Husain, Merceron, 1985-6);

- **i test brevi, ideati per uno screening preliminare, vanno integrati da esami successivi** che verifichino e approfondiscano i risultati;

- **i risultati vanno utilizzati integralmente**: soprattutto in test proiettivi, alcuni operatori maldestri concentrano l'attenzione su una parte degli indicatori, trascurandone altri a loro arbitrio, oppure attribuiscono a ogni stimolo un valore simbolico che si suppone uguale per tutti i soggetti (Rossel, Husain, Merceron, 1985-6); utilizzare i dati parzialmente è un abuso;

- se i risultati dei test vengono utilizzati a scopo predittivo, **bisogna tener conto anche delle condizioni ambientali** e non misconoscere «un dato essenziale della psicologia, secondo il quale "il comportamento è certo funzione della persona ma anche dell'ambiente"» (K. Lewin, cit. da Blatt, 1968); è un abuso utilizzare i test a "scopo divinatorio, come se si trattasse di "leggere" il comportamento manifesto in una sfera di cristallo» (Rossel, Husain, Merceron, 1985-86);

- **le variabili misurate dal test (o dal gruppo di test usati) vanno commentate tenendo conto delle relazioni tra una variabile e l'altra**; è un abuso interpretarle una per una (Gardner, 1982);

- **i risultati del test vanno interpretati tenendo conto delle caratteristiche personali del soggetto (età, sesso, ecc.)** che, prese in debita considerazione, possono indurre a un'interpretazione diversa da quella standard;

- nell'interpretare i risultati, **l'operatore deve assumere le necessarie salvaguardie da dinamiche di "proiezione" di caratteristiche personali** (cfr. Anzieu, 1965, 1967), precauzioni che comportano come minimo grande accuratezza nel rispettare le norme standard di somministrazione e codifica e uno studio accurato dei dati sull'attendibilità e la validità del test.

V. Punteggi ottenuti dai soggetti in due diverse occasioni:

- è legittimo confrontare **risultati ottenuti dallo stesso soggetto in occasioni diverse** solo se si tiene conto di eventuali differenze fra i test usati nelle due occasioni (p.es.: test dalla denominazione simile possono misurare variabili diverse).

Punteggi ottenuti in due o più test, o in diverse occasioni:

- vanno interpretati tenendo conto di eventuali **differenze tra le norme**;

- **non devono essere interpretati come se differenze di qualunque entità tra due punteggi fossero statisticamente significative**, ma tenendo conto dell'errore tipico delle differenze (cfr. cap. VI).

VI. Rapporto sui risultati:

- **il rapporto dev'essere redatto in termini comprensibili** agli interessati o ai genitori (se i soggetti sono bambini);
- **per i test di profitto**: è un abuso consegnare i risultati agli alunni senza discuterne le **implicazioni in termini di miglioramento del profitto** (Gardner, 1982).

- **per i test di abilità e di personalità**: è un abuso redigere il rapporto senza tener conto della sensibilità dei soggetti e delle loro reazioni personali.

Per alcuni aspetti sono state presentate sintesi critiche della letteratura di ricerca, che non era stato possibile collocare nei capitoli precedenti e che peraltro sono direttamente attinenti alla corretta interpretazione dei test. In particolare, sono stati presentati in questo capitolo il dibattito fra predizione clinica e statistico-attuariale, la problematica delle differenze tra sessi nelle prestazioni ai test, della stabilità o meno di significato in funzione dell'età dei soggetti e delle loro condizioni fisiche, dell'influsso rispettivo dei fattori biologici e culturali, dell'interferenza tra variabili "cognitive" e "non cognitive", della falsificazione più o meno volontaria dei risultati

ai test, dei rapporti psicologici generati da computers. Di queste problematiche così ampie è stato peraltro considerato solo l'impatto che possono avere, a livello pratico, nell'interpretazione dei risultati dei test e nella stesura del "rapporto psicologico".

Questo capitolo è stato scritto, come il precedente, soprattutto per gli studenti delle scuole di specializzazione in Psicologia clinica (o scuole affini) e per gli studenti di Psicologia alle prese con i primi tentativi guidati di pratica psicologica. Nonostante la sua lettura in alcune parti non sia facilissima, questa è forse la parte del Manuale più tipicamente "introduttiva", destinata com'è ad integrarsi con gli insegnamenti psicologici applicativi e con l'attività pratica.