

Psicologia dell'Intelligenza Umana

Teorie, misure e differenze individuali

Enrico Toffalini

Dipartimento di Psicologia Generale
Università di Padova

Indice

Introduzione	1
I Capire l'intelligenza	3
1 Perché studiare l'intelligenza	4
2 Che cosa chiamiamo intelligenza	13
II Modelli e test dell'intelligenza	23
3 Il fattore generale e la positive manifold	24
4 Abilità multiple e modelli gerarchici	32
5 Dai primi test al testing di massa	46
6 Le batterie moderne di intelligenza	57
III Strumenti I · Punteggi e distribuzioni	69
7 Distribuzioni, punti z e punteggi standardizzati	70
IV Basi biologiche, ambiente e cambiamento	85
8 Geni, ambiente ed ereditabilità	86
9 Che cosa aggiunge la genetica molecolare	98
10 L'effetto Flynn e i cambiamenti storici	107

11	Cervello e intelligenza	118
12	Si può potenziare l'intelligenza?	126
V	Strumenti II • Correlazione e qualità della misura	139
13	Correlazione e differenze individuali	140
14	Quando una misura è buona	155
VI	Scuola, salute e lavoro	166
15	Intelligenza, scuola e condizioni sociali	167
16	Salute e longevità	175
17	Intelligenza e lavoro	182
VII	Intelligenza e condizioni del neurosviluppo	192
18	Dalle categorie alle dimensioni	193
19	Valutare l'intelligenza nel neurosviluppo	203
VIII	Approfondimenti: oltre g	214
20	Altre intelligenze o vecchi costrutti con un nome nuovo?	215
21	Si può spiegare la positive manifold senza un fattore generale?	220
	Riferimenti bibliografici	227

Introduzione

Lo studio scientifico dell'intelligenza in Italia non gode della stessa attenzione e maturità teorica che ha raggiunto in altri paesi. Eppure, psicologhe e psicologi si trovano quotidianamente a valutare il funzionamento intellettivo di bambini, adulti, pazienti, in ogni genere di contesto, utilizzando strumenti complessi come le scale di QI e altri test cognitivi. Farlo senza un'adeguata consapevolezza dei modelli teorici, della storia, dei limiti e dei punti di forza della misurazione rischia di ridurre la valutazione a un gesto tecnico, anziché renderla un atto scientifico e interpretativo.

Cos'è l'intelligenza? Come si misura? È innata o appresa? Cosa significa il QI? Stiamo diventando più intelligenti o meno? Esistono le intelligenze multiple? Questo libro, e il progetto più ampio da cui nasce, mira a fornire gli strumenti concettuali per capire queste domande e le loro risposte, nonché a fornire una base di competenze pratiche per chi deve utilizzare i test quotidianamente.

Come leggere questo libro

Il libro segue il filo di un corso universitario e parte dalle idee: che cosa chiamiamo intelligenza, se è una o molte, come sono nati i test, da dove vengono le differenze fra le persone, che cosa prevedono nella vita reale. Questi capitoli si leggono senza preparazione statistica: quando serve un concetto tecnico, lo spiegano in modo intuitivo.

Le basi tecniche, che pure sono la parte più utile a chi i test li usa davvero, sono raccolte in due parti a sé, segnalate in verde e chiamate **Strumenti**: la prima su distribuzioni e punteggi, la seconda su correlazione e qualità della misura. Stanno al centro del libro, nel punto in cui diventano necessarie, ma non sono un passaggio obbligato. Si possono studiare da sole, anche

prima del resto, oppure saltare, se si conoscono già o se interessano solo le parti concettuali. Dai capitoli concettuali, brevi rimandi in verde indicano dove approfondire.

Che cosa tratta

I capitoli esplorano le teorie classiche e contemporanee sull'intelligenza, la sua storia e le sue basi biologiche, la misurazione psicometrica, la genetica delle differenze individuali, i fenomeni sociali e culturali che la influenzano e i test più utilizzati in ambito clinico, scolastico e di ricerca. Un'attenzione particolare è dedicata all'interpretazione dei punteggi, alla comprensione critica dei limiti dei test e ai temi più attuali, come la neurodiversità, la plusdotazione, il funzionamento intellettivo limite e la disabilità. L'obiettivo è promuovere una riflessione informata e critica sul modo in cui la psicologia studia, misura e comunica l'intelligenza.

Parte I

Capire l'intelligenza

1 Perché studiare l'intelligenza

L'intelligenza è uno dei primi temi affrontati dalla psicologia scientifica e, oggi, uno dei più maturi. Ha prodotto metodi che usa l'intera disciplina, ha una rete fitta di collegamenti con la genetica, le neuroscienze e lo sviluppo, e si misura ogni giorno in ambulatori, scuole, aziende e tribunali. Eppure resta uno degli argomenti su cui anche chi lavora in psicologia sa meno di quanto crede.

Capita di sentire professionisti che chiedono come interpretare un QI calcolato con l'età mentale, un metodo abbandonato da decenni. Capita di sentire chiedere quale questionario si usi per misurare l'intelligenza, come se bastasse domandare alle persone quanto si sentono brillanti. Capita perfino di sentir dire che i test di QI siano un mito pseudoscientifico. Sono tre errori diversi, ma hanno un'origine comune: si usano strumenti di cui non si conoscono le basi.

Questo capitolo spiega perché vale la pena colmare la lacuna. Le ragioni sono pratiche, teoriche e storiche. E c'è un motivo meno ovvio: capire l'intelligenza significa capire come la psicologia studia le differenze fra le persone.

1.1 Test ovunque, spesso sotto un altro nome

Le batterie di intelligenza sono fra gli strumenti più usati della professione psicologica. Raramente però si presentano con il loro nome. Si parla di «profilo cognitivo» a un genitore, di «abilità di ragionamento» in una richiesta a una scuola, di «test di problem solving» in una selezione del personale. Spesso la scelta è sensata: evita di caricare la valutazione di significati che il punteggio non ha. Ma il cambio di etichetta non cambia lo strumento.

I contesti d'uso sono molti:

- nello **sviluppo**, per distinguere traiettorie tipiche e atipiche e per valutare i disturbi del neurosviluppo;
- nella **scuola**, per comprendere difficoltà di apprendimento e bisogni educativi;
- nella **clinica**, per descrivere risorse e limiti di una persona;
- nell'**invecchiamento**, per distinguere un declino atteso da uno patologico;
- nel **lavoro**, per prevedere l'adattamento a compiti complessi;
- in ambito **forense**, per perizie su imputabilità, capacità di partecipare al processo o danno neuropsicologico, dove validità e attendibilità di una misura vanno difese davanti a un giudice.

Molte batterie classificate come «cognitive» o «neuropsicologiche» sono, nei fatti, test di intelligenza. Le scale Wechsler ne sono l'esempio più noto.

Il problema è che la formazione si concentra spesso su somministrazione e scoring. Si impara a leggere le istruzioni e a convertire i punteggi grezzi, ma non a rispondere a domande più importanti. Che cosa sta misurando davvero lo strumento? Quanto è affidabile il punteggio ottenuto? Una discrepanza fra due indici è abbastanza grande da essere interpretata? Senza queste risposte l'interpretazione diventa automatica, a volte delegata alle piattaforme di scoring. Il risultato è un numero preciso fino al decimale e opaco nel significato.

1.2 Una palestra per tutta la psicologia

La seconda ragione riguarda la ricerca. Molti metodi oggi diffusi in ogni area della psicologia sono nati o si sono sviluppati per studiare le abilità cognitive. L'analisi fattoriale è stata proposta da Spearman per spiegare le correlazioni fra prestazioni scolastiche e sensoriali (Spearman, 1904). Dalla stessa tradizione discendono l'idea di **variabile latente**, i modelli di equazioni strutturali e buona parte della teoria della misura.

Chi conosce l'intelligenza capisce quindi meglio gli strumenti con cui si analizzano personalità, psicopatologia o benessere. Il percorso inverso è più faticoso: si imparano le tecniche e poi si scopre che erano state costruite per rispondere a domande precise sulle abilità mentali.

C'è anche un debito storico. La psicologia delle differenze individuali e la psicometria si sono formate attorno ai primi test cognitivi. Concetti oggi ovvi, come il punteggio standardizzato o il confronto con un gruppo normativo, sono stati inventati per misurare l'intelligenza. Sapere da dove vengono aiuta a capire perché funzionano in un certo modo e dove smettono di funzionare.

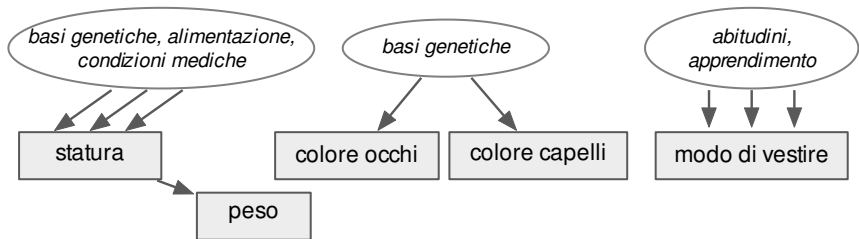
1.3 Due modi di fare psicologia

Le origini della psicologia scientifica si collocano di solito alla fine dell'Ottocento. In Germania Wilhelm Wundt apre a Lipsia, nel 1879, il primo laboratorio di psicologia sperimentale. L'interesse è rivolto ai processi generali della mente: percezione, memoria, attenzione, coscienza. Le differenze fra le persone sono un disturbo da contenere.

Negli stessi anni, in Gran Bretagna, la biologia evoluzionista di Darwin porta Francis Galton a una domanda diversa. Se la variabilità è la materia prima dell'evoluzione, allora le differenze individuali non sono rumore: sono l'oggetto da studiare. Galton raccoglie misure fisiche e sensoriali su migliaia di visitatori del suo laboratorio antropometrico. Karl Pearson formalizza la correlazione. Spearman propone l'analisi fattoriale.

Che cosa significa studiare le differenze individuali si capisce partendo da quelle che notiamo a colpo d'occhio: statura, peso, colore degli occhi e dei capelli, modo di vestire, tono di voce, modo di camminare. Viene subito da chiedersi da che cosa dipendano, e le risposte sono diverse (Figura 1.1). La statura dipende da basi genetiche, alimentazione e condizioni mediche, e a sua volta influisce sul peso. Il colore degli occhi e dei capelli dipende quasi solo da basi genetiche, in parte comuni. Il modo di vestire dipende da abitudini e apprendimento. Alcune caratteristiche sono stabili, altre cambiano in fretta. Le stesse domande valgono in psicologia: lungo quali dimensioni

le persone differiscono, da che cosa dipendono quelle differenze, quanto sono stabili.



Cambiano o sono stabili? Quanto rapidamente cambiano?

Figura 1.1: Differenze individuali visibili a colpo d’occhio e alcune delle loro cause. Cause diverse producono differenze diverse; alcune differenze ne influenzano altre; alcune cambiano nel tempo, altre sono stabili.

Nasce così una distinzione che Cronbach, molti decenni dopo, descrisse come quella fra «due discipline» della psicologia scientifica (Cronbach, 1957):

Tabella 1.1: Due prospettive nello studio della mente.

	Processi generali	Differenze individuali
Domanda	Come funziona la mente in media?	Lungo quali dimensioni le persone differiscono, e con quali conseguenze?
Disegno tipico	Esperimento controllato	Studio correlazionale e longitudinale
Campione preferito	Omogeneo	Ampio e rappresentativo
Variabilità fra persone	Rumore	Oggetto di studio

	Processi generali	Differenze individuali
Esempi	Memoria, percezione, linguaggio	Intelligenza, personalità, psicopatologia

Talvolta l'oggetto è lo stesso, ma la domanda cambia. Si prenda la memoria di lavoro. Chi studia i processi generali si chiede se una componente fonologica e una visuospatiale siano distinte perché un compito secondario interferisce con l'una e non con l'altra. Chi studia le differenze individuali si chiede se siano distinte perché alcune persone vanno meglio nell'una e altre nell'altra. Sono due evidenze diverse. Non è detto che diano la stessa risposta.

Oggi le due tradizioni si contaminano sempre di più. Un esperimento può stimare sia l'effetto medio di una manipolazione sia quanto quell'effetto varia da persona a persona. Un'illusione percettiva, un trattamento o una traiettoria di sviluppo hanno un andamento tipico e una dispersione attorno a esso. Studiare solo il primo significa perdere metà della storia.

1.4 La foresta delle differenze individuali

Le abilità cognitive sono una delle tre grandi aree in cui la psicologia descrive le differenze fra persone. Le altre due sono la personalità e la psicopatologia. In tutte e tre si è arrivati a modelli simili: molti tratti specifici, raggruppati in dimensioni più ampie, a loro volta correlate.

- Per le abilità cognitive il modello di riferimento è gerarchico, con abilità specifiche, ampie aree come intelligenza fluida e cristallizzata e, al vertice, un fattore generale.
- Per la personalità i cinque grandi fattori riassumono decine di tratti più circoscritti.
- Per la psicopatologia la proposta HiTOP organizza sintomi e sindromi in dimensioni internalizzanti, esternalizzanti e di altro tipo (Kotov et al., 2017); alcuni autori hanno ipotizzato anche un fattore generale, chiamato p per analogia con g (Caspi et al., 2014).

Queste aree condividono strumenti e problemi. Si misurano con test e punteggi; mostrano basi genetiche distribuite su moltissime varianti; hanno correlati neurobiologici; cambiano lungo lo sviluppo; interagiscono con l'esperienza. L'intelligenza è il punto in cui questi temi sono stati studiati più a lungo e con più dati. Per questo funziona bene come porta d'ingresso.

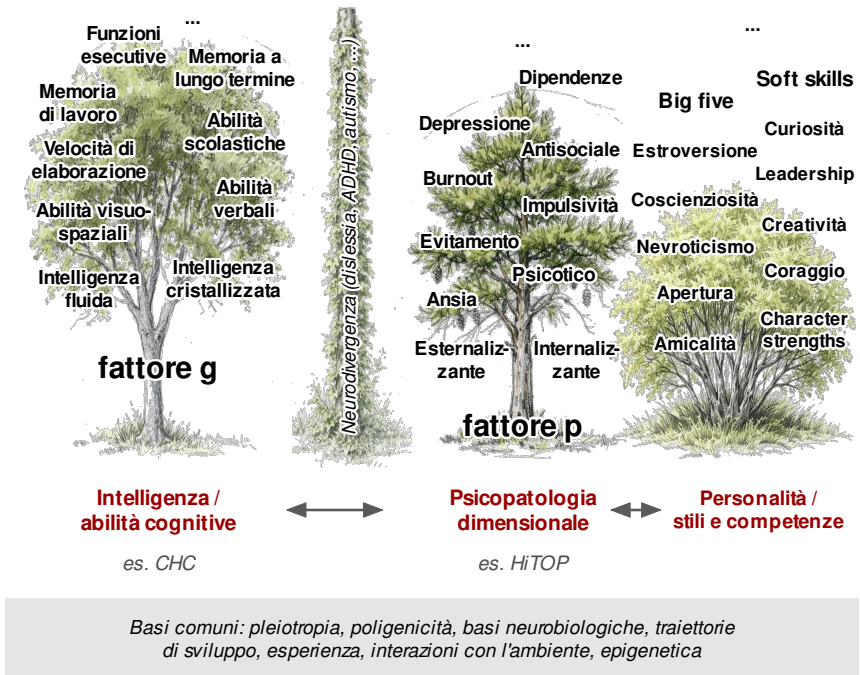


Figura 1.2: La «foresta» delle differenze individuali. Abilità cognitive, psicopatologia e personalità sono tre chiome di tratti specifici correlati; per le prime al tronco c'è il fattore g, per la psicopatologia è stato ipotizzato un fattore generale p. Le tre aree sono collegate fra loro, attraversate dalla neurodivergenza, e poggiano su basi comuni. Illustrazioni dell'autore.

Fra le chiome delle abilità cognitive e della psicopatologia corre la neurodivergenza: dislessia, ADHD e autismo hanno a che fare con entrambe, e se ne parlerà nella parte sul neurosviluppo.

L'intelligenza, del resto, sta a un crocevia (Figura 1.3). Ha dato alla psicomетria i suoi strumenti di base: test, punteggi, modelli statistici. Riceve dalla genetica del comportamento, con gli studi sui gemelli e le analisi genomiche, la prova che le sue differenze hanno basi ereditabili distribuite su migliaia di varianti. Scambia domande e risposte con le neuroscienze cognitive, che cercano le basi cerebrali delle differenze individuali. E alimenta lo studio dello sviluppo e dell'arco di vita, oltre a una lunga serie di applicazioni cliniche, scolastiche e organizzative.

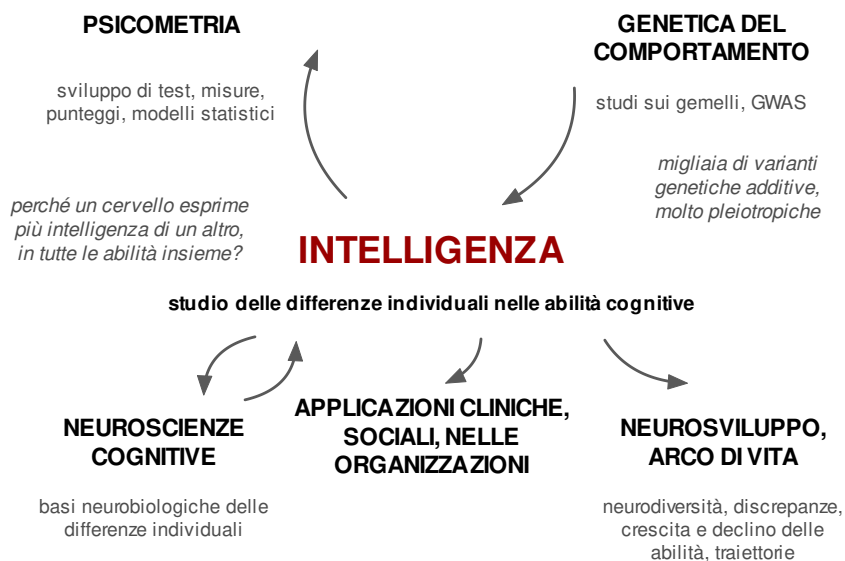


Figura 1.3: L'intelligenza, intesa come studio delle differenze individuali nelle abilità cognitive, al crocevia fra discipline: fornisce strumenti alla psicomетria, riceve dati dalla genetica del comportamento, scambia domande con le neuroscienze cognitive, alimenta lo studio dello sviluppo e le applicazioni.

1.5 Perché sembra più difficile di quanto sia

L'intelligenza non è un tema più complicato della memoria o del linguaggio. Eppure sembra essere compresa con più fatica. Una parte del problema sta nei pregiudizi. Tutti hanno un'idea di che cosa sia l'intelligenza, e molte di queste idee contraddicono ciò che la ricerca ha accertato. Il caso non è unico, ma qui la distanza fra intuizione comune e consensus scientifico è particolarmente ampia.

L'altra parte del problema è emotiva. Le obiezioni si somigliano: il QI serve a discriminare; non si possono ridurre le persone a numeri; siamo tutti uguali; nella vita conta l'impegno, non il QI; ognuno è un genio a modo suo. Alcune sono slogan. Altre contengono un nucleo fondato, che i capitoli successivi prenderanno sul serio.

Questa resistenza non nasce dal nulla. Nella prima metà del Novecento i test di intelligenza sono stati usati per giustificare politiche eugenetiche, restrizioni all'immigrazione e sterilizzazioni forzate. Sono fatti storici, non caricature polemiche, e ne parleremo nel capitolo sulle origini dei test. Ignorarli sarebbe disonesto. Concludere che il costrutto sia per questo privo di valore sarebbe però un errore di altro tipo: confondere l'uso di uno strumento con la sua validità.

Un costrutto capace di generare polemiche così accese e abusi così gravi è, evidentemente, un costrutto potente. Lo è nel bene e nel male. Le stesse misure che sono state usate per escludere servono oggi a riconoscere una disabilità intellettiva, a individuare un bisogno educativo, a documentare un declino cognitivo, a stimare l'impatto di un'esposizione ambientale nociva. La conclusione ragionevole non è rinunciare a misurare, ma imparare a misurare bene e a interpretare con giudizio.

1.6 L'impostazione di questo libro

Il percorso che segue cerca di offrire una visione equilibrata e allineata al consenso scientifico prevalente, senza nascondere le controversie, gli errori storici e gli abusi. Quando l'evidenza è solida, lo diremo chiaramente.

te. Quando è debole, contraddittoria o fraintesa, lo diremo con la stessa chiarezza.

Le domande sono quelle che chiunque si pone: che cos'è l'intelligenza, come si misura, se è una o molte, che cosa significa il QI, quanto contano geni e ambiente, dove si trova nel cervello, se stiamo diventando più o meno intelligenti, se si può potenziare, che cosa predice nella vita reale, come si valuta nelle condizioni del neurosviluppo. Per rispondere servirà anche un minimo di linguaggio quantitativo: correlazioni, distribuzioni, punteggi, qualità della misura. Non è un pedaggio da pagare per arrivare ai temi interessanti. È il modo in cui quei temi diventano discutibili con argomenti invece che con opinioni.

In breve

- I test di intelligenza sono fra gli strumenti più usati in psicologia, spesso sotto altri nomi; usarli bene richiede di conoscerne le basi concettuali, non solo le procedure di scoring.
- Lo studio dell'intelligenza ha generato metodi, come l'analisi fattoriale e le variabili latenti, che servono a tutta la psicologia.
- La psicologia dei processi generali e quella delle differenze individuali pongono domande diverse anche sugli stessi fenomeni.
- Abilità cognitive, personalità e psicopatologia si descrivono con modelli simili, tratti specifici raggruppati in dimensioni più ampie; l'intelligenza è l'area studiata più a lungo e sta al crocevia fra psicomетria, genetica, neuroscienze e sviluppo.
- La storia dei test include abusi reali. Riconoscerli non implica rinunciare a misurare, ma pretendere misure valide e interpretazioni prudenti.

Riferimenti bibliografici

2 Che cosa chiamiamo intelligenza

Possiamo dire che gli esseri umani sono intelligenti? Siamo l'unica specie che si è data da sola il nome di *sapiens*. Siamo anche una specie che altera il proprio ambiente, sa di farlo e non riesce a smettere. Il paradosso è antico, e Sternberg ha curato un intero volume su una sua versione in piccolo: perché persone intelligenti fanno a volte cose tanto stupide (Sternberg, 2002).

La domanda sembra imbarazzante solo perché alla parola «intelligenza» chiediamo troppo: la rapidità del ragionamento, la saggezza delle scelte, talvolta perfino il carattere. Nessun concetto regge un incarico così ampio. Questo capitolo restringe il campo partendo da dove parte chiunque, il senso comune, e arriva a una definizione di lavoro, a un numero, il QI, e alle obiezioni che quel numero si porta dietro.

2.1 Gli indizi di tutti i giorni

A dieci anni, come si riconosceva un compagno di classe molto intelligente? Prendeva voti alti, parlava quasi come un adulto, capiva subito le regole di un gioco nuovo, finiva il compito per primo, era bravo in matematica. Quando spiegava qualcosa agli altri, rendeva semplici cose difficili.

Altri indizi sono più interessanti. Non è che non sbagliasse mai: sapeva trovare gli errori e correggerli. Non è che non si mettesse mai nei guai: sapeva come uscirne. Non era necessariamente il più obbediente. E se copiava, lo faceva in modo efficace.

Questi indizi non sono sbagliati. Hanno però un difetto: mescolano cose diverse. Possedere una conoscenza, acquisirla in fretta e sapere quando

usarla sono operazioni distinte; l'attenzione e l'obbedienza sono altro ancora. Anche la buona condotta va tenuta fuori: capire le conseguenze di un'azione non obbliga a scegliere quelle migliori per tutti, e si può ragionare in modo eccellente al servizio di uno scopo pessimo.

Non è pedanteria. Se la definizione includesse ogni qualità desiderabile, una buona decisione sarebbe intelligente per definizione e una cattiva sarebbe stupida per definizione. Avremmo un'etichetta per commentare i risultati e nessuna possibilità di spiegare come si producono.

2.2 Che cosa dicono le parole

L'etimologia offre un primo aiuto. *Intelligentia* viene dal latino *inter* e *legĕre*: leggere fra le cose, cogliere ciò che non si vede subito. Altre lingue hanno percorsi analoghi, in cui conoscere si intreccia con vedere, percepire e capire (Figura 2.1).

Le parole, però, portano con sé anche sfumature morali, caratteriali e sociali. Warne ne raccoglie alcuni esempi (Warne, 2020). Nella cultura cinese l'intelligenza include l'uso etico della conoscenza, vicino alla saggezza. In Africa orientale, in India e altrove la tenacia nel portare avanti un compito è considerata una capacità cognitiva. I Cree del Canada vi includono riflessività, pazienza e perseveranza; in Zimbabwe vi rientrano competenze sociali come prendersi cura di amici e sconosciuti.

Conviene allora ancorarsi al significato comune? La biologia e la medicina hanno risposto di no da tempo: si sono date un gergo tecnico, spesso latino e greco, che nessuno confonde con il linguaggio di tutti i giorni. Nessun medico discute se un'epatite sia «davvero» un'inflammatione nel senso del bar. La psicologia scientifica avrebbe potuto fare lo stesso con più decisione. Non l'ha fatto, e paga il prezzo di usare una parola che tutti credono di capire.

latino: **inter** + **legĕre** → **intelligentia**, capacità di «leggere tra» le cose:
comprendere anche ciò che non si vede immediatamente

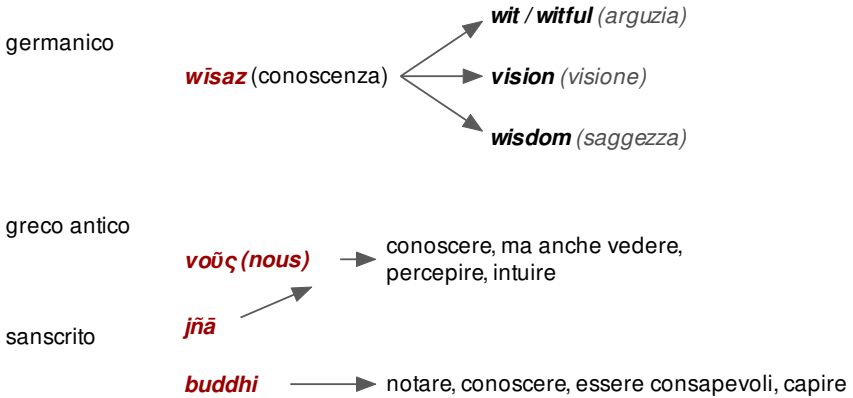


Figura 2.1: Le radici di alcune parole legate all'intelligenza: in lingue diverse conoscere si intreccia con vedere, percepire e capire.

2.3 Le definizioni degli esperti

Se il senso comune non basta, si può chiedere agli esperti. Il guaio è che ciascuno ha la sua definizione. Nel 1921 il *Journal of Educational Psychology* interpellò diversi studiosi (Dearborn, 1921; «Intelligence and Its Measurement», 1921; Woodrow, 1921). Per Pintner l'intelligenza era la capacità di adattarsi a situazioni relativamente nuove; per Woodrow, quella di acquisire abilità; per Dearborn, quella di trarre profitto dall'esperienza; per Terman, quella di condurre un ragionamento astratto. Settantasette anni dopo, nel 1998, un dibattito italiano sulla rivista *Sistemi Intelligenti* fra studiosi di discipline diverse, dalla psicologia all'informatica, produsse una varietà simile, dal comportamento che massimizza il beneficio dell'organismo alla capacità di costruirsi rappresentazioni interne del mondo (Cornoldi, 2007).

Sotto la varietà, però, c'è una traiettoria. Da sempre le definizioni insistono

2 Che cosa chiamiamo intelligenza

su **adattamento** e **apprendimento**: rispondere a richieste nuove quando le abitudini non bastano, e migliorare con l'esperienza. Dagli anni Cinquanta la rivoluzione cognitiva ha aggiunto le **rappresentazioni mentali**, le operazioni interne, la soluzione di problemi attraverso l'elaborazione di informazione. Chi si rappresenta la logica di un problema può confrontare possibilità e scartare un'ipotesi prima di provarla.

Sono concetti utili, perché danno finalmente una base teorica. Ma non sono mai stati strettamente necessari per studiare e misurare bene l'intelligenza. I test funzionavano prima della rivoluzione cognitiva, e funzionerebbero anche se le sue teorie venissero sostituite domani.

La formulazione oggi più citata è di Gottfredson, firmata da decine di ricercatori: l'intelligenza è una capacità mentale generale che, tra le altre cose, implica ragionare, pianificare, risolvere problemi, pensare in termini astratti, comprendere idee complesse, apprendere rapidamente e apprendere dall'esperienza (Figura 2.2). Non è l'apprendimento dai libri né l'abilità nel sostenere test; riflette una capacità più ampia di afferrare ciò che ci circonda, di dare senso alle cose e capire che cosa fare (Gottfredson, 1997, p. 13).

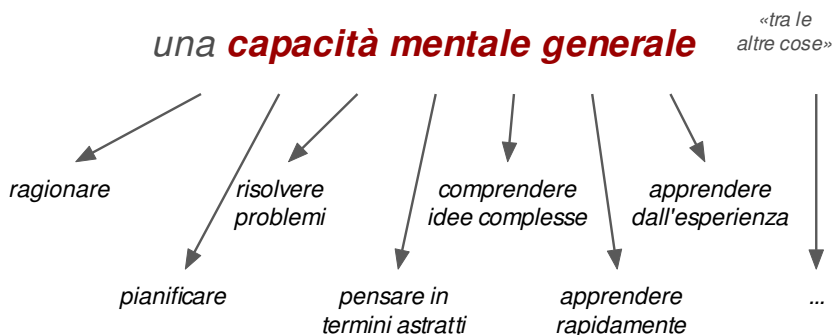


Figura 2.2: Una capacità mentale generale e alcune delle attività in cui interviene, «tra le altre cose». Elaborazione da Gottfredson (1997).

«Generale» indica l'ampiezza del campo di applicazione, non l'uniformità dei risultati. La stessa capacità interviene in compiti diversi, e una persona

può riuscire meglio in alcuni che in altri. Se l'intelligenza sia una cosa sola o molte è la domanda del capitolo sul fattore *g*.

2.4 Un'onesta definizione operazionalista

Nel 1923 Boring propose una via d'uscita più spiccia: l'intelligenza è ciò che misurano i test di intelligenza (Boring, 1923). Sembra una battuta, ed è stata spesso citata come tale. In realtà è un'ammissione onesta. Storicamente l'interesse a *misurare* l'intelligenza è sempre stato molto più forte dell'interesse a *definirla*, e per ragioni pratiche solide: scuole, eserciti e cliniche avevano bisogno di decidere prima che i teorici si mettessero d'accordo.

Il rischio, naturalmente, è la circolarità: un test misura l'intelligenza perché l'intelligenza è ciò che il test misura. Se ne esce solo mettendo alla prova le misure con dati indipendenti, come si vedrà nel capitolo sulla qualità della misura.

La stessa diffidenza verso la parola spiega i nomi alternativi. In genetica del comportamento, psicologia del lavoro, psicomетria applicata ed economia si preferisce spesso parlare di **abilità mentale generale** (*GMA*, molto frequente), *abilità cognitiva generale*, *abilità generale* o semplicemente *g*. Quest'ultima sigla è classica nella letteratura psicometrica e ha un significato tecnico più ristretto: il fattore comune ricavato dalle correlazioni fra prove diverse. Per il resto del libro basta un glossario minimo:

- **abilità mentale (o cognitiva) generale**, *GMA* o *g*: la cosa che si vuole misurare;
- **QI**: l'indicatore numerico con cui la si stima.

Per l'indicatore, invece, la parola «intelligenza» resta quasi universale: il **quoziente di intelligenza**, o **QI**. È un indicatore di differenze individuali: esprime quanto bene una persona se la cava, rispetto ad altre della stessa età, su una scala standardizzata. Il QI non è *g*: *g* appartiene a un modello delle relazioni fra misure, il QI è il risultato concreto di uno strumento. Sono strettamente legati, ma non coincidono (Neisser et al., 1996, p. 78).

2.5 Perché un numero conta tanto

A questo punto l'obiezione è legittima. Perché una cosa che non sappiamo bene che cosa sia, che non sappiamo nemmeno come chiamare e che forse non è neppure «una cosa» dovrebbe essere così importante?

La ragione, prima ancora che teorica, è operazionalista. In un'ora o poco più, con i test, si ottiene una stima robusta, affidabile e stabile di un indicatore che è fra i più potenti predittori di esiti a lungo termine disponibili in psicologia delle differenze individuali (Deary, 2012a; Warne, 2020):

- **successo scolastico:** voti, completamento degli studi, anni di istruzione, con correlazioni dell'ordine di 0,40–0,60 (Roth et al., 2015); in parte è ovvio, perché i primi test furono costruiti proprio per questo scopo;
- **prestazione lavorativa:** rapidità di apprendimento e produttività, soprattutto nei lavori di complessità medio-alta, con correlazioni fra circa 0,20 e 0,50 (Schmidt & Hunter, 2004);
- **posizione sociale raggiunta:** reddito e status socioeconomico, anche al netto del punto di partenza familiare (Strenze, 2007);
- **salute e longevità:** meno psicopatologia, minore mortalità, migliore salute fisica; le associazioni sono più piccole, ma una deviazione standard in più di QI in età giovanile si accompagna a un rischio di morte inferiore di circa un quarto (Calvin et al., 2011).

Ognuno di questi legami ha un capitolo dedicato nella parte su scuola, salute e lavoro, con i suoi limiti. Qui conta un punto solo: prevedere bene un esito è una prova importante a favore di una misura, ma non spiega perché la previsione funziona. Qualità della misura e spiegazione causale sono domande diverse.

Strumenti → Come si legge una correlazione di 0,50, quando è attendibile e perché non basta a stabilire una causa: *Correlazione e differenze individuali*.

2.6 Come si arriva al QI

Proprio perché la capacità da misurare è generale, nessun test particolare è ideale da solo. Si preferisce comporre prove di natura diversa: completare una tabella di figure scegliendo il pezzo mancante, rispondere a domande di cultura generale o spiegare il significato di una parola, ripetere una serie di cifre, ricordare una posizione, riconoscere quali figure sono uguali a un modello dopo una rotazione, costruire un disegno con dei cubi, associare simboli a cifre il più in fretta possibile (Figura 2.3). Ogni prova ha i suoi difetti particolari; mettendone insieme molte, i difetti pesano meno.

Il punteggio di ciascuna prova viene **standardizzato**, cioè confrontato con quello di un campione di riferimento della stessa età; i punteggi standardizzati vengono sommati e il totale viene riportato sulla scala del QI. Su quella scala la media è 100 e la deviazione standard, di solito, 15. La maggior parte delle persone si concentra intorno alla media, e i livelli prestativi molto bassi o molto alti sono rari (Figura 2.4).

Due conseguenze vanno fissate subito. La prima: il QI è una posizione relativa. Ha senso solo nel confronto con le **norme**, cioè con il campione su cui il test è stato tarato, e il confronto è legittimo solo se quella popolazione è pertinente alla persona valutata (American Educational Research Association et al., 2014, standard 5.8). La seconda: la scala non ha uno zero naturale né un'unità di intelligenza. Un QI di 120 non significa aver risolto il 120% dei problemi, né avere il 20% di intelligenza in più di chi ha 100.

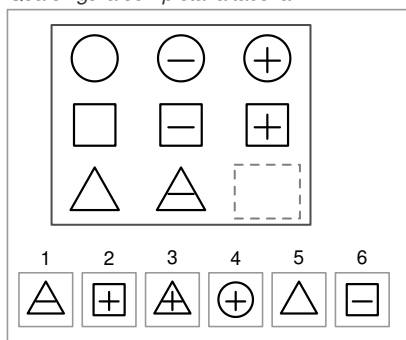
Strumenti → Perché il QI si distribuisce a campana, come si passa da un punteggio grezzo al QI e che cosa sono punti *z*, punteggi ponderati e percentili: *Distribuzioni, punti *z* e punteggi standardizzati.*

2.7 Obiezioni ai test

Chi vede per la prima volta un test di intelligenza di solito protesta. Perché queste banalità dovrebbero rappresentare la mia intelligenza? Perché non proporre problemi davvero complessi e articolati? Perché non valutare che

2 Che cosa chiamiamo intelligenza

Quale figura completa la tabella?



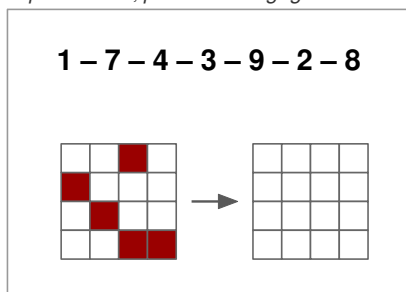
Rispondi alle domande

A quale evento storico è associato l'anno 1789?

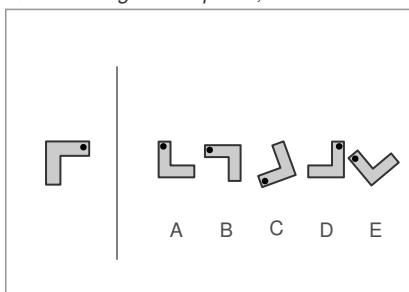
Che cosa significa «ineffabile»?

- A) Che non può essere negato
- B) Che non può essere espresso a parole
- C) Caratterizzato da lunghe pause
- D) Superfluo, inutile

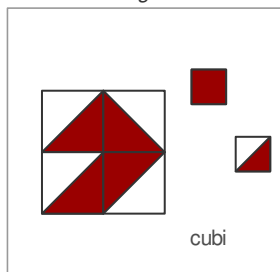
Ripeti la serie, poi ricorda la griglia



Quali sono uguali alla prima, se ruotate?



Costruisci la figura



Completa col simbolo (tempo limitato)

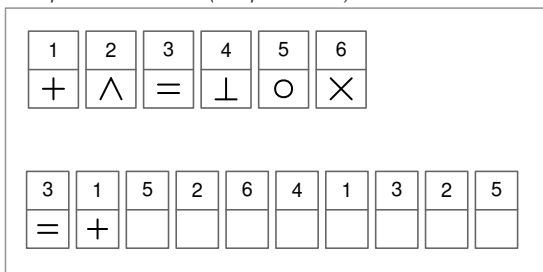


Figura 2.3: Esempi di item di prove cognitive diverse. Sono item costruiti per questo libro: riprendono la logica di prove molto diffuse (matrici, cultura generale, vocabolario, ripetizione di cifre, memoria visuo-spaziale, rotazione mentale, disegno con cubi, codifica di simboli), ma non riproducono item di test pubblicati. Soluzioni: matrice, figura 3; rotazione, A, C ed E (B e D sono immagini speculari).

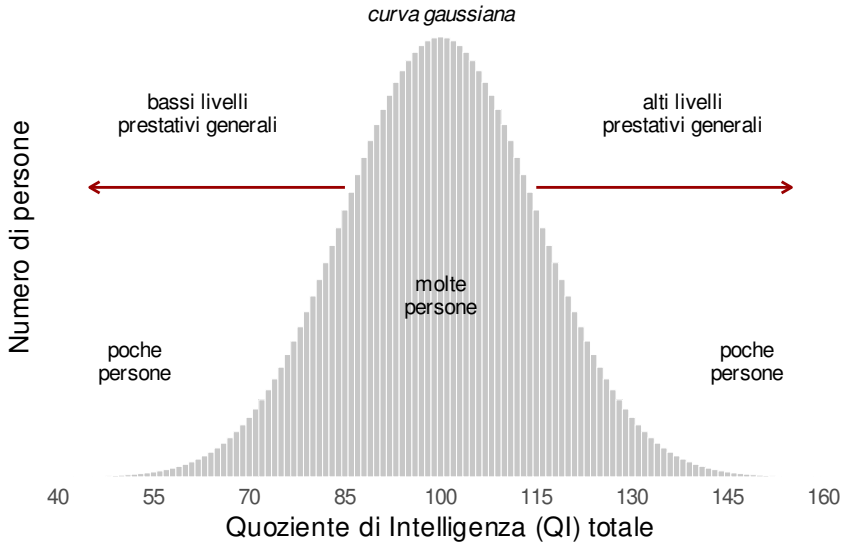


Figura 2.4: Il QI totale si distribuisce come una curva gaussiana: molte persone intorno a 100, poche agli estremi. Distribuzione teorica con media 100 e deviazione standard 15.

cosa una persona sa fare e produrre nella realtà? E perché affidarsi a vecchi test comportamentali, invece di sfruttare le neuroscienze?

La prima risposta è che nessuna delle prove viste sopra è l'intelligenza. L'intelligenza si riflette in tutte, e in molte altre cose. Una singola prova è un campione piccolo e imperfetto; molte prove diverse sono un campione migliore.

La seconda riguarda il formato. Si preferiscono compiti standardizzabili, con risposte giuste ben identificate, somministrabili a molte persone in tempi ridotti. Un problema molto articolato fornisce un solo risultato dove servirebbero decine di osservazioni: poca granularità, poca attendibilità, e la difficoltà di valutare processi e risultati in modo uguale per tutti.

La terza spiega perché molti compiti sembrano strani. L'obiettivo è vedere la potenza del motore, non quanta strada abbia già fatto l'automobile.

Un compito che misura soprattutto ciò che una persona ha già avuto occasione di imparare dice molto sulla sua storia e poco sulla sua capacità di affrontare il nuovo. L'eccezione importante è la competenza verbale: il vocabolario è appreso, eppure si rivela un buon indicatore della capacità generale. Una spiegazione plausibile è che la maggior parte delle parole non viene insegnata: se ne ricava il significato dai contesti, e farlo è già un piccolo problema da risolvere.

C'è infine la filosofia operazionalista: l'importante è che i test funzionino. È una posizione difendibile solo se il funzionamento viene dimostrato, con dati, per ogni uso che se ne fa. «Sembra artificiale» e «sembra realistico» sono impressioni; la qualità di una misura si giudica dalle evidenze.

Quanto alle neuroscienze e alla genetica, il loro contributo concettuale è eccellente: hanno mostrato, per esempio, che *g* ha correlati biologici reali. La loro capacità predittiva, però, resta modesta. Il miglior indicatore cerebrale singolo è ancora un dato quasi banale come il volume complessivo del cervello, che correla con il QI intorno a 0,24 (Pietschnig et al., 2015). E il criterio con cui si giudicano queste misure resta comunque il comportamento: un indice cerebrale interessa perché predice la prestazione, non il contrario. Il capitolo su cervello e intelligenza riprende la questione.

In breve

- «Intelligenza» si può definire come abilità mentale generale (GMA): una capacità che sostiene molte capacità cognitive diverse e non coincide con ogni qualità desiderabile.
- Il termine è poco tecnico; è lecito sostituirlo con alternative come GMA o *g*, tenendo distinti la capacità da misurare e il QI, che ne è l'indicatore.
- La misura interessa perché riassume una dimensione principale delle differenze individuali e predice molti esiti diversi; prevedere, però, non equivale a spiegare.
- Il QI si ottiene aggregando molte prove di natura diversa, standardizzate e riportate su una scala con media 100 e deviazione standard 15.

Riferimenti bibliografici

Parte II

Modelli e test dell'intelligenza

3 Il fattore generale e la positive manifold

Chi va bene in un test cognitivo tende ad andare bene anche negli altri. Non in tutti allo stesso modo, non sempre, ma con una regolarità che più di un secolo di ricerca non è riuscito a smentire. Vocabolario e rotazione mentale, memoria di cifre e ragionamento con le matrici, aritmetica e velocità nel confrontare simboli: compiti che non si assomigliano in nulla producono punteggi correlati positivamente. Questo fatto ha un nome, **positive manifold**, ed è il punto di partenza obbligato di ogni teoria dell'intelligenza.

***Strumenti** → Per seguire questo capitolo basta l'idea intuitiva: una correlazione positiva fra due prove significa che chi va bene nell'una tende ad andare bene anche nell'altra. Come se ne misura e se ne giudica la grandezza: **Correlazione e differenze individuali**.*

La domanda è se l'intelligenza sia una cosa sola o molte. Assomiglia alla personalità, che si descrive con diverse dimensioni distinte? Oppure tutto tende a convergere su un'unica dimensione? E il QI riflette qualcosa di reale, con basi biologiche, o è solo la somma di abilità diverse a cui si è dato un nome solo? Lo studio dell'intelligenza ha oscillato a lungo fra un approccio **unitario** e uno **multiplo**. Questo capitolo presenta il primo, nella forma che gli diede Charles Spearman. Il prossimo presenta le alternative e la sintesi gerarchica oggi prevalente.

3.1 Perché le prestazioni correlano?

Un'associazione fra due misure ammette poche spiegazioni di principio: A causa B, B causa A, le due si influenzano a vicenda, oppure una terza variabile influenza entrambe. Appliciamole a due prove cognitive. Che il vocabolario causi direttamente la capacità di ruotare mentalmente una figura

è implausibile, almeno nel presente di un adulto. Che accada il contrario, altrettanto. Resta l'ipotesi più semplice: una causa comune. Un'influenza reciproca accumulata durante lo sviluppo non è esclusa, ed è la base delle alternative discusse negli approfondimenti finali; ma richiede una storia più lunga da raccontare.

Nel 1904 Spearman pubblicò un'analisi delle correlazioni fra misure molto eterogenee, raccolte su bambini delle scuole: voti in varie materie, discriminazioni sensoriali, giudizi degli insegnanti. Notò che tutte le correlazioni erano positive e che il loro schema poteva essere spiegato da un unico fattore sottostante, comune a tutte le misure, più una componente specifica di ciascuna (Spearman, 1904). Chiamò il primo **fattore generale**, o *g*; le seconde fattori specifici, o *s*.

Gli indicatori usati da Spearman erano rudimentali. Pochi anni dopo Binet e Simon avrebbero costruito le prime batterie di prove cognitive simili a quelle moderne. Ma l'idea era già completa: le prestazioni osservate sono indizi di una capacità comune che non si osserva direttamente.

*in basso, le componenti
«s» specifiche, senza
vera dignità teorica*

*«g»: il vero nucleo
dell'intelligenza;
oggi diremmo:
fattore latente*

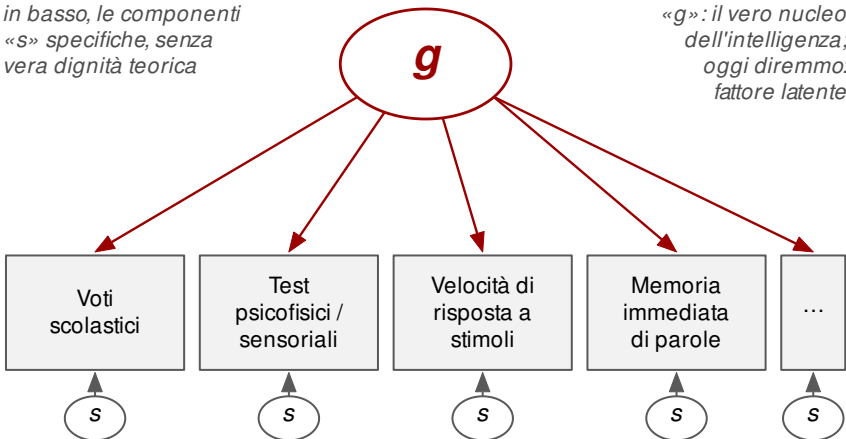


Figura 3.1: Il modello di Spearman: un fattore generale *g*, comune a tutte le misure, e una componente specifica *s* per ciascuna. Le misure indicate sono del tipo usato da Spearman, ancora molto rudimentali. Elaborazione da Spearman (1904).

3.2 Un fattore latente, spiegato con lo sport

L'idea di fattore latente è più intuitiva di quanto sembri. Prendiamo un esempio che con l'intelligenza non ha nulla a che fare. Misuriamo in un ampio gruppo di persone il tempo per nuotare una vasca, il tempo per percorrere dieci chilometri in bicicletta, la lunghezza di un salto, la precisione nei tiri liberi a basket. Nella popolazione generale queste prestazioni saranno correlate. Chi nuota velocemente tende anche a pedalare velocemente e a saltare lontano.

Perché? Non perché il nuoto migliori il salto. Esistono caratteristiche comuni che favoriscono tutte queste attività: forza muscolare, coordinazione, composizione corporea, flessibilità, età. Possiamo chiamarle insieme «atleticità». Ogni prova dipende inoltre da componenti specifiche, come l'allenamento in quella disciplina o la statura per il basket. Un modello con un fattore generale e tanti fattori specifici spiega le correlazioni in modo parsimonioso. Fra atleti olimpici, molto selezionati e specializzati, le correlazioni potrebbero quasi sparire; nella popolazione generale sono forti. La Figura 3.2 traduce l'esempio in un modello, e la Figura 3.1 mostra che è lo stesso schema proposto da Spearman.

Lo stesso schema vale per la personalità. Le risposte ad affermazioni come «mi piace conoscere persone nuove», «mi riesce facile iniziare una conversazione con uno sconosciuto», «mi piace essere al centro dell'attenzione» sono correlate, ed è su questa correlazione che si costruiscono i questionari di personalità. Il fattore comune si chiama estroversione (Figura 3.3). Dietro c'è probabilmente un intreccio di livelli diversi: la sensibilità dei sistemi di ricompensa agli stimoli sociali, la ricerca di sensazioni, migliaia di varianti genetiche, poche delle quali identificate. Nessuno osserva direttamente il fattore; lo si inferisce dalle correlazioni e lo si usa perché spiega e prevede. Un item formulato al contrario, come «mi sento a disagio fra molta gente», ha un peso negativo: chi è più estroverso tende a dare punteggi più bassi.

in basso, le componenti
«s»: allenamento
specifico, familiarità
con la disciplina,
statura...

cause comuni: forza,
coordinazione,
composizione corporea,
flessibilità, età...

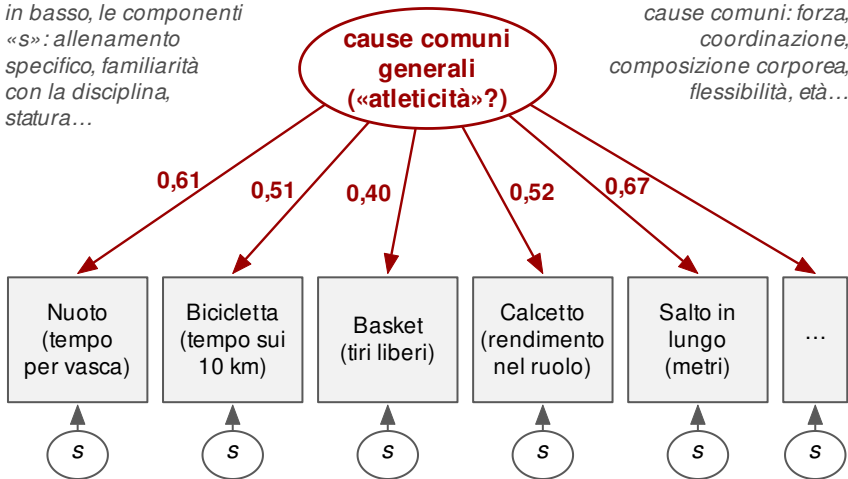


Figura 3.2: Un fattore latente fuori dalla psicologia. Cause comuni generali, riassumibili come «atleticità», spiegano perché prestazioni sportive diverse sono correlate; ogni prova dipende anche da componenti specifiche *s*, come l'allenamento nella disciplina o la statura. Il numero di prove non ha un limite definito. I pesi sono inventati a scopo illustrativo.

3.3 Che cos'è, e che cosa non è, g

La teoria di Spearman è il primo caso storico di definizione di un fattore latente in psicologia. Oggi i fattori latenti si usano ovunque, dalla psicomètria all'economia alla biostatistica. Conviene quindi chiarire bene che cosa si intende.

g si chiama **latente** perché non si osserva direttamente. Nessuno dei punteggi ottenuti nei test è la cosa che interessa; tutti servono solo da misure indirette. Nell'esempio sportivo si potrebbe in teoria misurare la forza o la coordinazione. Nel caso dell'intelligenza, in pratica, no.

g non è necessariamente una cosa sola. Anzi, si ipotizza che sia l'aggregato di moltissimi fattori: varianti genetiche, caratteristiche del cervello, esperienze, condizioni di sviluppo. Lo si tratta come unitario perché questi fattori, nei dati comportamentali, non si lasciano separare: agiscono tutti

3 Il fattore generale e la positive manifold

in basso, ε : tendenze proprie della singola situazione, non comuni alle altre

dietro: sistemi di ricompensa, ricerca di sensazioni, migliaia di varianti genetiche

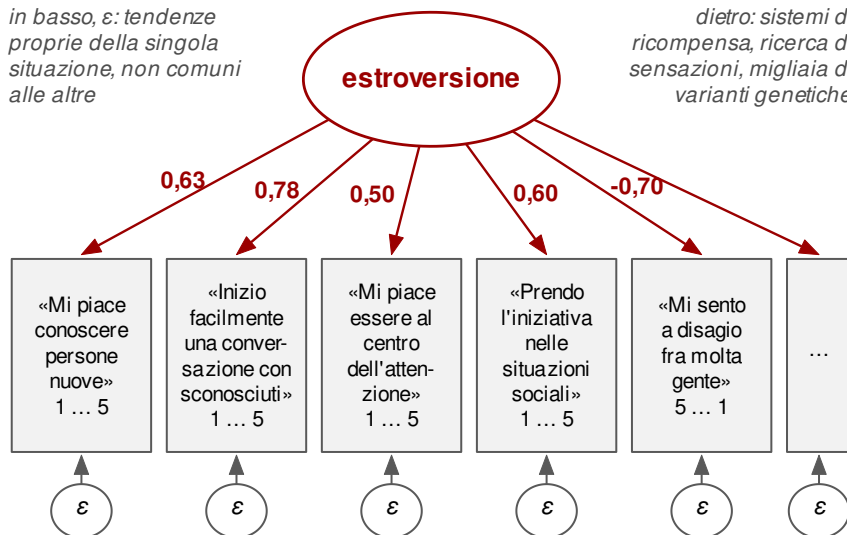


Figura 3.3: Lo stesso modello applicato a un questionario di personalità. L'estroversione spiega le correlazioni fra le risposte agli item; ε indica la parte di ogni risposta legata alla situazione specifica e non condivisa con le altre. L'item formulato al contrario ha peso negativo. Item e pesi sono inventati a scopo illustrativo.

nella stessa direzione su tutte le prove. Il «fattore» è un riassunto statistico di cause comuni, non un organo o una sostanza.

Infine, g è utile perché è una spiegazione parsimoniosa e perché prevede cose rilevanti. Vale finché ha questa validità predittiva. Se domani un modello con più dimensioni spiegasse meglio le correlazioni fra prove e gli esiti nella vita reale, g andrebbe rivisto.

3.4 «Se non vedo, non credo»

A molti l'idea di lavorare con qualcosa che non si vede sembra sospetta. È una diffidenza rispettabile, condivisa da tradizioni importanti: il comportamentismo radicale, l'operazionalismo, alcune teorie moderne che spiega-

no le correlazioni fra abilità senza ricorrere a un fattore comune, di cui si parla negli approfondimenti finali.

Concentrarsi su ciò che si osserva è giusto. Ma spesso la spiegazione dei fenomeni impone di postulare qualcosa che non si osserva. Il comportamentismo radicale entrò in crisi proprio di fronte a fenomeni come l'apprendimento latente e l'acquisizione del linguaggio, che richiedevano di ipotizzare rappresentazioni interne; da quella crisi nacquero le scienze cognitive. Fuori dalla psicologia, la cosmologia attribuisce gran parte della massa dell'universo a una «materia oscura» mai osservata direttamente, ma necessaria per spiegare la dinamica delle galassie. Il paragone con g è naturale: qualcosa che deve esserci per dare conto dei dati, anche se non si riesce ancora a osservarlo direttamente.

L'analogia ha un limite a favore di g . Per l'intelligenza si conoscono già alcune componenti: varianti genetiche comuni e caratteristiche del cervello associate al fattore generale. Spiegano però una quota ancora piccola delle differenze. Ne parleremo nei capitoli sulla genetica e sul cervello.

3.5 Quanto è solido il fattore generale

Che il fattore generale sia un risultato robusto lo mostrano due tipi di evidenza. Il primo è la sua ubiquità. Nella più ampia rianalisi mai condotta delle correlazioni fra test cognitivi, Carroll esaminò oltre quattrocento insiemi di dati raccolti in decenni di ricerca: la positive manifold e un fattore generale emergevano in modo sistematico (Carroll, 1993).

Il secondo è la sua indipendenza dalle prove usate per stimarlo. Se g fosse solo il fattore comune di una batteria particolare, batterie diverse dovrebbero produrre g diversi. In uno studio su adulti che avevano svolto tre batterie complete e diverse fra loro, i fattori generali stimati separatamente correlavano fra loro quasi perfettamente (Johnson et al., 2004). È la versione moderna dell'«indifferenza dell'indicatore» di Spearman: purché le prove siano abbastanza numerose e varie, il fattore generale che se ne ricava è sostanzialmente lo stesso.

Per questo *g*, pur con una parte di nebulosità teorica, resta una nozione centrale e largamente accettata nella ricerca sull'intelligenza (Deary, 2012a; Gottfredson, 1997). Oggi viene inserito in un quadro gerarchico, meno rigido dell'unitarismo originario, ma non è stato abbandonato.

3.6 I limiti della teoria unitaria

La teoria unitaria classica ha però tre limiti seri, che le teorie successive hanno cercato di superare.

Il primo è l'assenza di una teoria psicologica. Il fattore generale descrive che le prestazioni covariano, ma non dice quale processo mentale le accomuni. Velocità di elaborazione? Memoria di lavoro? Controllo dell'attenzione? Efficienza delle connessioni cerebrali? In senso minimo *g* indica solo «cause comuni». Per questo è stato sospettato di essere un epifenomeno statistico reificato, cioè un'astrazione trattata come se fosse una cosa. Il sospetto non tocca l'utilità pratica del *QI*, che nessuno contesta seriamente, ma il suo statuto teorico.

Il secondo è la dipendenza dalla scelta delle prove. Il fattore comune di una batteria è il fattore comune delle prove che contiene. Le prime scale storiche erano piene di prove verbali, e il loro fattore generale era sbilanciato di conseguenza. Il risultato sull'equivalenza dei *g* stimati da batterie diverse attenua il problema, ma solo per batterie ampie e ben bilanciate.

Il terzo è la sottovalutazione dei fattori specifici. Se tutto ciò che conta è *g*, le differenze fra le abilità di una persona diventano rumore. Ma un punto di forza nelle abilità visuospatiali o una debolezza selettiva nel linguaggio possono essere rilevanti per capire un caso clinico o una scelta di studi. L'esistenza di un fattore generale non implica che non esista nient'altro.

Questi limiti non sono stati ignorati. Hanno alimentato un secolo di teorie alternative, con fortune molto diverse. Alcune hanno lasciato contributi duraturi; altre hanno avuto più successo nelle scuole che nei dati. Il prossimo capitolo le esamina e mostra come la ricerca sia arrivata a una sintesi.

In breve

- Prestazioni cognitive molto diverse sono sempre correlate positivamente: è la positive manifold.
- Spearman la spiegò con un fattore generale comune, *g*, più fattori specifici di ciascuna prova: il primo modello a variabile latente della psicologia.
- *g* non è un oggetto unico nel cervello, ma il riassunto di molte cause comuni che agiscono nella stessa direzione su tutte le prove.
- Il fattore generale è robusto: emerge in centinaia di insiemi di dati ed è quasi identico se stimato da batterie diverse.
- La teoria unitaria classica manca però di una spiegazione psicologica, dipende dalle prove scelte e trascura le abilità specifiche.

Riferimenti bibliografici

4 Abilità multiple e modelli gerarchici

Nessuno ha mai seriamente negato la positive manifold. Ciò che molti hanno negato è che la spiegazione migliore sia un fattore generale. Le prove correlano, sostengono, perché si sovrappongono: ruotare mentalmente una figura richiede anche memoria di lavoro, che a sua volta richiede velocità di elaborazione. Oppure perché l'intelligenza misurata dai test è solo una fra molte intelligenze. Oppure, in alcune teorie recenti, perché le abilità si sostengono a vicenda durante lo sviluppo.

Questo capitolo segue la storia delle alternative al fattore unico. Non è un catalogo: ogni teoria va giudicata per ciò che ha spiegato e per l'evidenza che la sostiene. Alcune hanno fallito in modo utile, altre hanno avuto più fortuna nelle aule che nei dati. Il punto d'arrivo è il modello gerarchico, oggi la sintesi più accreditata. Una sintesi che però non ha chiuso del tutto il contrasto di fondo: conta di più l'abilità generale o le abilità specifiche?

4.1 Thurstone e le abilità mentali primarie

Negli anni Trenta Louis Thurstone osservò che molti test dell'epoca erano impuri. Voti scolastici, problemi aritmetici, memoria di parole, completamento di sequenze numeriche: ciascuno mescolava diverse competenze mentali. Forse, pensò, *g* era un'illusione. I test sembravano diversi, ma chiamavano in causa sempre le stesse capacità di base, e per questo correlavano.

Nel 1938 propose una teoria radicalmente multicomponentiale. Il comportamento intelligente dipenderebbe da un insieme di **abilità mentali primarie**, teoricamente indipendenti (Thurstone, 1938). Fra queste c'erano la

comprensione verbale, la fluenza verbale, il ragionamento induttivo, il calcolo numerico, la visualizzazione spaziale, la memoria e la velocità percettiva. Se si fossero costruiti test puri di ciascuna abilità, questi non sarebbero stati correlati.

Thurstone costruì quei test e affinò le tecniche dell'analisi fattoriale per analizzarli. Il risultato fu un fallimento utile: le abilità primarie risultavano correlate fra loro. Thurstone stesso finì per ammettere un fattore generale di ordine superiore. La sua teoria non è più attuale, ma ha lasciato due eredità: test di abilità specifiche ancora usati nella ricerca e l'idea che fra il fattore generale e le singole prove esista un livello intermedio.

4.2 Sternberg e Gardner: allargare il concetto

Una seconda famiglia di teorie non contesta tanto la struttura statistica quanto l'ampiezza del concetto. Robert Sternberg ha proposto una **teoria triarchica** con tre forme di pensiero intelligente (Sternberg, 1985). L'intelligenza analitica corrisponde grosso modo a ciò che misurano i test. L'intelligenza pratica è la capacità di adattarsi alle richieste quotidiane usando al momento giusto ciò che si è appreso. L'intelligenza creativa è la capacità di trovare soluzioni nuove.

L'enfasi sul comportamento nei contesti reali è condivisibile. Il problema è la misura. L'intelligenza analitica si misura bene ed è un ottimo predittore degli apprendimenti e un buon predittore del successo lavorativo. Le misure di intelligenza pratica e creativa si sono rivelate più difficili da costruire e, quando funzionano, correlano con l'abilità generale più di quanto la teoria preveda.

Howard Gardner ha proposto le **intelligenze multiple**: linguistica, logico-matematica, spaziale, musicale, corporeo-cinestesica, interpersonale, intrapersonale, a cui ne ha poi aggiunte altre (Gardner, 1983). I criteri per riconoscere un'intelligenza erano ambiziosi: basi neurobiologiche localizzabili, operazioni cognitive dedicate, una storia evolutiva plausibile, uno sviluppo descrivibile, l'esistenza di talenti o deficit selettivi. Ogni persona avrebbe una combinazione unica di intelligenze.

La teoria è molto popolare in ambito educativo. Il suo sostegno empirico è scarso. Gardner non ha sviluppato misure standardizzate delle sue intelligenze, e quando altri ricercatori le hanno costruite hanno trovato che le prove delle intelligenze più «cognitive» correlano fra loro e con *g*, mentre quelle musicale e corporea dipendono in parte da abilità non cognitive (Visser et al., 2006; Waterhouse, 2006). Le intelligenze multiple, poi, non hanno tutte la stessa importanza. Una scarsa abilità musicale o corporeo-cinestesica pesa poco nella vita quotidiana; una scarsa abilità logico-matematica è spesso parte di una disabilità intellettiva. Chiamare tutto «intelligenza» rende l'etichetta più inclusiva e meno informativa.

C'è poi un problema di numero. Joy Guilford arrivò a ipotizzare oltre cento abilità distinte, poi quasi duecento (Guilford, 1967). Come ha osservato Alan Kaufman, se un'abilità sola era troppo poco, centoventi sono troppe (Kaufman, 2009).

4.3 Intelligenza fluida e cristallizzata

Una distinzione proposta da Raymond Cattell ha avuto invece un successo duraturo, perché poggia su dati solidi (Cattell, 1963).

L'intelligenza fluida (*Gf*) è la capacità di ragionare su problemi nuovi, indipendentemente dall'istruzione, ricavando regolarità da stimoli mai visti. Si misura preferibilmente con prove non verbali, come le matrici. **L'intelligenza cristallizzata** (*Gc*) è l'insieme delle conoscenze e delle strategie acquisite, insieme alla capacità di usarle quando servono. Dipende molto dall'istruzione e si misura soprattutto con prove verbali, come il vocabolario o le domande di cultura generale.

Le due forme sono fortemente correlate. Si distinguono soprattutto per il modo in cui cambiano nel corso della vita (Figura 4.1). L'intelligenza fluida cresce rapidamente durante lo sviluppo, raggiunge il massimo nella giovane età adulta e poi declina insieme alle abilità neurocognitive generali. L'intelligenza cristallizzata continua a crescere a lungo e può aumentare anche in età avanzata, finché si accumulano esperienze (McArdle et al., 2002).

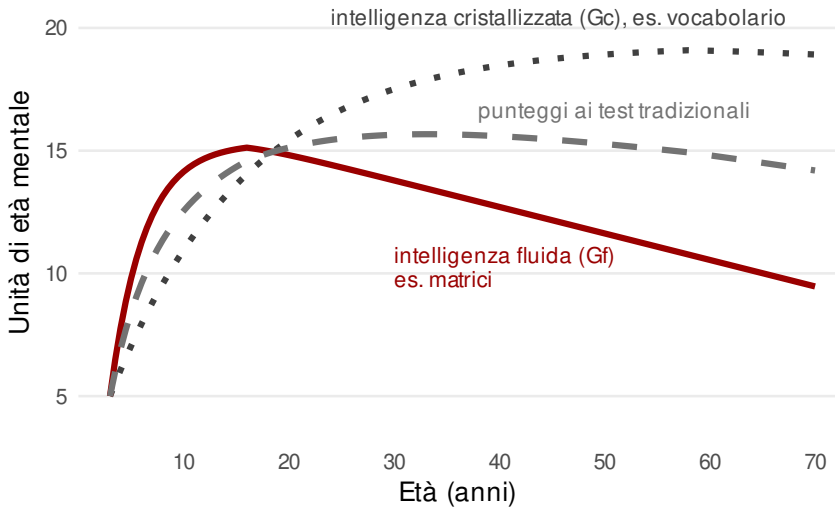


Figura 4.1: Andamento dell'intelligenza fluida, dell'intelligenza cristallizzata e dei punteggi ai test tradizionali, che le mescolano, nel corso della vita, in unità di età mentale. Curve schematiche ridisegnate dal grafico di Cattell (1987), con le tendenze confermate da McArdle et al. (2002); non sono dati osservati.

Le due forme interagiscono. Una pera e un fungo pesano insieme 200 grammi, la pera e una cipolla 240, la cipolla e il fungo 110: quanto pesano tutti e tre (Figura 4.2)? Chi non ha mai visto un sistema di equazioni può arrivare alla soluzione per tentativi, con un notevole impegno di intelligenza fluida. Chi lo riconosce come $X + Y = 200$, $X + Z = 240$, $Z + Y = 110$ somma le tre equazioni, ottiene $2(X + Y + Z) = 550$ e risponde 275 grammi in due passaggi. L'apprendimento fornisce strategie e strutture concettuali, cioè intelligenza cristallizzata, che rendono facili problemi che la sola intelligenza fluida affronterebbe con fatica.

Secondo la **teoria dell'investimento** di Cattell, l'intelligenza fluida emerge per prima e fornisce la capacità con cui, nel tempo, si costruisce quella cristallizzata.

Cattell non pensava davvero a due intelligenze separate: era sostanzial-

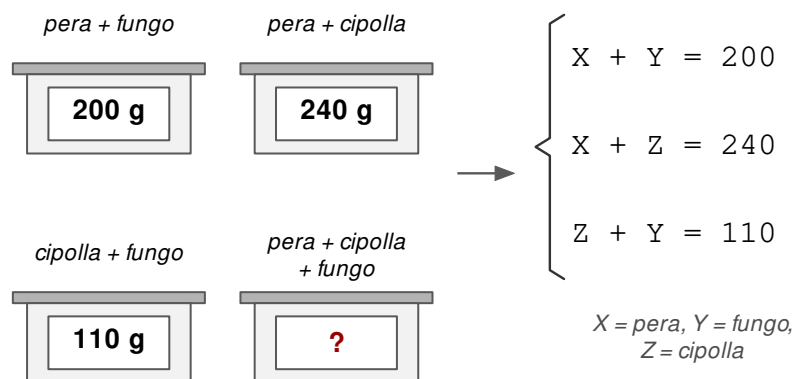


Figura 4.2: Lo stesso problema in due forme. Le pesate di tre oggetti a coppie si possono affrontare per tentativi, con la sola intelligenza fluida, oppure riconoscendovi un sistema di equazioni, una struttura appresa che rende la soluzione immediata.

mente un unitarista, e considerava Gf e Gc due modi in cui la stessa capacità si esprime. La sua distinzione ha però aperto la strada al passaggio dall'unitarismo iniziale al modello gerarchico.

4.4 La sintesi gerarchica

Il fallimento di Thurstone conteneva un'informazione preziosa. Prove del tutto diverse continuano a correlare, ma prove che coinvolgono materiale o processi simili correlano di più: verbale con verbale, spaziale con spaziale, memoria con memoria. Lo schema richiede almeno tre livelli: un fattore generale; fattori di dominio, che raggruppano prove simili; componenti specifiche di ogni prova.

Riprendiamo l'esempio sportivo del capitolo precedente. Oltre all'atleticità generale esistono capacità intermedie comuni a gruppi di discipline: la velocità per nuoto e ciclismo, il gioco di squadra per basket e calcetto. E restano componenti proprie di ciascuna disciplina (Figura 4.3). Un modello a più livelli è un compromesso fra teorie unitarie e multiple: il fattore

generale coesiste con fattori più specifici, e solo l'insieme descrive in modo completo le differenze individuali.

*componenti
«intermedie», comuni
a gruppi di prove*

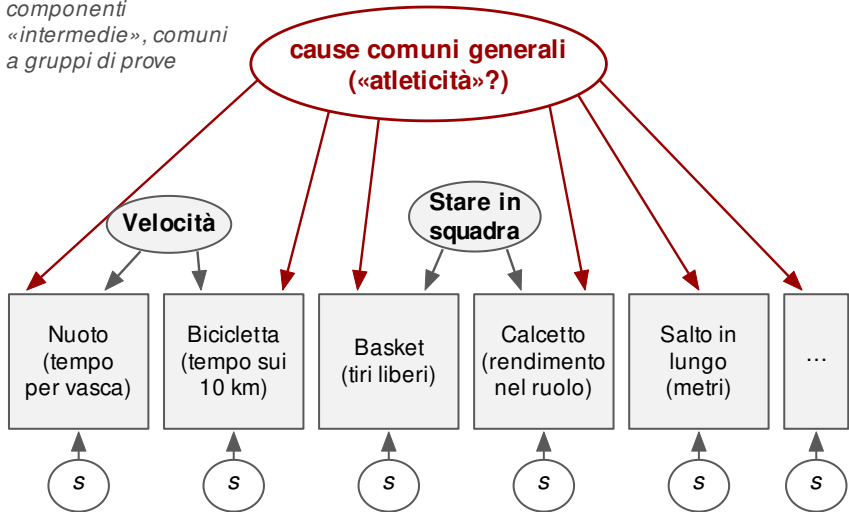


Figura 4.3: L'esempio sportivo con un livello in più. Oltre alle cause comuni generali, componenti intermedie, come la velocità o la capacità di stare in squadra, sono comuni solo a gruppi di prove; a ogni prova resta una componente specifica *s*. Esempio illustrativo, estraneo all'intelligenza.

Una delle prime formulazioni gerarchiche si deve a Philip Vernon, che sotto *g* collocava due grandi gruppi di abilità, verbali-educative e spaziali-pratiche-meccaniche, poi fattori di gruppo minori e infine le abilità specifiche di ogni prova (Vernon, 1950) (Figura 4.4).

La formulazione oggi più citata è il modello **Cattell-Horn-Carroll** (CHC). La sua storia combina tre contributi. Cattell introdusse la distinzione fra fluida e cristallizzata. John Horn, rifiutando il fattore generale, mostrò che con l'analisi fattoriale si potevano distinguere diversi altri domini. John Carroll, dopo aver rianalizzato centinaia di insiemi di dati, propose una sintesi a tre strati (Carroll, 1993). In seguito i modelli di Cattell-Horn e di Carroll sono stati integrati e ampliati (McGrew, 2009).

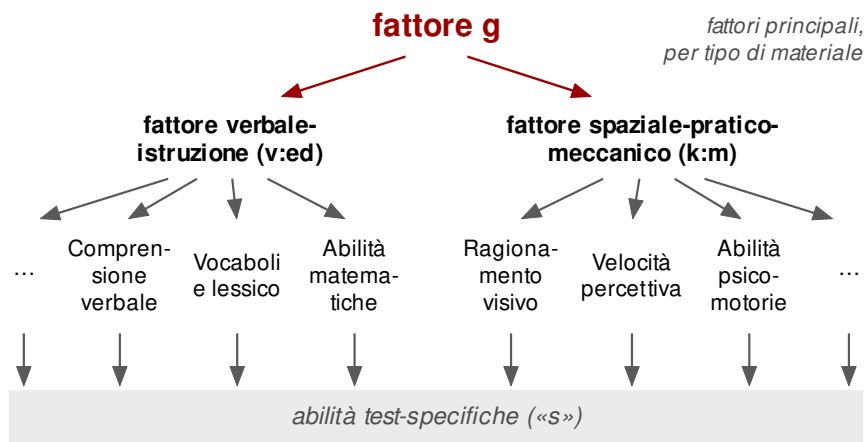


Figura 4.4: Il modello gerarchico di Vernon: il fattore g, due fattori principali che distinguono per tipo di materiale, fattori di gruppo minori e abilità specifiche di ciascuna prova. Elaborazione da Vernon (1950); elenco non completo.

Oggi «intelligenza gerarchica» è quasi sinonimo di CHC, e le batterie moderne sono costruite o riviste per riflettere questo modello. Due precisazioni sono però doverose. Non tutti i fattori hanno la stessa importanza: intelligenza fluida e cristallizzata sono centrali, altri domini, come la conoscenza quantitativa o la lettura e scrittura, sono più periferici o dipendono molto dal contesto. E il CHC è più una tassonomia delle abilità distinguibili che una teoria di come funzionano.

4.5 Leggere una prestazione su più livelli

Che cosa significa, concretamente, che l'intelligenza è gerarchica? Che la prestazione a una prova va letta su diversi livelli contemporaneamente.

Prendiamo il punteggio di una persona in una prova di vocabolario. Riflette in parte il fattore generale; in parte l'abilità verbale, comune alle prove linguistiche; in parte la conoscenza specifica delle parole presenti nella

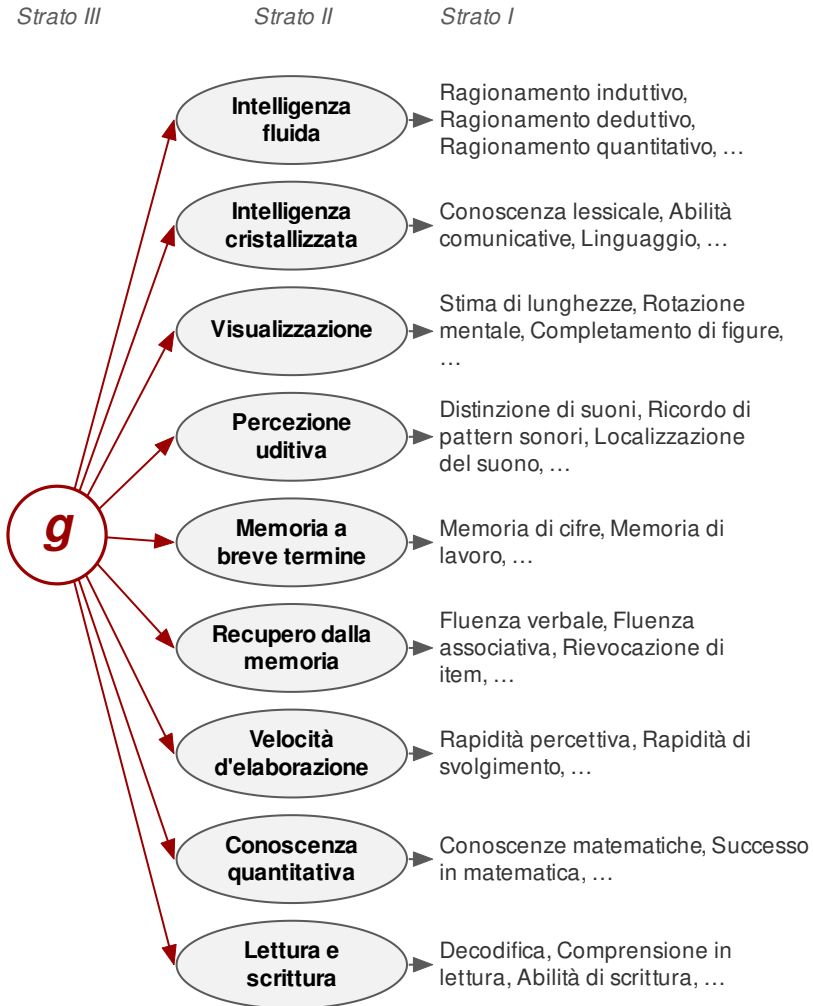


Figura 4.5: Il modello Cattell-Horn-Carroll (CHC): il fattore generale (strato III), nove abilità ampie (strato II) ed esempi di abilità ristrette (strato I). Elenco non esaustivo; i puntini indicano che ogni livello ne comprende altre.

prova. Il resto è errore: casualità, numero limitato di item, effetti transitori come una notte dormita male. Componente specifica ed errore, insieme, possono pesare per il 30–40% della varianza dei singoli subtest. L'errore da solo ne rappresenta il 10–20%: è la quota che manca a un'attendibilità, cioè alla coerenza del punteggio da una misurazione all'altra, tipicamente fra 0,80 e 0,90. Lo stesso vale per una prova di rotazione mentale, con il dominio visuospatiale al posto di quello verbale (Figura 4.6). I livelli alti sono quelli di interesse teorico; quelli bassi, spesso, non sono interpretabili.

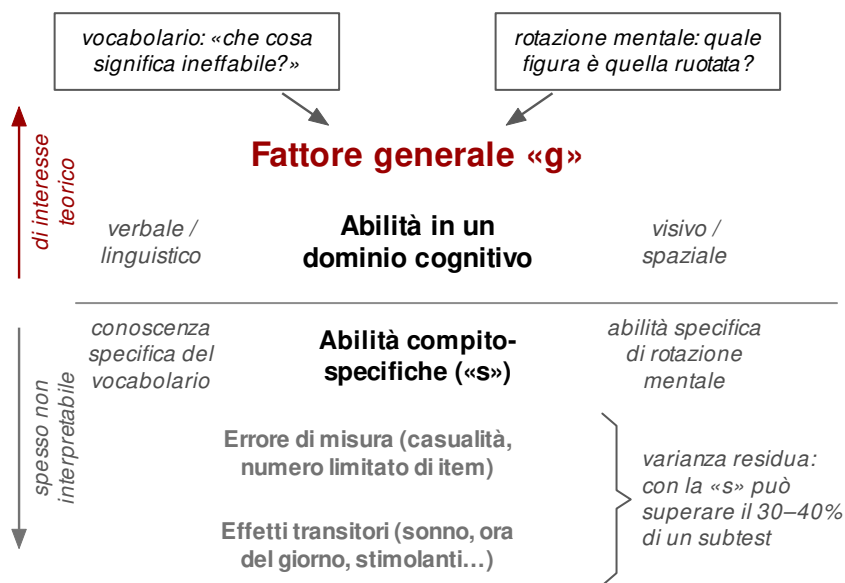


Figura 4.6: Che cosa spiega la prestazione a una prova cognitiva. Una prova di vocabolario e una di rotazione mentale riflettono lo stesso fattore generale, abilità di dominio diverse, abilità specifiche del compito e, infine, errore di misura ed effetti transitori. Andando verso l'alto cresce l'interesse teorico; verso il basso cresce la parte non interpretabile.

4.6 La prevalenza di g

In teoria c'è ampio accordo sul modello gerarchico. In pratica, la ricerca mostra che i diversi livelli non pesano allo stesso modo, e la corrente principale continua a preferire una visione di fatto unitaria (Tabella 4.1).

Tabella 4.1: Quanto conta ciascun livello della gerarchia in una batteria di intelligenza.

Livello	Peso nella batteria	Che cosa se ne può dire
Fattore g	dominante, misura attendibile	un QI di 115 è interpretabile, stabile, replicabile
Abilità di dominio	seguono molto g; al netto di g sono poco attendibili	indici come verbale = 122 e percettivo = 113 sono in gran parte ridondanti; ripetendo il test potrebbero perfino invertirsi
Abilità specifiche	pressoché non misurabili con una batteria	i singoli subtest, come matrici o cubi, non sono interpretabili da soli
Errore di misura	molto, ma senza interesse teorico	è rumore: si riduce con prove più lunghe o ripetute, spesso insostenibili nella pratica

Nelle batterie multicomponenziali il fattore generale spiega da solo la maggior parte della varianza comune ai subtest: nell'analisi della WISC-IV statunitense più di due terzi (Watkins, 2006). I fattori di dominio, al netto di g, ne spiegano poca e sono poco attendibili (Gignac & Kretzschmar, 2017); il capitolo sulla qualità della misura mostrerà come lo si verifica. È inoltre il fattore generale a prevedere con regolarità gli esiti della vita reale: risultati scolastici, prestazione lavorativa, salute e longevità (Deary, 2012a). Ed è la componente più chiaramente ereditabile: le stesse varianti genetiche agiscono in larga parte su abilità diverse (Plomin & Deary, 2015).

Questo vale per la popolazione nel suo insieme. Può non valere per individui con discrepanze molto ampie fra abilità, o in gruppi con condizioni del neurosviluppo, dove i profili di abilità sono informativi (Giofrè et al., 2017). Nella pratica clinica la gerarchia non è un dettaglio accademico: decide quanto peso dare al totale e quanto ai singoli indici.

Punto metodologico: dimenticare la causa comune

Ignorare il fattore generale può produrre conclusioni sbagliate. Supponiamo di voler capire quali abilità prevedono il successo scolastico e di inserire in un modello vocabolario, memoria di lavoro e velocità come predittori separati. Ciascuna sembrerà legata al successo, e si sarà tentati di concludere che tutte contano. Ma se le tre abilità correlano col successo soprattutto perché condividono g , il modello attribuisce a ciascuna un effetto che appartiene al fattore comune. Un modello che include il fattore generale riduce, o elimina, molti di questi effetti specifici apparenti. Lo stesso errore compare spesso in psicologia della personalità e sociale, ogni volta che si trascura la possibilità di cause comuni a più misure. La Figura 4.7 confronta i due modelli.

Per approfondire

Un esempio simulato, con il codice per riprodurlo, è nel post [Controllare per il latente](#) dell'autore.

4.7 Una sintesi ragionevole

Si può quindi rispondere alla domanda iniziale con una sintesi.

Che cos'è g ? È l'insieme delle cause comuni, in buona parte ancora ignote, che fanno sì che una persona vada meglio o peggio in molte abilità cognitive diverse. Si riassume come abilità mentale generale. Moltissimi fattori genetici, ambientali e di esperienza producono un assetto neurobiologico

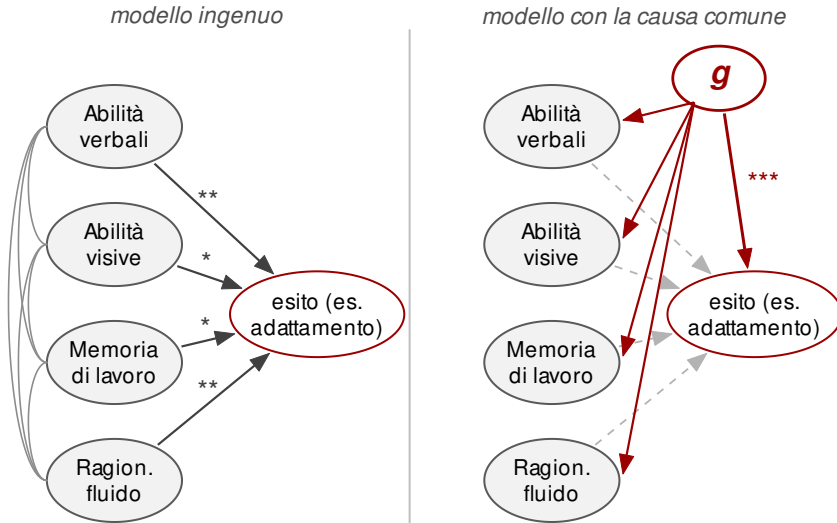


Figura 4.7: Due modelli per gli stessi dati. Nel primo, abilità correlate fra loro sembrano prevedere ciascuna l'esito per conto proprio (gli asterischi indicano effetti «significativi»). Nel secondo, una causa comune, g , sostiene le abilità e l'esito: gli effetti diretti delle singole abilità si attenuano o spariscono (linee tratteggiate).

più o meno favorevole alla prestazione cognitiva; statisticamente, il risultato si lascia riassumere da un singolo fattore. Ma g non è, ontologicamente, una cosa sola.

Come si misura? Con il QI, che aggrega le prestazioni in molte prove di tipo diverso.

Perché è importante? Perché si misura in modo affidabile, stabile e relativamente rapido, e prevede molti esiti rilevanti della vita reale, spesso rendendo superfluo l'approfondimento di questo o quel dominio specifico. Spesso, non sempre. Le abilità specifiche esistono, e in alcuni casi fanno la differenza. Ma chi vuole interpretarle deve prima fare i conti con il fattore generale.

4 Abilità multiple e modelli gerarchici

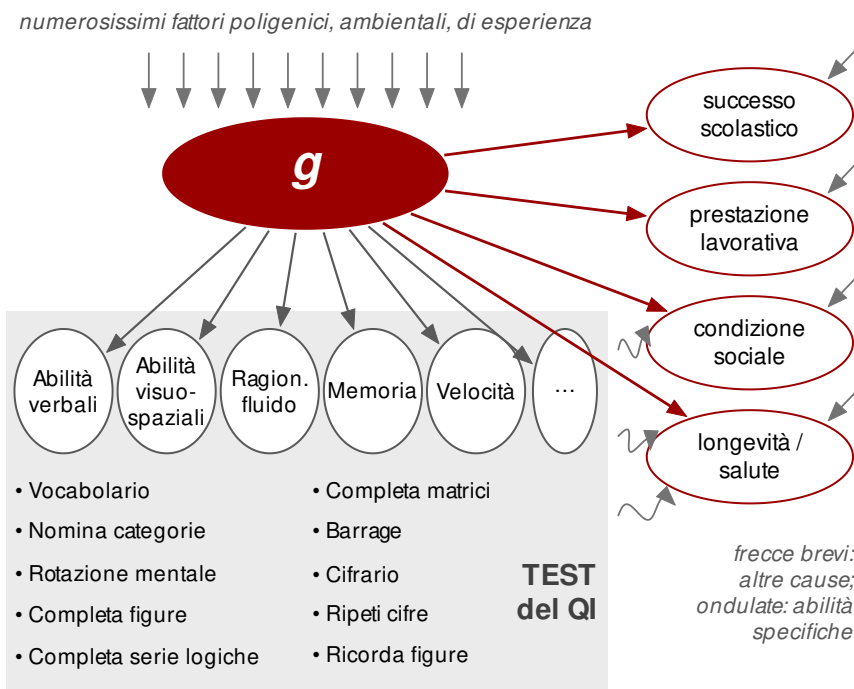


Figura 4.8: Una sintesi moderna. Numerosissimi fattori genetici, ambientali e di esperienza producono un assetto più o meno favorevole alla prestazione cognitiva generale, riassunto statisticamente da *g*; *g* sostiene le abilità di dominio, che si osservano nelle prove di un test del QI; lo stesso *g* prevede esiti importanti della vita reale, che dipendono anche da altre cause (frecce brevi) e, in misura minore, dalle abilità specifiche (frecce ondulate).

In breve

- Le teorie multiple hanno cercato di spiegare la positiva manifolds senza un fattore generale; Thurstone fallì in modo produttivo, mentre le intelligenze multiple di Gardner hanno scarso sostegno empirico.
- Intelligenza fluida e cristallizzata sono fortemente correlate,

ma seguono traiettorie diverse nel corso della vita.

- Il modello gerarchico CHC combina un fattore generale, abilità ampie e abilità ristrette: è oggi la tassonomia di riferimento.
- Nei dati il fattore generale prevale: spiega gran parte della varianza comune e degli esiti reali, mentre i fattori di dominio al netto di g sono deboli.

Riferimenti bibliografici

5 Dai primi test al testing di massa

L'intelligenza è il primo costrutto psicologico per cui sono stati costruiti test, e la storia di quei test coincide in gran parte con la nascita della psicomетria. Punteggi standardizzati, campioni normativi, somministrazioni collettive, risposte a scelta multipla: sono state inventate qui e poi estese a personalità, psicopatologia, apprendimento.

È una storia di innovazioni notevoli e di errori gravi. Alcuni errori erano tecnici, e sono stati corretti. Altri erano morali e politici, e hanno lasciato un'ombra che dura ancora. Questo capitolo segue le tappe principali fino alla Prima guerra mondiale e agli anni immediatamente successivi. Il prossimo riprende il filo dalle scale moderne.

5.1 Galton: un fallimento istruttivo

Negli anni Ottanta dell'Ottocento Francis Galton, cugino di Charles Darwin, aprì a Londra un laboratorio antropometrico. Per una piccola somma i visitatori potevano farsi misurare: diametro della testa, acutezza visiva e uditiva, capacità di discriminare pesi, tempi di reazione, forza della presa, capacità polmonare. Galton era convinto che l'intelligenza fosse in larga parte ereditaria e che si manifestasse nell'efficienza del sistema nervoso, quindi in capacità sensoriali e motorie elementari (Galton, 1869).

Il progetto aveva elementi davvero innovativi. Era il primo tentativo di misurare quantitativamente le differenze intellettive. Combinava una teoria, quella evoluzionista, con tecniche di misura. E cercava una forma di validità: voleva verificare se chi stava sopra la media nelle misure stava sopra la media anche nel successo nella vita.

Non funzionò. Le misure sensoriali e motorie correlavano poco fra loro e quasi nulla con indicatori di successo. Negli Stati Uniti James McKeen



Figura 5.1: Le tappe principali della storia dei test di intelligenza, raggruppate in quattro fasi: primi tentativi, svolta clinico-educativa, test di massa, Wechsler e clinica moderna. Le date in nero sono discusse in questo capitolo, quelle in grigio nel successivo.

Cattell riprese l'approccio con i suoi «test mentali», e un'analisi condotta nel 1901 su studenti universitari mostrò che i punteggi non correlavano con i voti (Wissler, 1901). L'idea che l'intelligenza fosse una questione di sensi acuti e riflessi pronti fu abbandonata.

C'è poi l'altro lato. Galton coniò il termine «eugenetica» e ne fu un convinto sostenitore, con posizioni apertamente razziste. Erano idee molto diffuse nell'élite scientifica del tempo. Questo spiega il contesto, non lo giustifica.

5.2 Binet: misurare per proteggere

La svolta arrivò da un problema pratico. In Francia l'istruzione primaria era diventata gratuita e obbligatoria per tutti. Nel 1904 il governo istituì una commissione per individuare i bambini che non riuscivano a trarre profitto dall'insegnamento ordinario e avevano bisogno di un sostegno particolare. Alfred Binet, psicologo e membro di una società per lo studio psicologico dell'infanzia, si dedicò alla questione.

L'obiettivo era stabilire se un bambino potesse essere inserito con profitto nella classe della sua età. Con Théodore Simon, Binet costruì una scala presentata per la prima volta nel 1905, anche a un congresso internazionale a Roma (Binet & Simon, 1904). Conteneva una trentina di prove, per lo più verbali: nominare oggetti o parti del corpo, ripetere frasi o cifre, definire parole astratte, trovare la parola mancante in una frase, copiare una figura, ordinare oggetti di peso diverso.

La differenza rispetto a Galton era netta. Binet non misurava processi elementari, ma capacità complesse come comprensione, giudizio e ragionamento, le stesse che servono a scuola. E insisteva su un punto che resta attuale: nessuna prova isolata può valutare l'intelligenza, ma solo l'insieme. Un bambino può avere una difficoltà specifica a trovare le rime o a usare le forbici senza che questo dica molto delle sue capacità generali.

Molti tipi di prova introdotti da Binet sono ancora in uso, in forme simili, nelle batterie moderne. Altri aspetti del suo pensiero sono meno noti. Binet credeva che le capacità intellettive potessero essere migliorate con esercizi appropriati, una sorta di «ortopedia mentale». Non considerava l'intelligenza una dotazione fissa. E pensava che il ragionamento del bambino seguisse processi diversi da quelli dell'adulto, anticipando un tema che Jean Piaget avrebbe sviluppato. Soprattutto, concepiva la valutazione come strumento di tutela: serviva a dare aiuto a chi ne aveva bisogno, non a escludere.

5.3 L'età mentale e il primo quoziente

Nella prima versione le prove venivano presentate in ordine di difficoltà crescente e il bambino veniva classificato in categorie qualitative, con etichette mediche dell'epoca oggi giustamente abbandonate. Nella revisione del 1908 Binet introdusse un'idea più elegante: l'**età mentale**. Se un bambino superava le prove che la maggior parte dei bambini di otto anni supera, la sua età mentale era otto anni, qualunque fosse la sua età anagrafica. Le prove erano assegnate ai livelli di età, da 3 a 13 anni, in base alla quota di bambini che le superava (Tabella 5.1).

Tabella 5.1: Esempi di prove della scala Binet-Simon del 1908, per livello di età. Elaborazione da Binet & Simon (1916, pp. 238–239).

Età	Alcune prove della scala del 1908
3 anni	indicare occhi, naso e bocca; nominare oggetti in una figura; ripetere 2 cifre; dire il proprio cognome
5 anni	dire quale di due scatole pesa di più; copiare un quadrato; contare 4 monete; ricomporre un rettangolo tagliato in due pezzi
7 anni	indicare che cosa manca in un disegno; copiare un triangolo e un rombo; ripetere 5 cifre; descrivere una figura
9 anni	dire la data completa; nominare i giorni della settimana; dare il resto di una piccola somma; ordinare 5 pesi
11 anni	trovare l'assurdità in una frase; usare 3 parole date in una sola frase; dire più di 60 parole in 3 minuti; definire termini astratti

Età	Alcune prove della scala del 1908
13 anni	ritagliare un foglio piegato e prevederne la forma; distinguere il significato di parole astratte

Le prove erano semplici da presentare e richiedevano comprensione e giudizio, non nozioni scolastiche. Nella prova delle assurdità, per esempio, l'esaminatore leggeva lentamente una frase come «Uno sfortunato ciclista si è rotto la testa ed è morto sul colpo; l'hanno portato all'ospedale e si teme molto che non possa guarire», e chiedeva che cosa ci fosse di sciocco [Binet & Simon (1916), p. 228; traduzione nostra].

Nel 1912 William Stern propose di dividere l'età mentale per l'età cronologica (W. Stern, 1912). Lewis Terman moltiplicò il rapporto per cento e lo adottò nella sua versione statunitense della scala. Nacque così il primo **quoziente di intelligenza**:

$$\text{QI di rapporto} = \frac{\text{età mentale}}{\text{età cronologica}} \times 100$$

Un bambino di 4 anni con le prestazioni tipiche dei 6 avrebbe un QI di 150; uno di 6 anni con le prestazioni tipiche dei 4 avrebbe un QI di 67.

Il nome «quoziente» viene da qui. Ma questa formula non si usa più da decenni, e il QI moderno ha un significato e un calcolo completamente diversi. Il QI di rapporto aveva difetti seri.

- **La metrica è disuguale.** Un anno di differenza pesa molto di più a 4 anni che a 11. Lo stesso scarto in età mentale non corrisponde allo stesso scarto in capacità.
- **Lo sviluppo non è lineare.** Le prestazioni nei test crescono rapidamente nell'infanzia e si appiattiscono nella tarda adolescenza. Per mantenere un QI di 150, un ragazzo di 15 anni dovrebbe avere un'età mentale di 22 anni e mezzo, un concetto privo di senso. Per gli adulti la formula non funziona.

- **Il concetto è troppo globale.** L'età mentale evoca una maturazione complessiva della persona. Dire che un adulto con sindrome di Down «ha un'età mentale di sei anni» confonde la prestazione in alcuni test con l'esperienza, gli interessi e l'autonomia di un adulto.

Oggi anche le scale per i primi anni di vita usano punteggi di deviazione dalla media dei coetanei, che vedremo nel prossimo capitolo.

Attenzione clinica: età equivalente non è età mentale

Il concetto di **età equivalente** è difendibile, con cautela, quando si riferisce alla prestazione in un test e non alla persona. Nella disabilità intellettiva moderata o grave i punteggi standardizzati arrivano spesso al limite inferiore della scala: tutti i risultati si schiacciano sul minimo, e il profilo di punti di forza e debolezza si perde. Confrontare una persona con una popolazione di riferimento da cui è molto distante ha inoltre poco senso teorico. In questi casi dire che la prestazione in un certo subtest corrisponde a quella tipica di una certa età può aiutare la descrizione e la comunicazione, soprattutto per singole aree di funzionamento. Calcolare un QI di rapporto, invece, non ha giustificazione in nessun caso.

5.4 Terman e il lato oscuro

Lewis Terman portò la scala di Binet all'Università di Stanford, la adattò, la standardizzò su un campione statunitense e la pubblicò nel 1916 come Stanford-Binet (Terman, 1916). Il test ebbe un successo enorme e fu rieditato più volte, fino alla quinta edizione del 2003.

Terman, però, ne sostenne l'uso a fini eugenetici. Nello stesso volume del 1916 attribuiva a cause razziali i punteggi bassi di bambini di origine messicana, afroamericana e di altri gruppi, ne raccomandava la separazione in classi speciali e presentava la loro presunta fertilità elevata come un problema eugenetico. Fu membro di organizzazioni che promuovevano la sterilizzazione obbligatoria delle persone con basso livello intellettuale, praticata

in molti stati statunitensi nei primi decenni del Novecento e avallata dalla Corte Suprema nel 1927.

Lo strumento di Binet, pensato per proteggere, veniva così usato per classificare e discriminare. Il contrasto illustra un punto generale: la validità di un test non dice nulla sulla legittimità dell'uso che se ne fa. E le premesse ideologiche di chi interpreta i punteggi pesano almeno quanto la qualità dello strumento.

La Stanford-Binet aveva anche limiti tecnici. Era nata per l'età evolutiva e restava legata all'età mentale, inadatta agli adulti. Era orientata a individuare deficit e distingueva male i livelli alti. E poggiava molto su prove verbali, che penalizzavano chi aveva avuto poca istruzione o non padroneggiava la lingua.

5.5 La guerra e il testing di massa

Nel 1917 gli Stati Uniti entrarono nella Prima guerra mondiale. L'esercito doveva assegnare rapidamente ruoli a oltre un milione e mezzo di reclute. Un test individuale come la Stanford-Binet richiedeva troppo tempo. Molte reclute erano analfabete o non parlavano inglese. E serviva una scala adatta agli adulti.

Robert Yerkes, allora presidente dell'American Psychological Association, riunì un gruppo di psicologi, fra cui Terman, per costruire test collettivi a carta e matita (Yerkes, 1921). Nacquero così due strumenti:

- **l'Army Alpha**, per chi sapeva leggere e scrivere in inglese, con prove di vocabolario, serie numeriche, comprensione di istruzioni verbali, analogie;
- **l'Army Beta**, per analfabeti e non anglofoni, con istruzioni mimate a gesti e prove non verbali: labirinti, completamento di figure, cifrari, sequenze di simboli.

Nel Beta l'esaminatore non spiegava a parole. Guardava un disegno alla lavagna, si mostrava perplesso, poi «capiva» e completava la parte mancante; le reclute dovevano fare lo stesso sul proprio foglio (Figura 5.2).

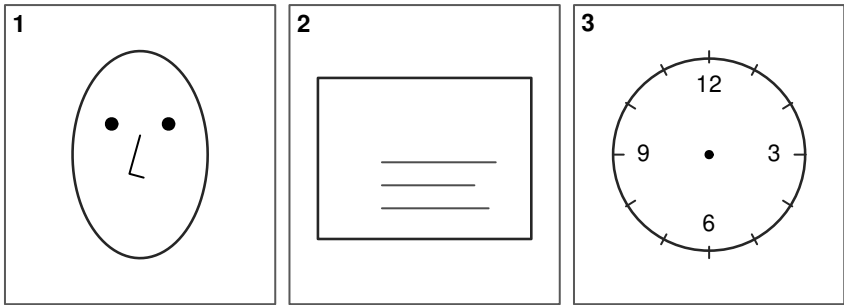


Figura 5.2: Tre item disegnati sul modello della prova «Cosa manca?» dell’Army Beta: a ogni figura manca un particolare da aggiungere. Disegni originali; il formato della prova è descritto da Yerkes (1921).

I punteggi grezzi, fino a 212 punti per l’Alpha e 118 per il Beta, venivano convertiti in fasce, dalla A, «molto superiore», alla E, corrispondenti grosso modo a percentuali fisse della distribuzione (Tabella 5.2). Le fasce orientavano l’assegnazione: ufficiali dalle prime, ruoli semplici o esonero dalle ultime. Era un sistema rudimentale di punteggi basati sulla posizione relativa, il primo passo verso i punteggi di deviazione.

Tabella 5.2: Le fasce di punteggio dei test dell’esercito statunitense nella Prima guerra mondiale. Elaborazione da Yerkes (1921).

Fascia	Etichetta	Quota approssimativa	Impiego suggerito
A	molto superiore	4–5%	ufficiali con elevate responsabilità
B	superiore	8–10%	ufficiali con minori responsabilità, sottufficiali
C+	media alta	15–18%	sottufficiali, alcuni ruoli di guida

Fascia	Etichetta	Quota approssimativa	Impiego suggerito
C	nella media	25%	soldati semplici con piccole responsabilità
C–	media bassa	25%	soldati semplici, compiti autonomi ma ripetitivi
D	inferiore	15%	soldati semplici, compiti semplici sotto supervisione
D–, E	molto inferiore	pochi	ai limiti del servizio regolare; ruoli di supporto o esonero

Le innovazioni furono notevoli: somministrazione collettiva, risposte a scelta multipla, scoring rapido, standardizzazione su numeri enormi, una prova non verbale per non sottovalutare chi non conosceva la lingua. Molte di queste soluzioni sono ancora alla base della valutazione su larga scala.

I limiti furono altrettanto evidenti. Anche maneggiare un modulo a carta e matita richiede familiarità con la scuola. Le condizioni di somministrazione erano spesso caotiche, e reclute che avrebbero dovuto svolgere il Beta finivano per svolgere l'Alpha. Alcune prove dell'Alpha erano domande di cultura generale statunitense, che misuravano soprattutto quanto a lungo si era vissuti nel paese. Nella prova di informazione, per esempio, bisognava sottolineare la risposta giusta fra quattro: «Il Wyandotte è una razza di: cavalli, polli, bovini, granito»; «Crisco è: un medicinale, un disinfettante, un dentifricio, un prodotto alimentare»; «Velvet Joe compare nelle pubblicità di: dentifricio in polvere, tessuti, tabacco, sapone» [Yerkes (1921), p. 228; traduzione nostra]. Per un immigrato arrivato da poco, sa-

pere che il Wyandotte è una razza di polli non aveva nulla a che fare con l'intelligenza.

5.6 Dai test ai pregiudizi

I dati dell'esercito furono poi usati per sostenere tesi sulla superiorità di alcuni gruppi nazionali. Nel 1923 lo psicologo Carl Brigham pubblicò un'analisi che classificava gli immigrati per paese d'origine, trovando punteggi più bassi fra quelli provenienti dall'Europa meridionale e orientale, italiani compresi (Brigham, 1923). I suoi stessi dati mostravano che il punteggio cresceva con gli anni di residenza negli Stati Uniti, un indizio evidente del peso della lingua e dell'esposizione culturale; Brigham lo interpretò invece come segno di un peggioramento delle ondate migratorie più recenti. Pochi anni dopo ritrattò pubblicamente quelle conclusioni (Brigham, 1930).

Nel 1924 il Congresso statunitense approvò una legge che limitava drasticamente l'immigrazione da quei paesi. Il ruolo dei test in quella decisione viene spesso esagerato: un'analisi dei dibattiti parlamentari ha mostrato che i risultati dei test furono citati raramente e che le motivazioni principali erano di altro tipo (Snyderman & Herrnstein, 1983). Ma il clima intellettuale in cui i test venivano interpretati era quello. E alcuni dei loro promotori lo alimentarono attivamente.

Brigham, intanto, aveva adattato l'esperienza dell'esercito a un uso civile. Nel 1926 nacque lo **Scholastic Aptitude Test** (SAT), concepito per una selezione universitaria che premiasse il merito anziché la famiglia d'origine. Le sue caratteristiche, cioè scelta multipla, somministrazione collettiva, standardizzazione su campioni ampi, sono state ereditate dalle grandi indagini educative contemporanee, come PISA e INVALSI. Queste ultime non valutano singoli individui, ma sistemi scolastici, e sono più legate ai programmi; i loro punteggi, tuttavia, sono dominati da un fattore generale molto simile a quello dei test cognitivi (Pokropek et al., 2022).

5.7 Che cosa resta

La prima stagione dei test lascia un'eredità doppia. Da un lato un insieme di soluzioni tecniche che hanno funzionato: prove complesse anziché elementari, aggregazione di molte prove, confronto con un gruppo di riferimento, versioni non verbali, somministrazione standardizzata. Dall'altro una serie di errori: un quoziente costruito su una metrica inadeguata, prove contaminate dalla lingua e dalla scolarità, interpretazioni razziali di differenze che riflettevano soprattutto esperienze diverse.

Chi oggi diffida dei test di intelligenza ha spesso in mente questa storia. Fa bene a ricordarla. Il passo successivo è chiedersi quali di questi errori siano stati corretti e quali richiedano ancora vigilanza. Il prossimo capitolo comincia da qui.

In breve

- Il tentativo di Galton di misurare l'intelligenza con capacità sensoriali e motorie fallì: le misure non prevedevano il successo scolastico o professionale.
- Binet costruì la prima scala efficace usando prove complesse e aggregandole, con lo scopo di individuare i bambini bisognosi di sostegno.
- Il QI di rapporto di Stern e Terman aveva difetti gravi e non si usa più; l'età equivalente di singole prestazioni resta utile solo in casi specifici.
- I test di massa dell'esercito statunitense introdussero innovazioni durature ma furono interpretati, con Terman e Brigham, in chiave eugenetica e razziale.

Riferimenti bibliografici

6 Le batterie moderne di intelligenza

Il QI che si legge oggi in un referto non ha quasi nulla in comune con il quoziente di Stern. Non è un rapporto fra età, non si basa su un'età mentale, e vale allo stesso modo a 8 e a 80 anni. È il risultato di una seconda rivoluzione nella storia dei test, iniziata alla fine degli anni Trenta in un ospedale psichiatrico di New York.

Questo capitolo descrive come sono costruite le batterie moderne, che cosa significano i loro punteggi e quali sono gli strumenti più diffusi. Non riproduce item di test commerciali: quando servono esempi, sono costruiti ad hoc per illustrare il tipo di prova, non equivalenti a item reali.

6.1 Un piccolo glossario

Il linguaggio dei test usa alcuni termini con un significato preciso (Tabella 6.1). Una **batteria** è lo strumento completo; i suoi **indici** sono i punteggi standardizzati che stimano un costrutto, dal QI totale agli indici di dominio; ogni indice aggrega alcuni **subtest**, cioè singole prove; ogni subtest è una sequenza di **item**, i singoli problemi, ordinati per difficoltà.

Tabella 6.1: Mini glossario dei test.

Termine	Che cos'è	Esempi
Batteria	Strumento completo per valutare un'area del funzionamento psicologico, articolato in indici e subtest; per tradizione spesso chiamato «scala»	WISC-V, IDS-2
Indice	Punteggio standardizzato che stima un costrutto latente, di solito aggregando più subtest	QI totale, indice di comprensione verbale
Subtest	Singola prova che misura una particolare abilità, fatta di item ordinati per difficoltà	Vocabolario, disegno con cubi
Item	Singolo problema di un subtest, a risposta giusta o sbagliata o valutato su pochi gradi di accuratezza	«Che cosa significa questa parola?»

6.2 Che aspetto hanno le prove

Gli item delle batterie in uso sono riservati: diffonderli renderebbe il test meno valido per chi lo incontrerà. Si può però mostrare la *logica* dei diversi tipi di prova con item costruiti apposta, che non coincidono con quelli di alcuno strumento pubblicato. La Tabella 6.2 ne raccoglie alcuni, ordinati per area del modello CHC; gli esempi visivi (matrici, rotazione mentale, disegno con cubi, codifica di simboli) sono nella Figura 2.3.

Tabella 6.2: Esempi di item costruiti per questo libro, sul modello di tipi di prova diffusi nelle batterie di intelligenza. Le soluzioni non sono uniche per tutti gli item verbali: nei test reali le risposte si valutano con criteri dettagliati, spesso con punteggi parziali (per esempio 0, 1 o 2 punti per una somiglianza vaga, concreta o astratta).

Area (CHC)	Tipo di prova	Esempio di item
Conoscenza (<i>Gc</i>)	Vocabolario	Che cosa significa «sporadico»?
Conoscenza (<i>Gc</i>)	Somiglianze	In che cosa si somigliano un ponte e una stretta di mano?
Conoscenza (<i>Gc</i>)	Informazione	Perché in inverno le giornate sono più corte?
Ragionamento fluido (<i>Gf</i>)	Serie numeriche	Quale numero continua la serie 2, 3, 5, 8, 12, ...?
Ragionamento fluido (<i>Gf</i>)	Analogie verbali	Quanto sta a mano come calzino sta a ...?
Ragionamento fluido (<i>Gf</i>)	Matrici	Quale figura completa la tabella? (risposta a scelta fra sei o otto)
Conoscenza quantitativa (<i>Gq</i>)	Problemi aritmetici a mente	Un libro costa 12 euro; con lo sconto del 25%, quanto si paga?
Memoria di lavoro (<i>Gwm</i>)	Cifre all'indietro	Ascolta «4 – 9 – 2 – 7» e ripeti la serie al contrario
Memoria di lavoro (<i>Gwm</i>)	Riordino di lettere e numeri	Ascolta «L – 3 – B – 1» e ripeti prima i numeri in ordine crescente, poi le lettere in ordine alfabetico
Elaborazione visiva (<i>Gv</i>)	Rotazione mentale	Quali di queste figure, ruotate, coincidono con il modello?
Velocità di elaborazione (<i>Gs</i>)	Ricerca di simboli	In due minuti, barra tutti i ▲ in una lunga fila di ▲ △ ■ ● □ ○
Memoria a lungo termine (<i>Gl</i>)	Rievocazione differita	Dopo venti minuti, ripeti le parole della lista ascoltata all'inizio

Due dettagli contano più dell'aspetto degli item. Il primo è l'ordine: dentro ogni subtest gli item vanno dal facilissimo al molto difficile, e la prova si interrompe dopo un certo numero di errori consecutivi. Così la stessa prova serve a un bambino di sei anni e a un adulto, e nessuno passa mezz'ora su problemi tutti troppo facili o tutti impossibili. Il secondo è il tempo: in molte prove visive e di velocità conta quanti item si completano entro un limite, e in alcune si guadagnano punti in più rispondendo in fretta. Per questo la stessa abilità, misurata con o senza limite di tempo, può dare risultati diversi.

6.3 Wechsler e il **QI di deviazione**

David Wechsler aveva lavorato da giovane alla somministrazione dei test dell'esercito statunitense. Negli anni Trenta era psicologo clinico al Bellevue Hospital di New York, dove doveva valutare adulti molto diversi per età, storia e sintomi: per stabilirne l'autonomia, la capacità di lavorare o di testimoniare, il bisogno di tutela, le possibilità di recupero. La Stanford-Binet non era adatta. Nel 1939 Wechsler pubblicò la Wechsler-Bellevue, la prima vera batteria clinica per adulti (Wechsler, 1939). Il campione di standardizzazione, circa 1.700 persone, era poco rappresentativo: per lo più newyorkesi, bianchi e con istruzione superiore alla media. Ma l'impianto era nuovo.

Le innovazioni erano due. La prima era l'equilibrio fra prove **verbali** e prove **di performance**, cioè non verbali, un'idea che riprendeva la distinzione fra Army Alpha e Army Beta. La batteria forniva un **QI verbale**, un **QI di performance** e un **QI totale**. Questa struttura è rimasta sostanzialmente invariata fino agli anni Novanta.

La seconda, più importante, era il **QI di deviazione**. Invece di confrontare età mentale e cronologica, Wechsler confrontava la prestazione di una persona con quella dei coetanei del campione normativo, e la esprimeva come distanza dalla loro media su una scala con media 100 e deviazione standard 15. È esattamente la logica dei punti *z*: la distanza dalla media si conta in deviazioni standard, cioè nella dispersione tipica dei punteggi attorno alla media. Il **QI di deviazione** funziona a ogni età e ha lo stesso significato in

qualunque punto della vita: 115 indica sempre una deviazione standard sopra la media dei coetanei. Questa metrica è diventata universale ed è usata ancora oggi per quasi tutti i test psicologici.

Strumenti → *Punti z, QI di deviazione, punteggi ponderati e percentili, con le loro trappole: Distribuzioni, punti z e punteggi standardizzati, subito dopo questa parte.*

Wechsler aveva un'ispirazione pragmatica e in fondo unitarista. Voleva anzitutto misurare bene il QI totale. Considerava i subtest misure diverse della stessa intelligenza, non misure di intelligenze diverse, e la divisione fra parte verbale e di performance solo uno dei modi possibili di raggrupparli (Wechsler, 1958). Gli bastava che la batteria funzionasse.

6.4 Dagli indici di Wechsler al CHC

Le scale Wechsler si sono moltiplicate e sono state riviste più volte: la WAIS per adulti, la WISC per l'età scolare, la WPPSI per l'età prescolare. WISC e WAIS sono oggi alla quinta edizione, la WPPSI alla quarta; le edizioni disponibili in Italia possono essere precedenti. Restano le batterie più usate al mondo per la valutazione cognitiva globale.

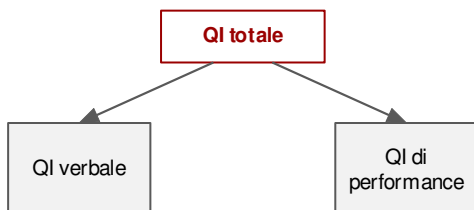
Nel tempo la struttura si è articolata. La distinzione fra QI verbale e QI di performance è stata sostituita, nelle edizioni degli anni Duemila, da quattro indici di ispirazione cognitiva e neuropsicologica: comprensione verbale, ragionamento percettivo, memoria di lavoro, velocità di elaborazione. Le quinte edizioni hanno separato il ragionamento percettivo in un indice visuospatiale e uno di ragionamento fluido, con un riferimento più esplicito al modello CHC. La Figura 6.1 riassume il passaggio.

Le convenzioni metriche sono ormai universali: i subtest hanno punteggi ponderati con media 10 e deviazione standard 3; gli indici, compreso il QI totale, media 100 e deviazione standard 15.

La struttura di una batteria come la WISC-V è dunque gerarchica: un QI totale, cinque indici primari, una decina di subtest principali e altri subtest supplementari. I principali tipi di prova sono questi.

1^a-3^a edizione

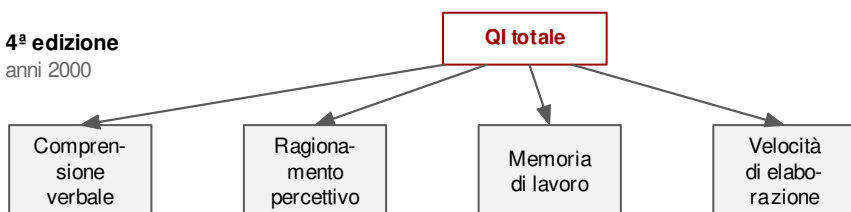
anni '30-'90



Wechsler: «misure diverse dell'intelligenza, non misure di tipi diversi di intelligenza»

4^a edizione

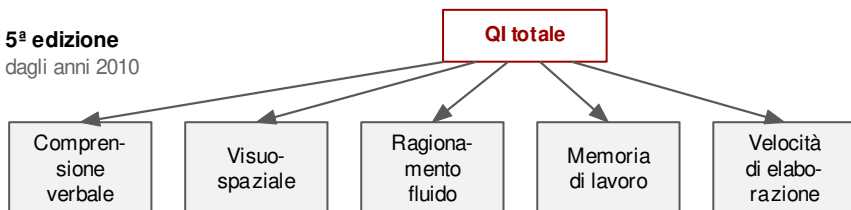
anni 2000



indici di ispirazione cognitiva e neuropsicologica: profilo più articolato, discrepanze interne

5^a edizione

dagli anni 2010



rimando esplicito al CHC (ragionamento fluido); memoria di lavoro verbale e visiva; alcuni subtest spostati

Figura 6.1: Come è cambiata la struttura delle scale Wechsler, con una progressiva aggiunta di indici e articolazioni fattoriali migliori: dal QI verbale e di performance delle prime tre edizioni ai quattro indici della quarta, ai cinque indici primari della quinta, più vicini al modello CHC. Le edizioni intermedie (una terza edizione orientata ai processi, estensioni della quarta) hanno preparato i passaggi.

Tabella 6.3: Indici primari della WISC-V, subtest e relativi tipi di prova. I subtest primari, in corsivo, servono a calcolare i cinque indici primari; il QI totale ne usa sette.

Indice	Che cosa valuta	Subtest (primari in corsivo)	Tipo di prova
Comprensione verbale	Conoscenza delle parole, dei concetti e del mondo	<i>Somiglianze, Vocabolario, Informazione, Comprensione</i>	Spiegare che cosa hanno in comune due concetti; definire parole
Visuospaziale	Analisi di dettagli visivi e relazioni spaziali	<i>Disegno con cubi, Puzzle</i>	Riprodurre un modello con cubi colorati; scegliere i pezzi che compongono una figura
Ragionamento fluido	Ragionamento induttivo e deduttivo, astrazione di regole	<i>Matrici, Confronto di pesi, Concetti illustrati, Ragionamento aritmetico</i>	Completare matrici di figure; trovare ciò che bilancia una serie di pesi
Memoria di lavoro	Mantenere e manipolare informazioni verbali o visive	<i>Memoria di cifre, Memoria di immagini, Riordinamento di lettere e numeri</i>	Ripetere sequenze di cifre in avanti, all'indietro o riordinate; riconoscere immagini viste
Velocità di elaborazione	Rapidità ed efficienza in compiti semplici	<i>Cifrario, Ricerca di simboli, Cancellazione</i>	Associare simboli a cifre; cercare simboli bersaglio

Un esempio originale chiarisce il tipo di richiesta senza riprodurre item reali. In una prova di somiglianze si potrebbe chiedere in che cosa si assomigliano una bussola e un orologio: entrambi sono strumenti che servono

a orientarsi, uno nello spazio e l'altro nel tempo. La risposta richiede di trovare la categoria astratta che li accomuna, non di descriverli.

La struttura a cinque fattori delle scale Wechsler più recenti è stata contestata. Alcune rianalisi dei dati normativi trovano che quattro fattori di dominio, più un fattore generale molto forte, descrivano i dati altrettanto bene, e che la varianza specifica di molti indici sia modesta (Canivez et al., 2017). È lo stesso problema incontrato con i modelli gerarchici, che il capitolo sulla qualità della misura riprenderà: i domini esistono, ma al netto di *g* pesano poco.

6.5 Leggere un profilo

Una batteria multicomponentiale restituisce molti numeri. La tentazione è leggerli tutti: un indice verbale di 112 e uno visuospaziale di 96 sembrano raccontare una storia. A volte la raccontano. Spesso, però, raccontano anche l'errore di misura.

Il QI totale è la stima più attendibile e stabile. Gli indici di dominio seguono in gran parte il fattore generale e, considerati da soli, sono meno attendibili: se si ripetesse la valutazione, la differenza fra due indici potrebbe ridursi o addirittura invertirsi. I singoli subtest, ancora meno affidabili, non vanno interpretati isolatamente. Il capitolo sulla qualità della misura mostrerà perché.

Questo non significa che i profili siano inutili. Possono essere informativi quando la discrepanza è ampia, quando è coerente con altre informazioni sul caso e quando esiste un profilo atteso documentato per una certa condizione. In queste circostanze, per esempio, si può preferire un indice di abilità generale che escluda memoria di lavoro e velocità, se queste componenti sono compromesse per ragioni specifiche e rischiano di sottostimare il ragionamento. Ne riparleremo nel capitolo sulla valutazione nel neurosviluppo.

6.6 Le matrici di Raven e i test monocomponentiali

Accanto alle grandi batterie esiste una tradizione diversa, di origine britannica. John Raven, allievo di Spearman, voleva un test di ragionamento complesso per la ricerca, che non dipendesse da competenze verbali. Nel 1938 pubblicò le **Standard Progressive Matrices**: una serie di matrici di figure in cui manca un elemento, da scegliere fra diverse alternative in base a una regola da scoprire (Raven, 2000).

Durante la Seconda guerra mondiale l'esercito britannico le usò per lo screening di massa, grazie alla facilità di somministrazione collettiva, con scopi simili a quelli dei test statunitensi della guerra precedente. Seguirono una versione a colori per bambini, anziani e persone con difficoltà cognitive, nel 1947, e una versione avanzata per livelli di abilità elevati, nel 1962. Le matrici sono oggi considerate una misura tipica di intelligenza fluida e uno dei migliori singoli indicatori di *g*.

Test come le matrici di Raven o il Culture Fair Test di Cattell sono **monocomponentiali**: forniscono una sola stima del livello generale. Sono rapidi, spesso somministrabili in gruppo, e utili per la ricerca e lo screening. Furono presentati a lungo come «liberi dalla cultura». Nessun test lo è davvero: il formato a scelta multipla, la familiarità con figure astratte e con la logica delle prove scolastiche contano. Il capitolo sull'effetto Flynn mostrerà che proprio le matrici sono fra le prove più sensibili ai cambiamenti storici. Meglio parlare di test **ridotti nella componente culturale**.

6.7 Test equi?

L'accusa più antica ai test di intelligenza è che siano ingiusti verso alcuni gruppi. La storia raccontata nel capitolo precedente le dà più di una ragione. Ma la domanda va posta nei termini giusti. Nel linguaggio della psicometria un test è **distorto** (*biased*) quando lo stesso punteggio non ha lo stesso significato in gruppi diversi: perché a parità di abilità un gruppo ottiene punteggi più bassi in alcune prove, cioè manca l'invarianza di misura, oppure perché il punteggio prevede gli esiti in modo diverso nei due gruppi (Meredith, 1993). Una differenza fra le medie, da sola, non è prova

di distorsione. Un termometro che segna temperature diverse a Palermo e a Bolzano non è per questo un termometro sbagliato.

Per gruppi che condividono lingua e scolarizzazione, le batterie moderne superano in genere queste verifiche: la struttura dei punteggi è la stessa e la loro capacità predittiva è simile (Warne, 2020). I problemi seri nascono quando lingua, scolarità e familiarità con le prove differiscono, cioè proprio nelle condizioni in cui fallirono i test dell'esercito statunitense. Le differenze medie fra gruppi definiti per etnia o nazionalità sollevano, oltre a queste, domande sulle cause a cui l'ereditabilità stimata dentro ciascun gruppo non può rispondere da sola; il caso del divario Nord-Sud, che si vedrà nel capitolo sulla scuola, mostra quanto fragili possano essere i dati su cui si costruiscono certe conclusioni.

Il caso dei sessi è istruttivo. Nell'abilità generale le differenze medie fra maschi e femmine sono trascurabili; ne esistono alcune in abilità specifiche, come un vantaggio maschile nella rotazione mentale e uno femminile in alcune prove verbali e di velocità (Halpern et al., 2007). Va però ricordato che le batterie vengono costruite scartando gli item che funzionano in modo diverso nei due sessi, sicché l'uguaglianza dei totali è in parte una scelta di costruzione. Più solido è un altro risultato. Nell'intera coorte scozzese testata nel 1932, quasi tutti i bambini di undici anni, le medie erano praticamente uguali, ma la variabilità era maggiore fra i maschi, più numerosi in entrambe le code (Johnson et al., 2008).

Una spiegazione molto popolare delle differenze fra gruppi è la **minaccia dello stereotipo**: chi appartiene a un gruppo descritto come meno capace renderebbe peggio quando lo stereotipo viene evocato, per ansia o distrazione. L'idea è plausibile e ha avuto un enorme successo. Le evidenze sono più deboli della reputazione. Per le ragazze in matematica, una meta-analisi ha trovato un effetto piccolo, con segni di pubblicazione selettiva (Flore & Wicherts, 2015); e gli effetti si riducono ancora negli studi che riproducono le condizioni di una valutazione reale (Shewach et al., 2019). È una traiettoria che il capitolo sul potenziamento cognitivo mostrerà più volte: un risultato elegante, studi piccoli, una letteratura che si ridimensiona quando arrivano i campioni grandi.

6.8 Il panorama italiano e le tendenze recenti

Oltre alle scale Wechsler, in Italia sono disponibili diversi strumenti. Alcune batterie multicomponenziali, come la IDS-2, sono costruite esplicitamente sul modello CHC e coprono anche aree diverse dall'intelligenza in senso stretto, come funzioni esecutive e competenze scolastiche. Esistono batterie brevi, con pochi subtest, che stimano in modo rapido e robusto il QI totale; strumenti interamente non verbali, anche nelle istruzioni, utili per chi ha difficoltà linguistiche o per alcune condizioni del neurosviluppo; test monocomponenziali di ragionamento fluido per l'età evolutiva o per la selezione del personale. Un elenco ragionato, con caratteristiche e limiti di ciascuno, è nella sezione del sito dedicata alla **professione**.

Due tendenze stanno cambiando il settore. La prima è la diffusione di modelli psicometrici più avanzati della teoria classica, come la **teoria della risposta all'item**, che stima l'abilità tenendo conto delle caratteristiche di ciascun item. Su questa base si costruiscono **test adattivi computerizzati**, che scelgono l'item successivo in funzione delle risposte già date e raggiungono una stima precisa con meno domande. La seconda è il passaggio allo scoring tramite piattaforme online. È comodo, ma crea un problema nuovo: quando le tabelle normative non sono pubblicate, il professionista non può verificare come il punteggio grezzo sia stato trasformato (Frinco et al., 2024; Zanon & Priftis, 2025). Un numero ottenuto da una procedura opaca è difficile da difendere, in clinica come in tribunale.

6.9 Uno strumento, non un oracolo

Le batterie moderne sono il risultato di un secolo di correzioni. La metrica di deviazione ha eliminato i difetti del QI di rapporto. La struttura gerarchica permette di stimare sia il livello generale sia i domini. La standardizzazione su campioni ampi rende i confronti più equi. Restano però limiti che nessuna revisione elimina del tutto: l'errore di misura, la debolezza degli indici specifici, la dipendenza dalla lingua e dalla scolarità, la necessità di norme aggiornate e trasparenti. Una batteria ben usata fornisce una

stima solida; che cosa quella stima significhi per la persona valutata non lo dice il manuale.

In breve

- Wechsler introdusse il QI di deviazione, con media 100 e deviazione standard 15 rispetto ai coetanei: una metrica valida a ogni età e oggi universale.
- Le batterie moderne sono gerarchiche: QI totale, indici di dominio, subtest e item; le edizioni recenti si richiamano esplicitamente al modello CHC.
- Il QI totale è la stima più solida; le discrepanze fra indici vanno interpretate solo se ampie e coerenti con altre evidenze.
- Un test è distorto quando lo stesso punteggio non ha lo stesso significato in gruppi diversi, non quando le medie differiscono; la minaccia dello stereotipo spiega poco delle differenze osservate.
- Nessun test è libero dalla cultura; norme trasparenti e aggiornate sono una condizione per un uso professionale difendibile.

Riferimenti bibliografici

Parte III

Strumenti I · Punteggi e distribuzioni

7 Distribuzioni, punti z e punteggi standardizzati

Strumenti. Le due parti «Strumenti» raccolgono le basi tecniche del libro. I loro capitoli si possono leggere da soli, in qualsiasi momento, oppure saltare: i capitoli concettuali restano comprensibili e vi rimandano, con brevi note in verde, quando un passaggio richiede uno strumento in più.

Questo capitolo tratta la curva gaussiana, i punti z e le scale dei punteggi (QI, ponderati, punti T, percentili), la grandezza di un effetto e i casi in cui la normalità è una scelta teorica più che un dato. Non presuppone nulla. Serve soprattutto per leggere un profilo di punteggi, per capire la grandezza dell'effetto Flynn e per le soglie diagnostiche del neurosviluppo.

Un QI di 130 non indica una quantità di intelligenza. Indica una posizione: due deviazioni standard sopra la media della popolazione di riferimento, un livello raggiunto o superato da circa due persone su cento. Tutta la misura psicologica funziona così. Non avendo un metro naturale per le capacità mentali, usa come unità la variabilità fra le persone.

In psicologia si lavora con **variabili** perché interessano cose che variano. A volte le si manipola sperimentalmente: ascoltare musica veloce o lenta mentre si studia cambia l'apprendimento? Un training di memoria di lavoro migliora la matematica dei bambini? Altre volte le si osserva correlazionalmente: essere musicisti si associa a punteggi più alti nei test cognitivi? Il tempo davanti agli schermi si associa a una minore attenzione in età prescolare? In entrambi i casi serve un modo per descrivere come si distribuiscono i valori e quanto una persona si discosta dalle altre. Questo capitolo parte dalla curva gaussiana, il modello di distribuzione su cui poggia gran parte della psicometria, e arriva alle scale che si incontrano nei referti: punti z, QI, punteggi ponderati, punti T, percentili. Poi mostra dove il modello è una buona approssimazione e dove è una scelta teorica che conviene riconoscere come tale.

7.1 Una curva che si trova ovunque

Moltissimi test e modelli statistici poggiano su un'assunzione di normalità. Il motivo è che molte variabili osservate in natura, fisiche e biologiche, si distribuiscono davvero in modo simile: la statura degli adulti, il peso alla nascita, la pressione sanguigna a riposo, le dimensioni di foglie e organi, la durata di vita delle cellule, l'attività di fondo del sistema nervoso, gli errori nelle misure ripetute di una stessa grandezza. La maggior parte dei casi si concentra attorno a un valore centrale; allontanandosi in entrambe le direzioni i casi diventano sempre più rari. Il grafico ha la forma di una campana (Figura 7.1). Col tempo è diventata un'assunzione generale, a volte un po' troppo aprioristica, sulla distribuzione di tutto.

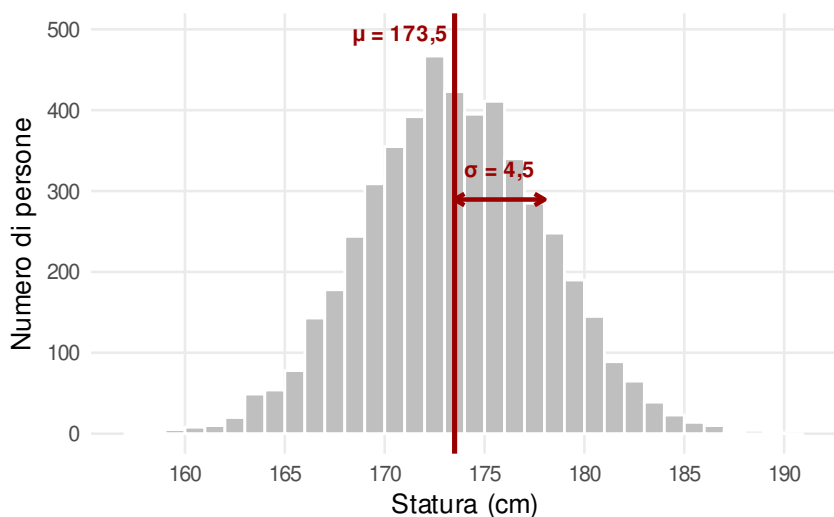


Figura 7.1: Statura simulata di 5.000 uomini adulti, con media $\mu = 173,5$ cm e deviazione standard $\sigma = 4,5$ cm (valori illustrativi). La linea indica la media; la freccia, una deviazione standard.

Questa forma è la **distribuzione gaussiana**, o normale. Ha tre proprietà importanti:

- è **simmetrica** attorno alla media;

- è descritta completamente da due parametri, la **media** (μ) e la **deviazione standard** (σ), che misura la dispersione dei valori attorno alla media;
- esistono corrispondenze fisse fra distanza dalla media, espressa in deviazioni standard, e proporzione di casi.

Circa il 68% dei casi cade entro una deviazione standard dalla media, circa il 95% entro due, quasi tutti entro tre. Oltre due deviazioni standard dalla media si trova circa il 4-5% dei casi, divisi fra le due code: circa il 2,3% per lato. Queste proporzioni valgono per qualunque gaussiana, qualunque siano media e unità di misura.

7.2 La variabilità come unità di misura

La statura si misura in centimetri, e un centimetro vale lo stesso ovunque. Le abilità mentali sono invece **variabili latenti**: non si vedono e non hanno un'unità naturale. Dieci risposte corrette in una prova non sono confrontabili con dieci risposte corrette in un'altra, perché dipendono dal numero e dalla difficoltà degli item. Non esistendo una quantità fisica da misurare, si usa come metrica la variabilità stessa.

Si esprime cioè ogni punteggio come distanza dalla media della popolazione di riferimento, usando la deviazione standard come unità. Questa trasformazione si chiama **standardizzazione** e produce i **punti z**:

$$z = \frac{x - \mu}{\sigma}$$

Un punto z di +1 indica un punteggio una deviazione standard sopra la media; -2 due deviazioni standard sotto. Lo zero è la media. La metrica non dice più «quanto» in senso assoluto, ma «quanto distante dagli altri». La distanza dalla media è diventata la metrica delle differenze individuali.

Quasi tutti i punteggi usati nei test psicologici sono punti z riscaldati, cioè moltiplicati per una costante e spostati per evitare decimali e numeri negativi (Figura 7.2):

7.3 Essere rari

- **QI e indici compositi** delle batterie di intelligenza: media 100, deviazione standard 15. Un QI di 115 corrisponde a $z = +1$.
- **Punteggi ponderati** dei singoli subtest: media 10, deviazione standard 3.
- **Punti T**, comuni nei questionari clinici: media 50, deviazione standard 10.
- **Percentili**: la percentuale di persone della popolazione di riferimento con un punteggio inferiore o uguale.

I percentili sono molto usati in clinica perché sono intuitivi e robusti alle violazioni della normalità. Hanno però un limite: non sono una scala a intervalli uguali. Al centro della distribuzione una piccola differenza di abilità sposta molti percentili; nelle code una grande differenza ne sposta pochi. Passare dal 50° all'84° percentile richiede una deviazione standard; passare dall'84° al 98° ne richiede un'altra. Per questo non ha senso fare la media dei percentili, e nella ricerca si preferiscono i punti z .

La stessa metrica servirà a leggere le correlazioni: nel capitolo sulla correlazione la previsione da una misura all'altra si scrive proprio in punti z .

7.3 Essere rari

Definire i punteggi come distanze dalla media ha una conseguenza fondamentale: la rarità statistica diventa, di per sé, un'informazione potenzialmente rilevante sul piano clinico.

La disabilità intellettiva non è definita da un livello di funzionamento basso in assoluto, ma basso rispetto a quello della popolazione, insieme a difficoltà nel comportamento adattivo. Una società organizza scuole, lavori e servizi attorno a ciò che la maggior parte delle persone riesce a fare: aspettative e requisiti sono tarati sulla media. Lo stesso vale per la dislessia: la scuola è costruita attorno a un ritmo medio di apprendimento della lettura, e chi ne è molto distante resta indietro. E per l'ADHD: livelli di attenzione o di impulsività diventano problematici soprattutto quando divergono da ciò che il contesto richiede.

7 Distribuzioni, punti z e punteggi standardizzati

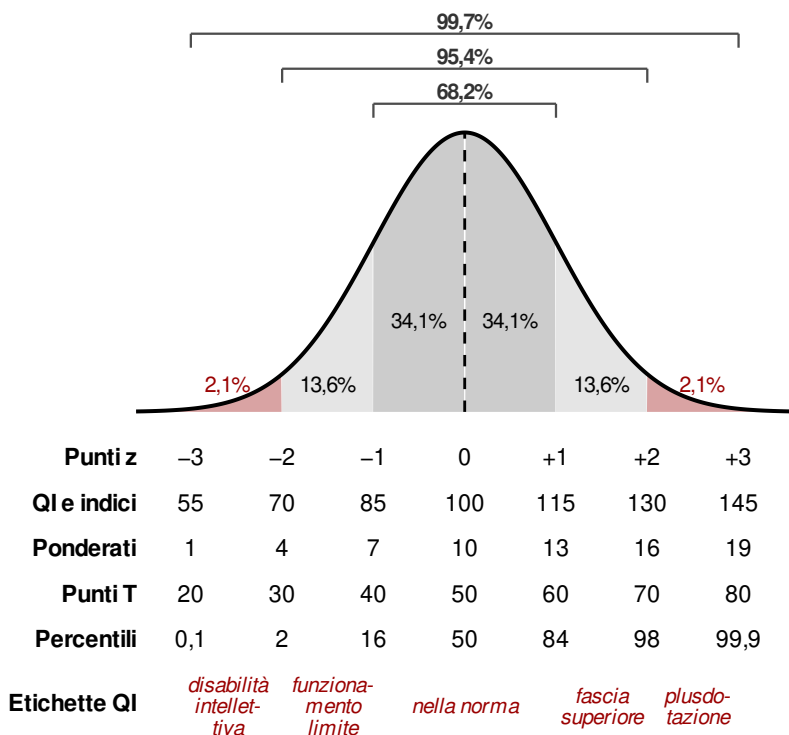


Figura 7.2: La curva gaussiana, le proporzioni di casi fra una deviazione standard e l'altra, e le principali scale di punteggio. Ogni riga riporta lo stesso punto della distribuzione in una metrica diversa; l'ultima indica le etichette descrittive usate di solito per il QI.

7.4 Da dove nasce la campana

Il ragionamento vale anche all'altro estremo. Un'abilità molto superiore alla media è rara, e la rarità può comportare un disallineamento con il contesto: un programma scolastico troppo lento, interessi poco condivisi con i coetanei. Essere statisticamente insoliti non è una diagnosi. È però una ragione per guardare meglio.

Punto metodologico: il d di Cohen

La stessa logica dei punti z serve a misurare quanto è grande la differenza fra due gruppi, per esempio clinico e non clinico, o fra due condizioni, per esempio prima e dopo un trattamento. Il **d di Cohen** è la differenza fra due medie divisa per la deviazione standard, combinando quelle dei due gruppi:

$$d = \frac{M_2 - M_1}{\sqrt{(DS_1^2 + DS_2^2)/2}}$$

La ragione è la stessa dei punti z : in psicologia il punteggio è spesso su una metrica arbitraria, e la variabilità fra persone diventa l'unità di misura. Un d di 0,5 indica che le due medie distano mezza deviazione standard (Figura 7.3). Le soglie tradizionali definiscono «piccolo» un d di 0,2, «medio» 0,5 e «grande» 0,8. Sono convenzioni proposte in assenza di dati migliori. Le revisioni degli effetti replicati in psicologia, dove i fenomeni sono molto rumorosi, suggeriscono che le soglie andrebbero abbassate (Funder & Ozer, 2019): nelle meta-analisi di psicologia sociale, per esempio, il d mediano è attorno a 0,36 (Lovakov & Agadullina, 2021). Come per ogni misura della grandezza di un effetto, il giudizio dipende dal contesto.

7.4 Da dove nasce la campana

Perché tante grandezze diverse assumono la stessa forma? Un'immagine utile è la passeggiata di una persona ubriaca che cerca di seguire una linea retta. A ogni passo barcolla un po' a destra o un po' a sinistra, a caso. Dopo un chilometro dove si troverà? Probabilmente non lontano dalla linea,

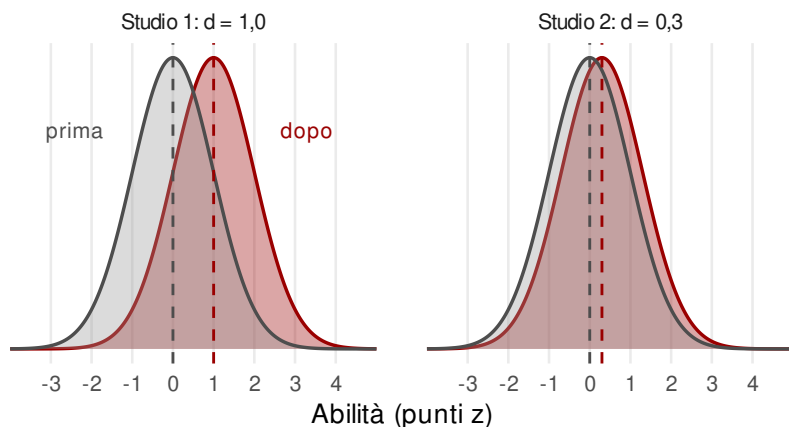


Figura 7.3: Due effetti di un trattamento espressi con il d di Cohen. Nel primo studio la media dopo il trattamento è una deviazione standard più alta ($d = 1,0$); nel secondo solo tre decimi di deviazione standard ($d = 0,3$), e le due distribuzioni quasi coincidono. Distribuzioni teoriche.

perché molte deviazioni si compensano. Talvolta, per una serie di sbandate nella stessa direzione, sarà parecchio distante. Ripetendo l'esperimento con moltissime passeggiate, le posizioni finali formano una campana (Figura 7.4). Un semplice accumularsi di effetti, che si sommano uno dopo l'altro tirando in direzioni diverse, produce alla fine una gaussiana. Lo stesso principio è reso visibile dalla tavola di Galton, in cui palline che rimbalzano a caso su file di chiodi si accumulano in una distribuzione a campana.

Il principio ha un nome: **teorema del limite centrale**. La somma di moltissime variabili diverse, ciascuna con un piccolo effetto, tende a distribuirsi in modo gaussiano, quasi indipendentemente da come sono distribuite le singole variabili (Figura 7.5). Ecco perché la gaussiana domina in psicologia: si assume che tratti, abilità e stati mentali siano il risultato finale di moltissimi processi additivi. Non dipenderebbero da una o poche cause, ma da innumerevoli fattori, piccoli o grandi, che si sommano.

Per l'intelligenza l'ipotesi additiva non è solo un'analogia. All'inizio era

7.4 Da dove nasce la campana

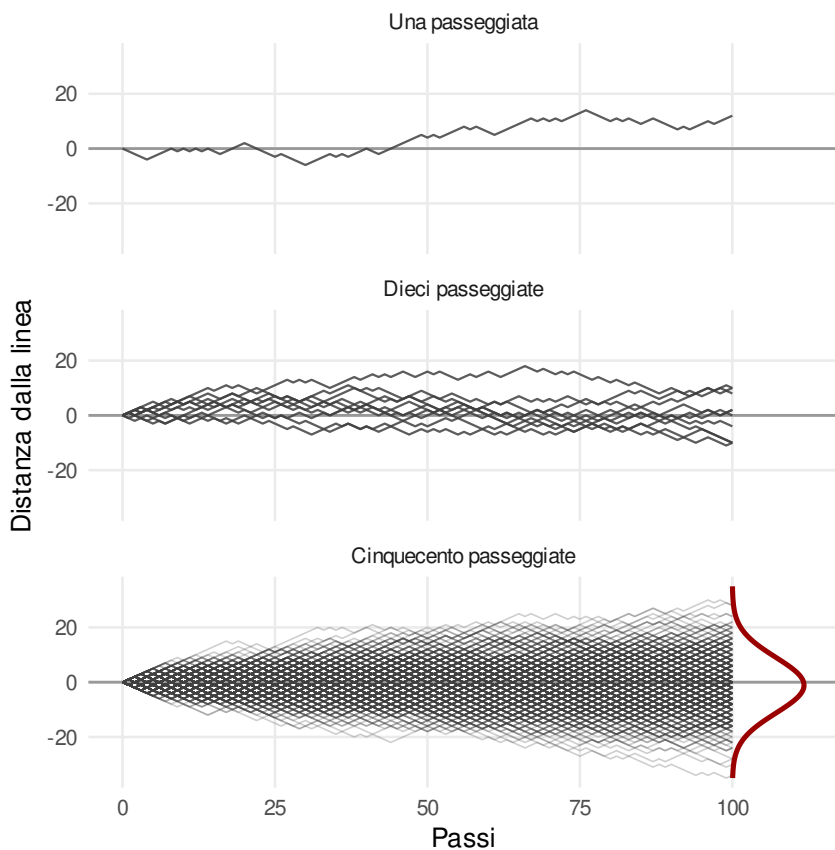


Figura 7.4: La passeggiata dell'ubriaco (random walk). A ogni passo ci si sposta a caso di un'unità in su o in giù. Una passeggiata sola ha un percorso imprevedibile; molte passeggiate si allargano a ventaglio; le posizioni finali di moltissime passeggiate (a destra, in rosso) formano una gaussiana. Dati simulati.

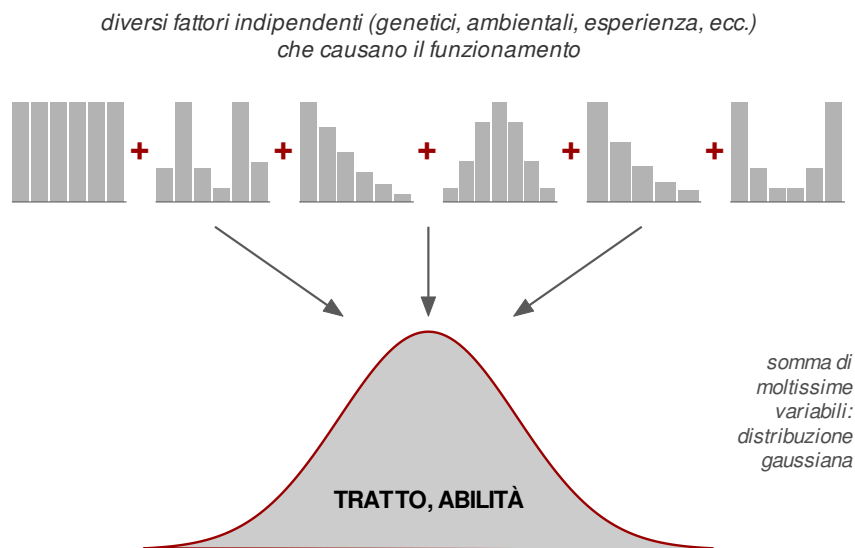


Figura 7.5: Il teorema del limite centrale. Fattori indipendenti (genetici, ambientali, di esperienza) distribuiti nei modi più diversi, piatti, asimmetrici, bimodali, discreti, sommati fra loro producono una distribuzione gaussiana: quella di un tratto o di un'abilità.

soltanto un'assunzione, ma oggi si comincia a vedere che cosa c'è sotto. Gli studi genomici stimano che le differenze genetiche associate al fattore generale dipendano da migliaia di varianti, ciascuna con un effetto minuscolo, di cui solo una piccola parte è stata identificata con certezza (G. Davies et al., 2018; Savage et al., 2018). A queste si aggiungono innumerevoli influenze ambientali, neurobiologiche, educative, motivazionali e situazionali, compresa un po' di casualità nelle risposte a un test. Le abilità e le disabilità cognitive sono quindi **poligeniche**, perché molti geni influiscono sullo stesso tratto, e per questo variano lungo dimensioni continue; e sono **pleiotropiche**, perché gli stessi geni influiscono su molte abilità, e per questo le abilità sono correlate fra loro (Kovas & Plomin, 2006). La poligenicità è stata perfino proposta come «legge» generale della genetica del comportamento: qualunque tratto comportamentale variabile dipende

7.5 Quando i dati non sono gaussiani

da moltissime varianti, ciascuna responsabile di una quota minima delle differenze (Chabris et al., 2015).

7.5 Quando i dati non sono gaussiani

La scelta di modellare abilità, tratti e stati mentali come gaussiani è nata molto prima che se ne conoscessero le basi genetiche o ambientali, perché sembrava naturale. Le evidenze successive l'hanno resa più plausibile, ma resta essenzialmente una **scelta teorica**. E c'è un colpo di scena: molti dati psicologici osservati non hanno affatto la forma di una campana, e non per ragioni profonde. Dipende dallo strumento di misura. Un test con pochi item facili produce molti punteggi massimi; uno con item difficili produce molti zeri. I tempi di risposta non possono scendere sotto lo zero e hanno una lunga coda di risposte lente. Le risposte su scala Likert sono poche categorie ordinate. Nei test cognitivi è frequente osservare distribuzioni dei punteggi grezzi asimmetriche, con un accumulo vicino al punteggio massimo.

Chi analizza i dati ha diverse strade. Può ignorare il problema e assumere comunque la normalità, sperando che non faccia danni: è la scelta più frequente, spesso per mancanza di alternative conosciute. Può usare test non parametrici, che però limitano fortemente le analisi possibili, per esempio con interazioni, effetti casuali o previsioni. Può usare modelli statistici che descrivono correttamente la distribuzione della risposta osservata, come i modelli lineari generalizzati. Oppure, se possibile, può raccogliere i dati con strumenti validati, standardizzati e normalizzati: è la prassi storica dei test di intelligenza.

Le procedure di trasformazione sono due (Figura 7.6). La **trasformazione lineare** in punti z sottrae la media e divide per la deviazione standard. È semplice e molto usata, ma conserva la forma originale: se i punteggi grezzi sono asimmetrici, lo sono anche i punti z , e le corrispondenze fra z e percentili non valgono più. Non è l'ideale, ma spesso si fa semplicemente così. La **normalizzazione** assegna invece a ogni punteggio grezzo il punto z corrispondente al suo percentile nella gaussiana. Il risultato si

7 Distribuzioni, punti z e punteggi standardizzati

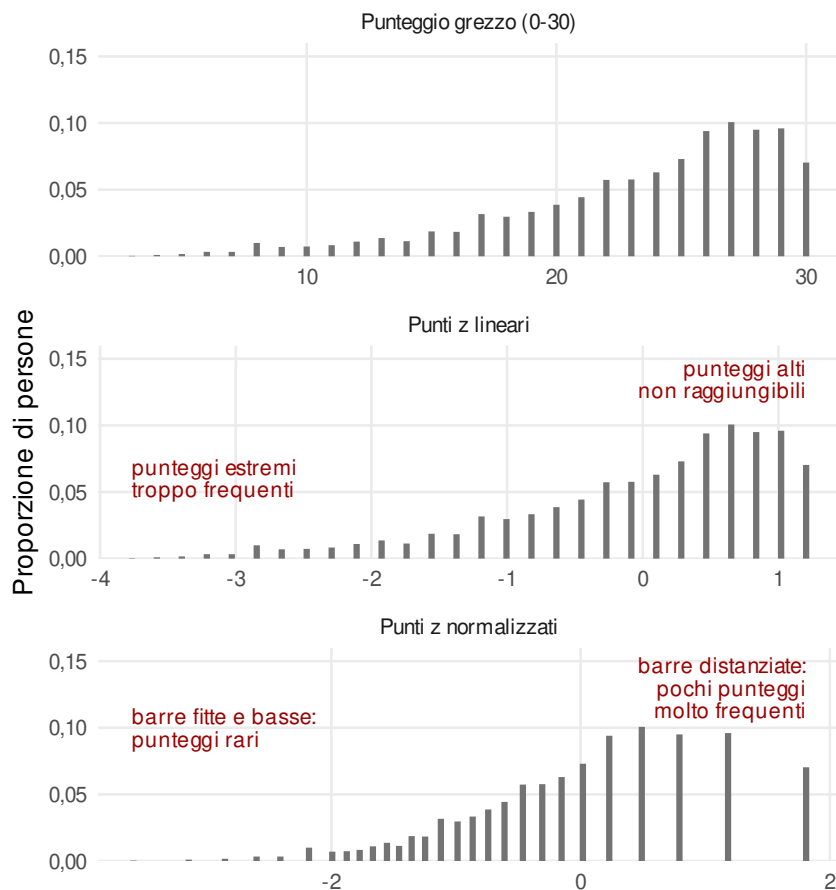


Figura 7.6: Punteggi grezzi simulati di una prova con effetto tetto (primo pannello) e due modi di trasformarli. Ogni barra è un punteggio grezzo possibile. La trasformazione lineare in punti z sposta e riscalda, ma conserva forma e spaziatura: i punteggi estremi bassi sono troppo frequenti e quelli alti non vanno oltre poco più di una deviazione standard. La normalizzazione assegna a ogni punteggio grezzo il punto z che occupa lo stesso percentile nella gaussiana: le barre non sono più equidistanti, si addensano dove i punteggi grezzi sono rari e si allontanano vicino al tetto, dove pochi punteggi molto frequenti devono coprire tutta la coda destra.

7.6 Che tipo di scala?

avvicina alla campana, ma non la riproduce: i punteggi grezzi restano pochi valori discreti, e vicino al tetto ciascuno raccoglie molte persone. Dove i punteggi sono rari, i punti z normalizzati si addensano; dove pochi punteggi molto frequenti devono coprire tutta la coda destra, si distanziano, e il massimo della prova può valere appena una o due deviazioni standard sopra la media. Una prova con effetto tetto, normalizzata, resta poco capace di distinguere fra le persone più abili. La normalizzazione è preferibile quando possibile, ma richiede competenze più avanzate ed è, di nuovo, una scelta teorica.

La normalizzazione è la procedura tipica delle tabelle normative dei test di intelligenza. Ogni tabella di conversione da punteggio grezzo a ponderato è, in sostanza, una normalizzazione calcolata separatamente per ogni fascia d'età o di scolarità: un certo numero di risposte corrette viene tradotto nel ponderato corrispondente alla stessa posizione nella distribuzione dei coetanei. In fondo a una tabella del genere c'è un'ipotesi esplicita (Figura 7.7): l'abilità latente, ciò che interessa davvero, è distribuita in modo gaussiano e causa il punteggio osservato, la cui forma irregolare riflette le caratteristiche dello strumento, come i limiti inferiore e superiore della scala, o del processo, come i tempi che non possono scendere sotto zero.

7.6 Che tipo di scala?

Le scale di misura si distinguono per le operazioni che consentono.

Tabella 7.1: Scale di misura e operazioni consentite.

Scala	Che cosa si può dire	Esempi
Categoriale	Uguale o diverso	Diagnosi sì/no; gruppo sperimentale o di controllo
Ordinale	Maggiore o minore	Risposta a un item Likert

Scala	Che cosa si può dire	Esempi
A intervalli	Distanze uguali; zero arbitrario; si può sommare e sottrarre	Temperatura in gradi Celsius; punti z ben costruiti
A rapporti	Anche «il doppio» e «la metà»; zero assoluto	Tempi; numero di errori; somma di risposte corrette; temperatura in kelvin

Il percorso tipico della misura psicologica è curioso. Si parte spesso da dati ordinali, come le risposte agli item di un questionario, o da dati a rapporti, come il numero di risposte corrette. Poi si convertono in punti z , che sono una scala a intervalli. Il motivo è sempre lo stesso: ciò che interessa è una variabile latente, il cui zero è ignoto. Ecco perché non ha senso dire che una persona con QI 140 è «il doppio più intelligente» di una con QI 70. Lo zero della scala non corrisponde all'assenza di intelligenza.

7.7 Oltre la gaussiana

La gaussiana deriva dal teorema del limite centrale: moltissime concause che si sommano. Alcune distribuzioni, però, nascono da processi diversi. Due eccezioni sono rilevanti.

La prima è la **mistura**. Il peso alla nascita è approssimativamente gaussiano, ma con un eccesso di neonati sottopeso dovuto a condizioni patologiche che agiscono come una popolazione a sé. L'intelligenza nella popolazione generale mostra un fenomeno simile: la coda inferiore contiene più casi del previsto, perché sindromi genetiche come quella di Down, mutazioni rare e danni neurologici precoci producono effetti molto grandi. In teoria anche questi effetti si sommano agli altri, ma sono così forti da formare un gruppo a parte.

La seconda è la distribuzione **log-normale**, prodotta da fattori che si moltiplicano anziché sommarsi. Si pensi a un esito sociale come il reddito: se

7.7 Oltre la gaussiana

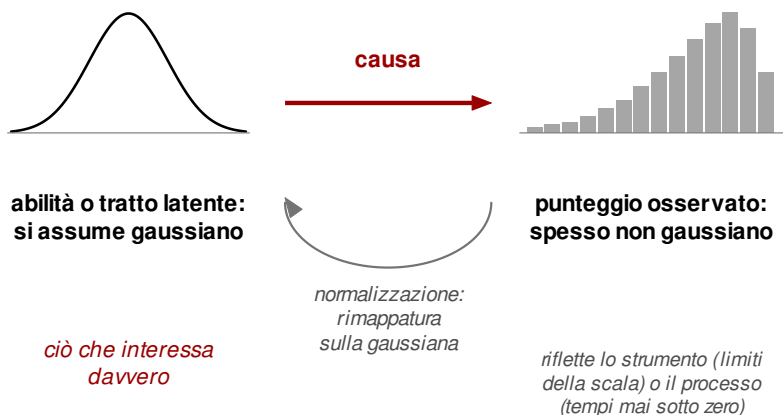


Figura 7.7: Riferirsi alla gaussiana come distribuzione della vera abilità latente è una scelta teorica. L'abilità, che si assume gaussiana, causa il punteggio osservato, che spesso non lo è; la normalizzazione rimappa il punteggio osservato sulla gaussiana.

istruzione, salute, opportunità, impegno e abilità interagiscono in modo moltiplicativo, basta che uno manchi quasi del tutto per azzerare il risultato, mentre una combinazione favorevole di tutti produce valori estremi. Da qui code molto lunghe; meccanismi in cui chi vince prende tutto, o le rendite che si accumulano, accentuano la concentrazione di ricchezza. Questo modello è più adatto agli esiti sociali che alle caratteristiche psicologiche di base.

Come sappiamo che l'intelligenza è il prodotto di effetti additivi e non moltiplicativi? In realtà non lo sappiamo con certezza. Le distribuzioni approssimativamente simmetriche dei punteggi grezzi nei test tarati sono un indizio debole, perché gli item vengono selezionati proprio per ottenere quella simmetria. Per alcune abilità, come lettura e matematica, sono stati ipotizzati meccanismi cumulativi, in cui chi parte avvantaggiato accumula vantaggi ulteriori. Qualcosa di simile, pur modesto, è stato proposto anche per l'intelligenza, attraverso il legame con la scolarizzazione. La gaus-

siana resta il modello di riferimento: utile, plausibile, ma non un fatto di natura.

In breve

- La gaussiana descrive bene grandezze prodotte dalla somma di molti piccoli effetti; le abilità cognitive, fortemente poligeniche, sono un caso plausibile.
- Non avendo un'unità naturale, la misura psicologica usa la deviazione standard come unità: QI, punteggi ponderati e punti T sono punti z riscaldati, e il d di Cohen esprime nella stessa unità la distanza fra due medie.
- I percentili sono intuitivi ma non hanno intervalli uguali; la rarità statistica è un indizio da approfondire, non una diagnosi.
- I punteggi grezzi spesso non sono gaussiani per ragioni legate allo strumento; la normalizzazione delle tabelle normative assume che lo sia l'abilità latente.
- Lo zero della scala del QI è arbitrario: rapporti come «il doppio» non hanno senso.

Riferimenti bibliografici

Parte IV

Basi biologiche, ambiente e cambiamento

8 Geni, ambiente ed ereditabilità

«Intelligenti si nasce o si diventa?» La domanda sembra legittima. In realtà è mal posta, e capire perché è il primo passo per orientarsi in uno dei temi più fraintesi della psicologia.

Abbiamo un'intelligenza umana perché abbiamo un genoma umano. Con il genoma di uno scimpanzé, di un'ape o di un delfino avremmo un'intelligenza completamente diversa. In questo senso l'intelligenza ha ovviamente **basi genetiche**. Ma di solito la domanda è un'altra: quanta parte delle differenze di QI fra le persone che ci circondano riflette differenze genetiche fra loro? Questa è una domanda sull'**ereditabilità**. Riguarda la piccola frazione del genoma che varia da persona a persona, e ha una risposta precisa solo se si precisa anche la popolazione a cui si riferisce.

8.1 Basi genetiche ed ereditabilità

Una caratteristica ha **basi genetiche** quando dipende da meccanismi biologici codificati nel genoma. Può essere identica in tutti gli individui di una specie: abbiamo cinque dita per mano, i cani scodinzolano. L'**ereditabilità**, invece, riguarda le differenze: indica quanto la variazione di una caratteristica fra gli individui di una popolazione riflette variazioni genetiche fra loro. I figli di persone alte tendono a essere più alti della media; i pulcini di galline che crescono in fretta tendono a crescere in fretta.

Una caratteristica con basi genetiche chiarissime può avere ereditabilità quasi nulla, se tutti condividono le stesse varianti. E una caratteristica fortemente ereditabile può essere molto influenzata dall'ambiente. Un esempio efficace è quello dei fenicotteri (Serpico, 2022). Il loro colore rosa ha una base genetica? Sì, ma si manifesta solo se si nutrono di crostacei ricchi di carotenoidi. Sono rosa per effetto dell'alimentazione, allora? Sì, ma solo

perché la loro genetica organizza il metabolismo e il deposito di quei pigmenti nelle piume. Un gabbiano può mangiare gli stessi crostacei e non diventerà mai rosa. Geni e ambiente non si sommano come ingredienti separati: agiscono insieme.

L'ereditabilità, indicata con h^2 , è una proporzione di varianza: al numeratore la variabilità del fenotipo che dipende da differenze genetiche fra individui, al denominatore la variabilità totale del fenotipo, attribuita a qualsiasi causa.

$$h^2 = \frac{\overbrace{\text{Var}(G)}^{\text{solo genetica}}}{\underbrace{\text{Var}(P_{\text{totale}})}_{\text{genetica} + \text{ambiente}}}$$

L'ambiente, quindi, è perfino nella formula: sta al denominatore. Per uno stesso tratto, se molta variabilità proviene dall'ambiente l'ereditabilità è bassa; se ne proviene poca, l'ereditabilità è alta. In una popolazione in cui le condizioni di vita sono molto diverse da persona a persona la varianza totale cresce e l'ereditabilità scende; dove sono simili, l'ereditabilità sale. L'ereditabilità non è una proprietà del tratto in assoluto, ma del tratto in una popolazione con una certa distribuzione di geni e di ambienti.

8.2 Come si stima: famiglie, adozioni, gemelli

Il fatto che figli di genitori intelligenti tendano a essere intelligenti non basta a stimare l'ereditabilità. Genitori e figli condividono metà della variabilità genetica, ma anche la casa, il reddito, i libri, la lingua, la scuola. La somiglianza potrebbe dipendere da entrambi.

Per separare le due fonti bisogna confrontare coppie di persone che condividono l'ambiente in misura simile ma la genetica in misura diversa, o viceversa. È la logica degli studi su famiglie, adozioni e soprattutto gemelli (Tabella 8.1). Le correlazioni fra i QI seguono con regolarità il grado di somiglianza genetica. I gemelli monozigoti cresciuti insieme correlano attorno a 0,86, i dizigoti attorno a 0,60, i fratelli non gemelli attorno a 0,47.

I gemelli monozigoti cresciuti separati restano molto simili, mentre i figli adottivi somigliano poco ai genitori adottivi (Plomin et al., 2001).

Strumenti → Che cosa significa, concretamente, una correlazione di 0,86 fra due gemelli, e perché una correlazione da sola non distingue geni e ambiente: *Correlazione e differenze individuali*.

Tabella 8.1: Correlazioni del QI fra persone con diversi gradi di parentela e di ambiente condiviso; *n* indica il numero di coppie. Elaborazione da Plomin et al. (2001).

Coppia di persone	Ambiente condiviso?	Variabilità genetica condivisa	Correlazione del QI
Gemelli monozigoti (<i>n</i> = 4.672)	sì	100%	0,86
Gemelli monozigoti separati alla nascita (<i>n</i> = 93)	no	100%	0,78
Gemelli dizigoti (<i>n</i> = 5.533)	sì	50%	0,60
Genitori e figli cresciuti in famiglia (<i>n</i> = 8.433)	sì	50%	0,42
Genitori e figli non cresciuti in famiglia (<i>n</i> = 720)	no	50%	0,24
Genitori e figli adottivi (<i>n</i> = 1.491)	sì	0%	0,19

Coppia di persone	Ambiente condiviso?	Variabilità genetica condivisa	Correlazione del QI
Fratelli con gli stessi genitori (n = 26.473)	sì	50%	0,47
Fratelli con gli stessi genitori cresciuti separati (n = 203)	no	50%	0,24
Fratelli adottivi (n = 714)	sì	0%	0,32

Che cosa significhino concretamente questi numeri lo mostra la Figura 8.1: con una correlazione di 0,86 i QI dei due gemelli monozigoti si dispongono attorno alla diagonale, con 0,42–0,60 la nuvola si allarga molto. Anche fra fratelli o fra gemelli dizigoti, cioè, i punteggi possono differire di parecchio.

Una precisazione sulle percentuali. Dire che i gemelli dizigoti condividono il 50% non significa che condividano metà del genoma: tutti gli esseri umani ne condividono quasi la totalità. Significa che condividono, in media, metà delle varianti che differiscono nella popolazione.

Il confronto più informativo è quello fra gemelli monozigoti e dizigoti cresciuti insieme. Hanno la stessa età, crescono nella stessa famiglia, ma condividono rispettivamente tutta e metà della variabilità genetica. Se i monozigoti si somigliano più dei dizigoti, la differenza va attribuita ai geni. Oggi si usano modelli di equazioni strutturali più sofisticati, ma la logica del calcolo è ben riassunta dalle formule di Falconer (Tabella 8.2), che usano solo le due correlazioni: $r_{MZ} = 0,86$ e $r_{DZ} = 0,60$.

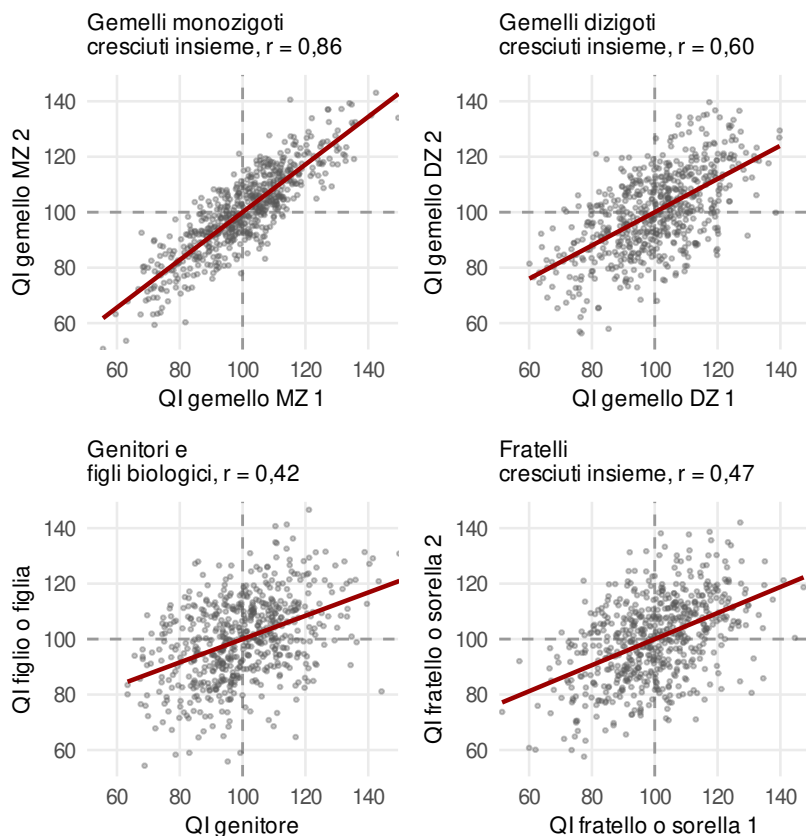


Figura 8.1: Che aspetto hanno le correlazioni della tabella: QI simulati di 600 coppie per quattro tipi di parentela, con le correlazioni riportate da Plomin et al. (2001). Ogni punto è una coppia; la linea rossa è la retta di regressione.

Tabella 8.2: Le formule di Falconer applicate alle correlazioni fra gemelli mono- e dizigoti (MZ) e dizigoti (DZ) cresciuti insieme. Le tre quote sommano a 100%.

Indice	Formula	Con i dati della tabella	Quota di varianza del QI attribuibile a...
h^2	$2 \times (r_{MZ} - r_{DZ})$	$2 \times (0,86 - 0,60) \approx 52\%$	variazioni genetiche additive (ereditabilità)
c^2	$2 \times r_{DZ} - r_{MZ}$	$1,20 - 0,86 \approx 34\%$	ambiente condiviso
e^2	$1 - r_{MZ}$	$1 - 0,86 \approx 14\%$	esperienze ambientali diverse (ambiente non condiviso) ed errore di misura

La logica è semplice. I mono- e dizigoti condividono il doppio della variabilità genetica dei dizigoti: se la loro correlazione è più alta, la differenza misura metà dell'effetto genetico, e la si raddoppia. Ciò che rende diversi due gemelli mono- e dizigoti cresciuti insieme, identici nei geni e nella famiglia, può venire solo da esperienze non condivise o da errori di misura.

L'**ambiente condiviso** comprende tutto ciò che rende simili i membri di una famiglia: reddito, quartiere, stile educativo, libri in casa. L'**ambiente non condiviso** comprende ciò che li rende diversi: amicizie, insegnanti, malattie, eventi casuali. Con questi dati, che mescolano età diverse, circa metà delle differenze di QI è attribuibile a differenze genetiche.

8.3 L'ereditabilità cresce con l'età

Il dato più sorprendente è che l'ereditabilità del QI non è fissa. L'evidenza convergente di molti studi e meta-analisi indica un'ereditabilità attorno al 20–40% nell'infanzia, al 40–60% nell'adolescenza e al 60–80% nell'età adulta (Haier, 2017; Plomin & Deary, 2015; Warne, 2020). Non è alta «da subito»: cresce progressivamente fino a coprire gran parte della varianza (Figura 8.2). In un grande studio su gemelli di diversi paesi, per esempio, è passata da circa 0,41 a 9 anni a 0,55 a 12 e 0,66 a 17 (Haworth et al., 2010). Nel frattempo l'effetto dell'ambiente condiviso si riduce fin quasi a scomparire (Plomin & Deary, 2015).

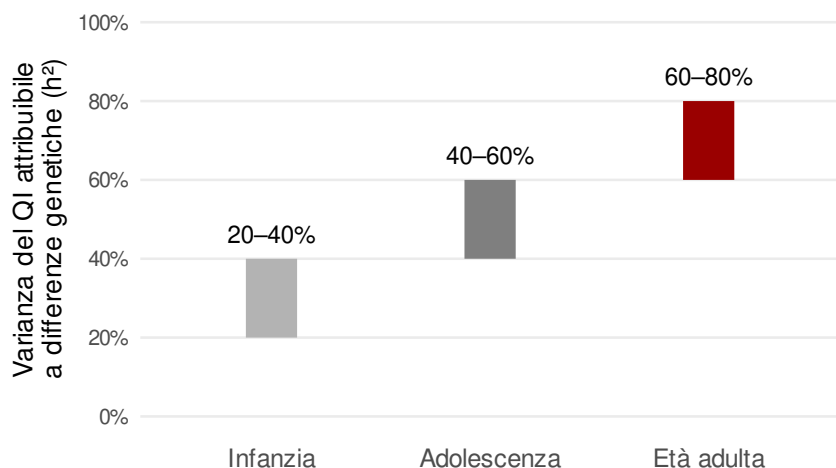


Figura 8.2: L'ereditabilità del QI cresce con l'età. Le barre indicano gli intervalli tipici delle stime nelle tre fasi della vita, secondo le sintesi di Plomin e Deary (2015), Haier (2017) e Warne (2020).

Ci si aspetterebbe il contrario: con gli anni si accumulano esperienze, e l'ambiente dovrebbe pesare di più. Perché accade? Sono state proposte diverse spiegazioni.

- **Nuovi geni che entrano in gioco?** Sembra di no. Le influenze genetiche sul QI sono in larga parte le stesse alle diverse età; ciò che cambia è il loro peso (Briley & Tucker-Drob, 2013; Plomin & Deary, 2015).
- **Correlazione gene-ambiente.** È l'ipotesi più accreditata. Piccole differenze iniziali vengono amplificate perché bambini diversi suscitano risposte diverse da genitori e insegnanti e, crescendo, scelgono attivamente ambienti in linea con le proprie inclinazioni.
- **Riduzione dell'ambiente condiviso.** Crescendo si condivide sempre meno con la famiglia d'origine.
- **Stabilizzazione del QI.** Nell'infanzia i punteggi oscillano di più, per i diversi ritmi di maturazione cerebrale; poi si stabilizzano.

Un esempio chiarisce il meccanismo della correlazione gene-ambiente. Due bambini partono quasi uguali: uno esattamente nella media, l'altro un po' sopra. Il secondo ottiene voti leggermente migliori e preferisce attività un po' più impegnative. Nell'adolescenza sceglie una scuola più esigente, riceve più attenzione da qualche insegnante, frequenta amici con interessi simili. Da adulto sceglie un'università selettiva e un lavoro intellettualmente stimolante. Il primo fa scelte tipiche e arriva dove era partito. Il secondo, partito un po' sopra la media, arriva molto sopra. L'effetto genetico iniziale è stato amplificato dall'ambiente, e l'analisi lo attribuirà ai geni. La Tabella 8.3 riassume l'esempio.

Tabella 8.3: Un esempio di correlazione gene-ambiente: l'effetto genetico iniziale viene amplificato dall'ambiente.

	Persona A: abilità generale esattamente nella media	Persona B: abilità generale un po' sopra la media
Infanzia	contesto scolastico tipico; voti medi; attività extrascolastiche tipiche	contesto scolastico tipico; voti buoni; attività extrascolastiche un po' più stimolanti

	Persona A: abilità generale esattamente nella media	Persona B: abilità generale un po' sopra la media
Adolescenza	scuola superiore e attività tipiche	a volte classi più stimolanti; scuola superiore più impegnativa; più attenzione dagli insegnanti; amici selezionati; attività complesse
Giovinezza	scelte di vita tipiche	università selettiva; contesti sociali e tempo libero sempre più stimolanti
Età adulta	scelte di vita tipiche	lavoro intellettualmente stimolante
Esito	parte nella media, arriva nella media	parte un po' sopra la media, arriva molto sopra, accumulando vantaggi

8.4 Che cosa l'ereditabilità non dice

Un'ereditabilità di 0,6–0,8 negli adulti suscita tre domande ricorrenti.

Significa che l'intelligenza è decisa dai geni e che impegno, istruzione e cambiamenti sociali sono inutili? No. L'ereditabilità descrive l'origine delle differenze fra individui in una popolazione, in un certo momento storico. Non dice nulla sul livello medio, né su che cosa accadrebbe cambiando l'ambiente di tutti. L'ambiente può avere effetti enormi sull'intelligenza dell'intera società senza che l'ereditabilità cambi, come si vedrà nel capitolo sull'effetto Flynn. E il restante 20–40% della varianza non è poco.

Il dato è definitivo? Non può esserlo per principio. L'ereditabilità esprime il peso dei geni in rapporto all'ambiente. Se le condizioni di vita diventassero più diverse da persona a persona, l'ereditabilità diminuirebbe; se diventassero più uniformi, aumenterebbe.

È meglio vivere in un mondo in cui l'ereditabilità dell'intelligenza è alta o bassa? Istintivamente si preferisce la seconda. Ma conviene pensarci. Un mondo con bassa ereditabilità è probabilmente un mondo di estremi, in cui alcuni crescono in condizioni molto avvantaggiate e altri in condizioni molto svantaggiate, e le differenze ambientali sono così grandi da schiacciare quelle genetiche. Un'ereditabilità alta è, almeno in parte, il segno che le opportunità di base sono distribuite in modo relativamente uniforme.

8.5 Quando l'ambiente è povero

Non solo un ambiente disuguale abbassa l'ereditabilità. Anche un ambiente mediamente povero sembra farlo. In uno studio su gemelli statunitensi valutati a 7 anni, l'ereditabilità del QI era attorno al 10% nelle famiglie con lo status socioeconomico più basso e molto alta in quelle con lo status più alto (Turkheimer et al., 2003). In un ambiente svantaggiato il potenziale genetico ha poco margine per esprimersi, e piccole differenze di opportunità, come una famiglia stimolante o trascurante, fanno una grande differenza. In un ambiente benestante le opportunità di base sono comunque sufficienti e le differenze genetiche emergono.

Il risultato, però, non è universale. Una meta-analisi ha trovato questa moderazione negli Stati Uniti ma non in Europa occidentale e in Australia, dove l'ereditabilità non varia in modo apprezzabile con lo status socioeconomico (Tucker-Drob & Bates, 2016). Una spiegazione plausibile è che un accesso più universale a cure mediche, istruzione di qualità e sostegno al reddito renda anche le condizioni più povere non così povere, almeno per lo sviluppo cognitivo. È un'ipotesi, non un risultato dimostrato; ma è un'ipotesi con implicazioni politiche evidenti.

8.6 I limiti degli studi sui gemelli

Gli studi sui gemelli stimano quanta variabilità di un tratto riflette variabilità genetica, non quali geni siano coinvolti. E poggiano su alcune assunzioni discutibili.

La prima è che gemelli monozigoti e dizigoti condividano l'ambiente nella stessa misura. Forse i monozigoti vengono trattati in modo più simile, perché si somigliano di più. In tal caso una parte della loro somiglianza maggiore sarebbe ambientale e l'ereditabilità risulterebbe sovrastimata.

La seconda è che i dizigoti condividano esattamente metà della variabilità genetica. Questo è vero se gli individui si accoppiano a caso. Ma le persone tendono a scegliere partner simili, e l'intelligenza è uno dei tratti per cui questo accade di più: la correlazione fra i QI dei coniugi è attorno a 0,4 (Plomin & Deary, 2015). Se i dizigoti condividono più del 50%, l'ereditabilità risulta sottostimata. Viceversa, se contano interazioni fra geni o effetti di dominanza, i dizigoti condividono meno del 50% delle combinazioni rilevanti, e l'ereditabilità risulta sovrastimata.

Questi errori vanno in direzioni opposte e la loro entità è discussa. Il quadro generale, però, è sostenuto anche da disegni diversi, come gli studi sulle adozioni e sui gemelli cresciuti separati. Il prossimo capitolo mostra che cosa aggiungono i metodi della genetica molecolare, e perché le loro stime non coincidono con quelle dei gemelli.

In breve

- Basi genetiche ed ereditabilità sono concetti diversi: la seconda riguarda le differenze fra individui in una popolazione specifica.
- Studi su famiglie, adozioni e gemelli indicano che le differenze genetiche spiegano circa metà della varianza del QI considerando tutte le età.
- L'ereditabilità cresce con l'età, fino a 0,6–0,8 negli adulti, probabilmente perché le differenze genetiche iniziali vengono amplificate dall'ambiente.
- Un'ereditabilità alta non implica immutabilità e non permette, da sola, di trarre conclusioni sulle differenze medie fra epoche o gruppi; dipende dalla distribuzione degli ambienti.
- La moderazione da parte dello status socioeconomico è stata osservata negli Stati Uniti ma non in Europa occidentale e Australia.

Riferimenti bibliografici

9 Che cosa aggiunge la genetica molecolare

Per decenni il «gene dell'intelligenza» è stato cercato come si cerca una chiave smarrita: con la convinzione che da qualche parte ci fosse. Non c'è. Non esiste un gene dell'intelligenza, così come non esiste un gene della statura o della depressione. Esistono migliaia di varianti, ciascuna con un effetto minuscolo. Arrivare a questa conclusione è stato uno dei risultati più importanti della genetica degli ultimi vent'anni.

Gli studi sui gemelli dicono quanta variabilità di un tratto è associata a variabilità genetica. Non dicono quali geni siano coinvolti né come agiscano. Con il sequenziamento del genoma umano è diventato possibile cercare direttamente nel DNA. Questo capitolo spiega che cosa si è trovato, e perché i risultati hanno sollevato problemi nuovi invece di chiudere quelli vecchi.

9.1 Dal DNA alle varianti

Il DNA è una lunga sequenza di nucleotidi, che differiscono per una base azotata: adenina, timina, guanina o citosina. Un gene è un tratto di questa sequenza, lungo spesso migliaia di nucleotidi. Il genoma umano contiene circa ventimila geni che codificano proteine, e quasi tutti esistono nella popolazione in versioni leggermente diverse, chiamate **alleli**. Poiché ogni cromosoma è presente in due copie, una ereditata da ciascun genitore, ogni persona porta due alleli per ogni gene.

La forma di variazione più semplice e più studiata è il **polimorfismo a singolo nucleotide (SNP)**: in una certa posizione del genoma, alcune persone hanno una base e altre un'altra. Il genoma umano contiene milioni di SNP comuni.

Gli **studi di associazione sull'intero genoma** (GWAS) si pongono una domanda semplice per ciascuno di questi SNP: chi porta una certa variante ha, in media, un QI più alto, più basso o uguale? Di solito si assume che gli effetti siano additivi: portare due copie della variante ha il doppio dell'effetto di portarne una.

9.2 Aghi in un pagliaio

Il problema è statistico. Gli SNP sono milioni, e i tratti come l'intelligenza sono **estremamente poligenici**: migliaia, forse decine di migliaia di varianti contribuiscono, ciascuna con un effetto piccolissimo. Per un tratto comportamentale tipico è raro che una singola variante comune spieghi più di una frazione infinitesimale della varianza. È stata proposta come una vera e propria «quarta legge» della genetica del comportamento (Chabris et al., 2015).

Per distinguere effetti così piccoli dal rumore servono campioni giganteschi. Uno dei GWAS di riferimento sull'intelligenza ha analizzato i dati di quasi 270.000 persone e ha identificato poco più di duecento regioni del genoma associate in modo statisticamente affidabile ai punteggi (Savage et al., 2018). Un altro, su circa 300.000 persone, ha ottenuto risultati simili (G. Davies et al., 2018). Le regioni identificate sono molte meno delle varianti realmente coinvolte: molte non superano le soglie di significatività, che per tenere conto dei milioni di confronti sono molto severe. E una regione può contenere molti SNP correlati fra loro, che contano come un solo segnale.

9.3 Il punteggio poligenico

Sommando gli effetti stimati in un GWAS sulle varianti presenti nel genoma di una persona si ottiene un **punteggio poligenico** (PGS): una stima della sua predisposizione genetica per un tratto. È un numero che si può calcolare alla nascita, o prima.

Quanto prevede? Poco, per ora. Un punteggio poligenico dell'intelligenza spiega in campioni indipendenti alcuni punti percentuali della varianza dei test cognitivi. Un punteggio costruito sui GWAS degli anni di istruzione, che dispongono di campioni ancora più grandi, arriva attorno al 7–10% (J. J. Lee et al., 2018; Plomin & von Stumm, 2018). È un risultato scientificamente notevole: una misura biologica, presente dal concepimento, prevede una differenza psicologica. Ma è troppo poco per qualsiasi conclusione affidabile sul singolo individuo. Due persone con lo stesso punteggio poligenico possono avere QI molto diversi. Le aziende che offrono previsioni genetiche sull'intelligenza dei figli vendono una precisione che non esiste.

9.4 L'ereditabilità mancante

Invece di cercare le varianti una per una, si può chiedere se la somiglianza genetica complessiva fra persone non imparentate, calcolata su tutti gli SNP comuni, si accompagna a una maggiore somiglianza nei punteggi. È l'idea della **analisi del tratto complesso sull'intero genoma** (GCTA). Fornisce un'ereditabilità stimata direttamente dal DNA, chiamata ereditabilità SNP (G. Davies et al., 2011).

Qui nasce un problema. Per l'intelligenza l'ereditabilità SNP si colloca attorno al 20–30% (Plomin & von Stumm, 2018). Gli studi sui gemelli, negli adulti, indicano il 60–80%. Dov'è finito il resto? È il problema dell'**ereditabilità mancante**, che riguarda quasi tutti i tratti complessi (Manolio et al., 2009).

Le spiegazioni proposte sono diverse e non si escludono.

- **Varianti rare.** Gli studi molecolari analizzano soprattutto SNP comuni, presenti in più dell'1% della popolazione. Le varianti rare sono moltissime e potrebbero spiegare una parte importante della variabilità. È un'ipotesi abbastanza accreditata.
- **Effetti non additivi.** I GWAS stimano effetti additivi. Ma possono esistere interazioni fra geni, effetti di dominanza, interazioni fra geni e ambiente. Sono effetti genetici reali, che si manifestano nei gemelli

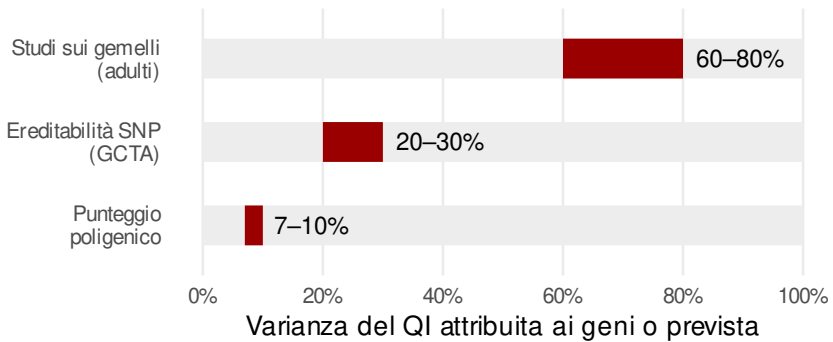


Figura 9.1: L'ereditabilità mancante. Quota di varianza del QI attribuita alle differenze genetiche dagli studi sui gemelli (adulti) e dall'analisi di tutti gli SNP comuni, e quota prevista oggi da un punteggio poligenico costruito sui GWAS degli anni di istruzione. Intervalli indicativi dalle fonti citate nel testo.

monozigoti, identici per l'intero genoma, ma non in due sconosciuti che condividono un singolo allele.

- **Sovrastima negli studi sui gemelli.** Se i gemelli monozigoti condividono più ambiente dei dizigoti, una parte della loro somiglianza è stata attribuita ai geni per errore.
- **Limiti dei dati.** Gli SNP genotipizzati catturano solo in parte le varianti causali, e la qualità imperfetta delle misure cognitive usate nei grandi campioni abbassa le stime. La potenza conta soprattutto per i punteggi poligenici: campioni più grandi stimano meglio i singoli effetti e producono previsioni migliori.

Un risultato recente complica ulteriormente il quadro. Confrontando fratelli all'interno della stessa famiglia, che hanno ereditato varianti diverse ma condividono l'ambiente familiare, le associazioni genetiche con le abilità cognitive e con l'istruzione risultano sensibilmente più deboli che nel confronto fra famiglie diverse (Howe et al., 2022). Una parte di ciò che un GWAS attribuisce ai geni di una persona riflette quindi altro: l'ambiente creato dai genitori, che dipende in parte dai loro geni; la tendenza a scegliere partner simili; differenze di ascendenza correlate con condizioni sociali. È un richiamo alla cautela, non una smentita del ruolo dei geni.

9.5 Geni e ambiente insieme

Geni e ambiente non agiscono come due addendi separati. Due esempi estremi, fuori dall'intelligenza in senso stretto, aiutano a capire.

La **fenilchetonuria** è una condizione genetica recessiva. Chi eredita due copie alterate del gene che codifica un enzima non riesce a metabolizzare la fenilalanina, un aminoacido presente in molti alimenti. La sostanza si accumula e danneggia il cervello, causando una disabilità intellettiva grave. Ma la causa diretta è l'accumulo dovuto alla dieta. Se la condizione viene individuata alla nascita, come accade con lo screening neonatale, e si segue una dieta povera di fenilalanina, la disabilità non si sviluppa (Blau et al., 2010). Una condizione interamente genetica nella causa ha una soluzione interamente ambientale nel trattamento.

Un esempio più sfumato riguarda alcune varianti del gene FTO, fra le più associate al peso corporeo. Il loro effetto si riduce sensibilmente nelle persone fisicamente attive (Kilpeläinen et al., 2011). La variante aumenta il rischio, ma il rischio si esprime in misura diversa a seconda dello stile di vita.

Per l'intelligenza, l'interazione con lo status socioeconomico vista nel capitolo precedente è un caso dello stesso tipo. Per semplicità si scrive «genetica + ambiente → tratto», ma la realtà è molto più intricata (Figura 9.2). Molti geni e molti fattori ambientali, dall'ordine di nascita all'ambiente prenatale, dalla salute allo status socioeconomico, dallo stile educativo dei genitori alla scuola, si intrecciano lungo lo sviluppo individuale. Ciascuno modula gli effetti degli altri, e tutti passano per l'organizzazione progressiva del sistema neurocognitivo. Il tratto che ne risulta, poi, non è un punto d'arrivo: un bambino più o meno capace suscita risposte diverse dai genitori, riceve rinforzi diversi, frequenta coetanei e percorsi scolastici diversi. È la correlazione gene-ambiente incontrata nel capitolo precedente, vista dall'interno.

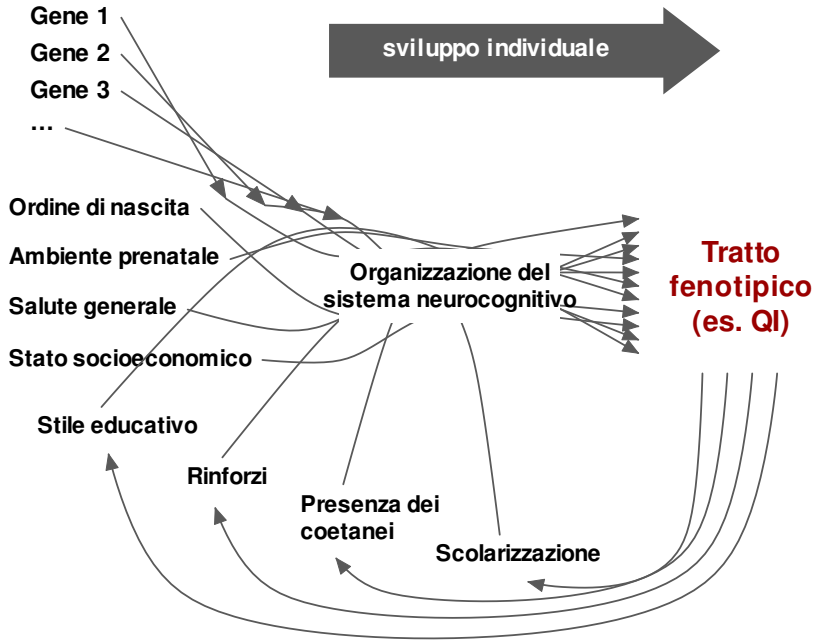


Figura 9.2: Genetica e ambiente non si sommano: si intrecciano lungo lo sviluppo individuale, passando per l'organizzazione del sistema neurocognitivo, e il tratto che ne risulta torna a modificare l'ambiente (freccie in basso). Adattamento da uno schema di Turkheimer (2000).

9.6 Pleiotropia: gli stessi geni, molti tratti

Le varianti associate all'intelligenza non sono solo numerosissime: tendono a ricorrere in tratti diversi. È la **pleiotropia**, cioè la capacità dello stesso gene di influenzare più caratteristiche.

Questo spiega in parte la positive manifold. Le basi genetiche di abilità cognitive diverse si sovrappongono in larga misura: le stesse varianti contribuiscono alla memoria, al linguaggio, al ragionamento. Si parla per questo di **geni generalisti** (Kovas & Plomin, 2006). Gli stessi geni influenzano lo sviluppo di diverse aree del cervello o caratteristiche neurobiologiche,

e ciascuna di queste influenza numerosi processi cognitivi (Figura 9.3). I processi finiscono così per correlare fra loro perché hanno cause comuni. Le cause comuni ipotizzate da Spearman hanno, almeno in parte, una base genetica.

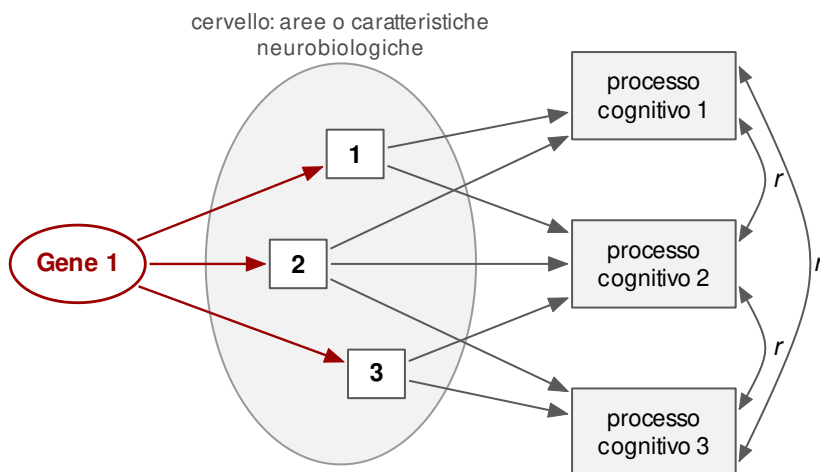


Figura 9.3: Geni generalisti e pleiotropia. Uno stesso gene influenza più aree o caratteristiche del cervello, e ciascuna di queste partecipa a più processi cognitivi: i processi risultano correlati (r) perché hanno cause comuni. Elaborazione da Kovas e Plomin (2006).

La sovrapposizione si estende oltre la cognizione. Le varianti associate all'intelligenza sono fortemente associate anche agli anni di istruzione e, in misura minore, a molte condizioni psichiatriche e neurologiche. Con alcune la correlazione genetica è negativa, come per ADHD e malattia di Alzheimer; con altre, come l'autismo, è debolmente positiva (Savage et al., 2018). Una correlazione genetica indica che gli stessi fattori genetici contribuiscono alle differenze in entrambi i tratti. Non indica che un tratto causi l'altro.

Per approfondire: la correlazione genetica nei gemelli

Anche gli studi sui gemelli permettono di stimare quanto due tratti abbiano basi genetiche comuni. Si usano le **correlazioni incrociate**: il punteggio di un gemello in lettura con il punteggio dell'altro gemello in matematica. Se questa correlazione è più alta nei monozigoti che nei dizigoti, una parte di ciò che lega lettura e matematica è genetica. La **correlazione genetica** r_G dice quanto sono sovrapposte le basi genetiche dei due tratti, x e y . Si ottiene rapportando la covarianza genetica fra i tratti alla radice del prodotto delle due ereditabilità:

$$r_G = \frac{A_{12}}{\sqrt{A_{11} A_{22}}} = \frac{2 \left(\overset{\text{incrociata, MZ}}{\widehat{r_{xy}^{MZ}}} - \overset{\text{incrociata, DZ}}{\widehat{r_{xy}^{DZ}}} \right)}{\sqrt{\underbrace{h_x^2}_{\text{tratto } x} \underbrace{h_y^2}_{\text{tratto } y}}}$$

Un esempio con numeri plausibili. In lettura $r_{MZ} = 0,80$ e $r_{DZ} = 0,47$, quindi $h_x^2 = 2 \times (0,80 - 0,47) = 0,66$. In matematica $r_{MZ} = 0,72$ e $r_{DZ} = 0,48$, quindi $h_y^2 = 0,48$. Le correlazioni incrociate valgono in media 0,51 nei monozigoti (0,53 e 0,49, a seconda di quale gemello si prende per la lettura) e 0,305 nei dizigoti (0,32 e 0,29). La covarianza genetica è $2 \times (0,51 - 0,305) = 0,41$, e $r_G = 0,41 / \sqrt{(0,66 \times 0,48)} \approx 0,73$. Se la correlazione fra lettura e matematica nella popolazione è 0,60, circa due terzi di essa (0,41 su 0,60) hanno origine genetica.

9.7 Le leggi della genetica del comportamento

Nel 2000 Eric Turkheimer riassunse in tre «leggi» ciò che gli studi su gemelli e famiglie avevano stabilito (Turkheimer, 2000):

1. ogni tratto comportamentale umano è in parte ereditabile;
2. l'effetto dell'ambiente condiviso in famiglia è di solito minore di quello dei geni;
3. una parte cospicua della variabilità resta attribuibile all'ambiente non condiviso, a eventi casuali e all'errore di misura.

La terza legge implica che non sarà mai possibile prevedere completamente il comportamento a partire da geni e ambiente familiare. La genetica molecolare ha poi aggiunto la quarta: ogni tratto dipende da moltissime varianti di effetto minuscolo (Chabris et al., 2015).

Queste leggi chiudono una vecchia polemica e ne aprono di nuove. Che i geni contino non è più in discussione. In discussione sono i meccanismi: come migliaia di varianti producano differenze nel cervello, come queste si traducano in differenze nelle prestazioni, e come l'ambiente amplifichi, attenui o indirizzi il processo. Le risposte richiedono di unire genetica, neuroscienze e psicologia dello sviluppo. Prima di arrivare al cervello, però, conviene fermarsi su un fatto che sembra contraddire tutto questo: in poche generazioni i punteggi di intelligenza di intere popolazioni sono saliti di decine di punti. È il tema del prossimo capitolo.

In breve

- Non esiste un gene dell'intelligenza: migliaia di varianti contribuiscono, ciascuna con un effetto minuscolo.
- I GWAS su centinaia di migliaia di persone hanno identificato circa duecento regioni del genoma associate all'intelligenza.
- I punteggi poligenici prevedono oggi una quota modesta della varianza, utile per la ricerca ma non per previsioni individuali.
- L'ereditabilità stimata dal DNA è molto inferiore a quella dei gemelli; varianti rare, effetti non additivi e limiti dei disegni sono le spiegazioni discusse.
- Le stesse varianti influenzano molte abilità e molti tratti: la pleiotropia fornisce una base genetica alla positive manifold.

Riferimenti bibliografici

10 L'effetto Flynn e i cambiamenti storici

Le differenze di QI fra adulti sono in gran parte ereditabili. Già Galton ne era convinto, e gli studi sui gemelli lo hanno confermato. Ne segue che la capacità intellettuale è fissata nei geni? No. L'ereditabilità descrive le differenze in una certa società, in un certo momento storico. Ma la società e il momento storico possono cambiare. E possono cambiare moltissimo.

Il caso più spettacolare è quello che oggi porta il nome di James Flynn. Interessato al dibattito sulle differenze di QI fra gruppi, Flynn si accorse di un fatto che quasi tutti avevano trascurato: ogni volta che un test veniva ritariato su un nuovo campione, le persone ottenevano punteggi grezzi medi più alti di quelle della taratura precedente (Flynn, 1984). Il ritmo corrispondeva a circa tre punti di QI per decennio. In un secolo, trenta punti: due deviazioni standard.

10.1 Una crescita in tutto il mondo

Flynn mostrò che il fenomeno non era limitato agli Stati Uniti: lo stesso accadeva in molti paesi industrializzati, talvolta a ritmi ancora più elevati (Flynn, 1987). Una meta-analisi successiva, con dati raccolti fra il 1909 e il 2013 in 31 paesi, ha stimato un aumento medio di quasi tre punti per decennio nel QI totale e di circa quattro nelle prove di intelligenza fluida, con grandi differenze fra paesi ed epoche (Pietschnig & Voracek, 2015).

Che cosa significa, in concreto? Se valutassimo la popolazione di oggi con le norme di inizio Novecento, il QI medio sarebbe attorno a 130 (Figura 10.1). Viceversa, valutata con le norme di oggi, la persona media di inizio Novecento otterrebbe un punteggio attorno a 70, al confine con la disabilità intellettuale.

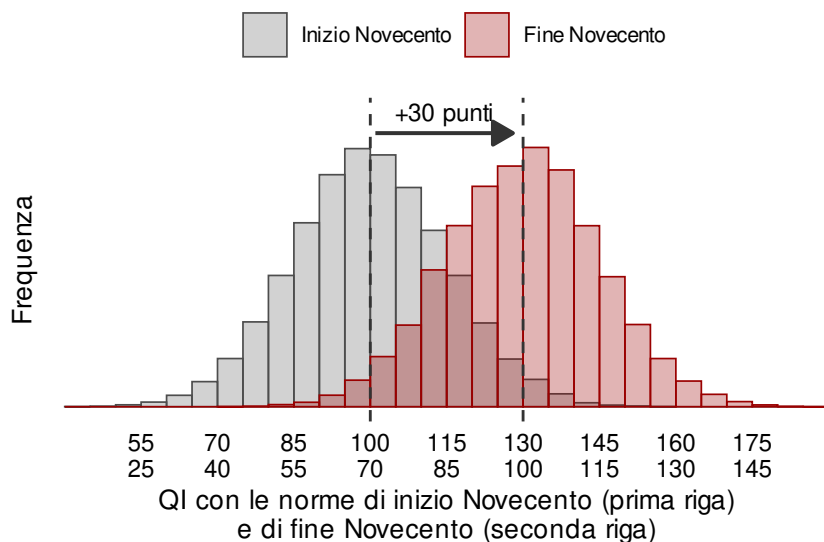


Figura 10.1: Grafico puramente illustrativo della grandezza dell'effetto Flynn: distribuzione dei punteggi di una popolazione di inizio Novecento e di una di fine Novecento. La prima riga dell'asse orizzontale esprime il QI con le norme di inizio Novecento, la seconda con quelle di fine Novecento. Il valore di 30 punti è un ordine di grandezza, non una stima per un paese specifico.

Il secondo numero fa capire perché la scoperta fu così sconvolgente. I nostri bisnonni non erano, in maggioranza, persone con una disabilità intellettiva. Lavoravano, gestivano famiglie e aziende, combattevano guerre, scrivevano libri. Eppure, per buona parte del Novecento, molti intellettuali dubitavano che le masse fossero capaci di svolgere ruoli cognitivamente complessi, e sostenevano l'istruzione di massa con scarsa convinzione (Flynn, 2020). L'effetto Flynn è anche la smentita di quel pessimismo.

10.2 Perché i punteggi sono cresciuti

La prima cosa da escludere è la genetica. La composizione genetica di una popolazione non cambia in modo apprezzabile in pochi decenni: l'effetto Flynn non è evoluzione. Le cause vanno cercate nell'ambiente, e conviene distinguerne tre famiglie.

La prima è l'**ambiente materiale**: alimentazione più ricca e varia, minori carenze nutrizionali, come quella di iodio, ridotta dall'introduzione del sale iodato; meno malattie infettive, grazie a igiene, vaccini e antibiotici; minore esposizione a neurotossine come il piombo. Questi fattori sono stati probabilmente decisivi nella prima metà del secolo, e rilevanti più a lungo per il piombo, eliminato da benzine e vernici solo negli ultimi decenni del Novecento.

La seconda è l'**ambiente sociale e culturale**: più anni di scuola per tutti, esposizione quotidiana a libri, giornali, radio e televisione, lavori sempre più cognitivamente impegnativi, famiglie più piccole, una vita quotidiana più complessa, varia e imprevedibile. Questi fattori sono rimasti attivi per tutto il secolo.

La terza sono i **bias di misura**: la crescente familiarità con il formato dei test, la pratica con prove simili. Questi aumentano i punteggi senza aumentare le capacità.

Flynn propose un'interpretazione suggestiva di come l'ambiente culturale abbia agito: cambiando le **abitudini mentali** (Flynn, 2020). Una persona di inizio Novecento, alla domanda su che cosa abbiano in comune un cane e un coniglio, poteva rispondere che il cane serve a cacciare il coniglio. Oggi si risponde che sono mammiferi. Davanti a un sillogismo con una premessa ipotetica, come «in questa regione non ci sono cammelli; questa città è in quella regione; ci sono cammelli in questa città?», la risposta poteva essere «non lo so, non ci sono mai stato». Il mondo di allora era ancorato all'utilità concreta delle cose. Il mondo di oggi abitua a classificare, astrarre, ragionare su ipotesi. Non a caso gli aumenti maggiori si osservano nelle prove come le matrici di Raven, che richiedono ragionamento ipotetico su stimoli astratti. Prove di intelligenza cristallizzata, come il vocabolario,

sono cresciute di meno, nonostante l'enorme aumento della scolarità. L'effetto sembra riflettere più un cambiamento nel modo di ragionare che un effetto diretto dell'istruzione formale.

Nessuna spiegazione singola regge da sola. L'istruzione spiega male perché i guadagni siano maggiori nell'intelligenza fluida. La tecnologia non spiega perché la crescita sia iniziata prima della sua diffusione di massa, né perché si osservi anche in contesti rurali poco tecnologici. La familiarità con i test non spiega perché i guadagni siano cominciati prima che i test si diffondessero. L'alimentazione e la salute prevederebbero guadagni maggiori nei bambini, mentre spesso sono maggiori negli adulti (Pietschnig & Voracek, 2015). La Tabella 10.1 riassume le principali ipotesi e la difficoltà maggiore di ciascuna. Il quadro più plausibile è quello di molti fattori che agiscono insieme e si rinforzano a vicenda, con pesi diversi in epoche e paesi diversi.

Tabella 10.1: Ipotesi sulle cause dell'effetto Flynn e loro limiti principali. Elaborazione dalla sintesi di Pietschnig e Voracek (2015).

Fattore ipotizzato	Difficoltà principale
Più istruzione	non spiega perché i guadagni siano maggiori nell'intelligenza fluida che in quella cristallizzata
Esposizione alla tecnologia	gli aumenti iniziano prima della sua diffusione di massa e si osservano anche in contesti poco tecnologici
Famiglie meno numerose	non spiega l'appiattimento recente né guadagni maggiori negli adulti
Familiarità e pratica con i test	gli aumenti iniziano prima della diffusione di massa dei test
Mescolanza genetica fra popolazioni	può spiegare solo una quota piccola, pochi punti in cinquant'anni

Fattore ipotizzato	Difficoltà principale
Meno piombo nel sangue	gli aumenti iniziano prima degli anni Settanta, quando il piombo è stato eliminato
Alimentazione migliore	prevede guadagni maggiori nei bambini, non è ciò che si osserva
Meno patogeni	come per l'alimentazione
Amplificazione sociale (circoli virtuosi di piccoli guadagni)	spiega l'amplificazione, non l'effetto iniziale; plausibile, ma manca una prova empirica forte
Sviluppo cognitivo più lento e più lungo	promettente, ma è più una cornice teorica che una prova empirica

Attenzione clinica: le norme invecchiano

L'effetto Flynn ha conseguenze pratiche immediate. Se una persona viene valutata con un test tarato decenni prima, viene confrontata con una popolazione che nel frattempo è cambiata. Durante il Novecento questo portava a **sovrastimare** il QI: un punteggio di 108 poteva corrispondere a 100 rispetto ai coetanei attuali, uno di 76 a 68. Nel secondo caso la differenza può decidere una diagnosi di disabilità intellettiva. Per questo è essenziale usare test con tarature recenti. Oggi che l'effetto si è attenuato o invertito in molti paesi, il problema non scompare: i profili medi delle diverse abilità possono essere cambiati in direzioni diverse.

L'effetto ha conseguenze anche per la ricerca. Negli studi trasversali sul declino cognitivo, gli anziani possono sembrare molto compromessi rispetto ai giovani in parte perché appartengono a coorti che hanno sempre avuto punteggi più bassi. L'effetto Flynn è anche un effetto di coorte.

10.3 Che cosa è aumentato davvero?

Il QI è un indicatore osservato di qualcosa che non si osserva. Se il QI medio aumenta, aumenta anche l'intelligenza? Non è scontato.

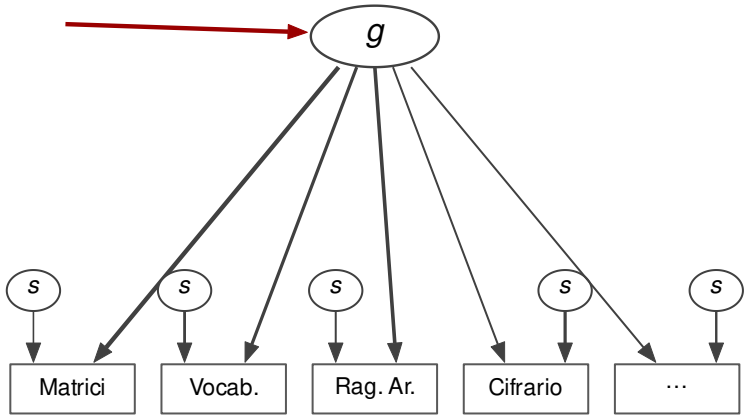
Il primo dubbio è quello dei bisnonni: un aumento di due deviazioni standard nella capacità reale è difficile da credere. Il secondo riguarda la genetica: se le differenze di intelligenza sono fortemente ereditabili e i geni non sono cambiati, può l'intelligenza essere cambiata? Sì, se si ricorda che l'espressione di una base genetica dipende dall'ambiente, come il rosa dei fenicotteri del capitolo sull'ereditabilità. No, se per intelligenza si intende solo il potenziale genetico.

Il terzo dubbio è il più serio. Gli aumenti sono molto disomogenei fra le prove: fortissimi nelle matrici e nelle prove di velocità, minori nel vocabolario, modesti nel ragionamento aritmetico. E le prove che meglio rappresentano il fattore generale non sono sempre quelle cresciute di più. Le analisi psicometriche mostrano che i guadagni fra coorti non rispettano l'invarianza di misura: le differenze fra generazioni non hanno la stessa struttura delle differenze fra individui della stessa generazione (Wicherts et al., 2004). Una lettura rigorosamente psicometrica suggerisce quindi che siano aumentate soprattutto abilità specifiche, più che g (te Nijenhuis & van der Flier, 2013).

Strumenti → *Che cos'è l'invarianza di misura e perché decide se due gruppi si possono confrontare con lo stesso test: Quando una misura è buona.*

Flynn arrivò a una posizione più sfumata. Se tutti i punteggi crescono, almeno un po', qualcosa dell'intelligenza deve essere cresciuto davvero. Non siamo diventati tutti geni, ma in molte abilità importanti la persona media raggiunge oggi livelli che a inizio Novecento sembravano impensabili. E anche se l'aumento non riguardasse g , mostrerebbe un'enorme plasticità delle capacità cognitive umane in risposta all'ambiente: un risultato forse più importante di g stesso. Una sintesi ragionevole è che i miglioramenti dell'ambiente materiale abbiano prodotto un aumento reale della capacità generale, soprattutto fino alla metà del secolo; che quelli dell'ambiente culturale abbiano prodotto un misto di capacità generale e abilità specifiche; e che i bias di misura abbiano aumentato solo i punteggi.

**Effetto
Flynn**



**Effetto
Flynn**

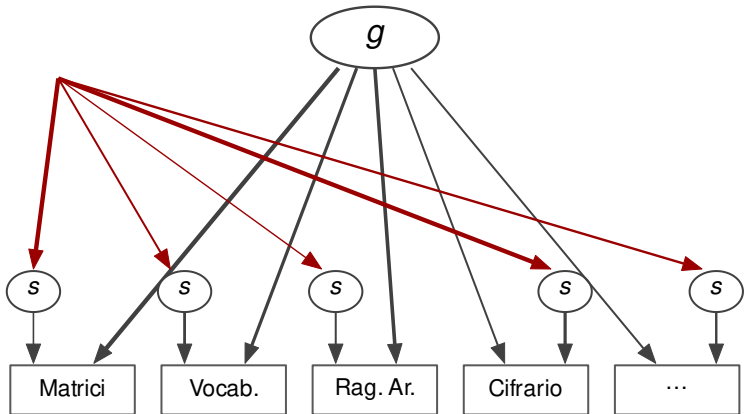


Figura 10.2: Che cosa è davvero aumentato: l'abilità mentale generale o solo abilità specifiche? Nel primo pannello l'effetto Flynn agisce su g , e i guadagni si trasmettono a tutte le prove in proporzione al loro legame con g . Nel secondo agisce sulle componenti specifiche s , con intensità diversa da prova a prova (freccie di spessore diverso). Una lettura rigidamente psicometrica fa propendere per il secondo scenario.



Figura 10.3: Le tre famiglie di cause dell'effetto Flynn e ciò che, secondo una sintesi abbastanza condivisa, ciascuna ha prodotto.

10.4 La fine, e l'inversione

L'effetto Flynn non è finito ovunque. Nella maggior parte del mondo, dove salute, nutrizione e scolarizzazione stanno ancora migliorando, i punteggi continuano a crescere (Pietschnig & Voracek, 2015). Nei paesi occidentali, però, la crescita si è fermata e in alcuni casi si è invertita.

Il primo segnale è venuto dalla Norvegia, dove tutti i ragazzi venivano valutati alla visita di leva: i dati mostravano un rallentamento a partire dagli anni Settanta e poi i primi segni di declino (Sundet et al., 2004). Oggi un **effetto Flynn inverso** è documentato in diversi paesi del Nord Europa, come Norvegia, Danimarca e Finlandia, oltre che in Francia e in altri contesti, con cali che in alcune stime arrivano a diversi punti di QI dagli anni Novanta (Dutton et al., 2016). Dati statunitensi raccolti online su quasi quattrocen- tomila adulti fra il 2006 e il 2018 mostrano cali in alcune abilità, come le serie di lettere e numeri e il ragionamento con matrici, ma un aumento nel-

la rotazione mentale tridimensionale (Dworak et al., 2023). Proprio perché riguarda item tra i più difficili, questo aumento rende poco plausibile che il calo dipenda solo da un minore impegno dei partecipanti.

Si è proposta anche una spiegazione genetica: in molte società le persone con QI più alto hanno in media meno figli, e questo produrrebbe un lento effetto «disgenico». L'associazione esiste, ma è debole e il suo effetto stimato è piccolo; inoltre agiva già durante il Novecento, quando era ampiamente compensata dall'effetto Flynn. Un'analisi decisiva è venuta ancora dalla Norvegia. Se il calo fosse dovuto a chi fa più figli, dovrebbe emergere fra famiglie diverse. Invece emerge **dentro** le famiglie: fratelli nati in anni successivi hanno in media QI più bassi dei loro fratelli maggiori, e il fenomeno non dipende dall'ordine di nascita (Bratsberg & Rogeberg, 2018). Sia l'effetto Flynn sia la sua inversione sono quindi di origine ambientale.

10.5 Perché scendiamo?

Le ipotesi sono numerose e quasi tutte da verificare. È probabile che si sia esaurita la spinta dei fattori che hanno fatto crescere i punteggi, e che ora emergano anche effetti contrari. Fra quelli materiali si citano la sedentarietà, la qualità della dieta, l'esposizione a nuovi inquinanti, la riduzione del sonno. Fra quelli culturali, scuole che chiedono meno memorizzazione e lettura approfondita, l'abitudine a delegare alla tecnologia ciò che prima si doveva ricordare, un'attenzione più frammentata. È possibile anche un problema di misura: alcuni test potrebbero essere invecchiati e non cogliere abilità oggi rilevanti.

Un indizio a favore di una spiegazione ambientale e culturale è che non tutte le abilità scendono allo stesso modo: nei dati statunitensi citati, serie di lettere e numeri e matrici calano mentre la rotazione mentale cresce (Dworak et al., 2023). Un declino biologico generale non produrrebbe andamenti così divergenti (Figura 10.4).

Nello stesso periodo si osserva un calo nei punteggi delle indagini internazionali sulle competenze scolastiche e, in alcuni studi, nelle prove di ragionamento formale di tipo piagetiano proposte agli adolescenti (Shayer & Ginsburg, 2009).

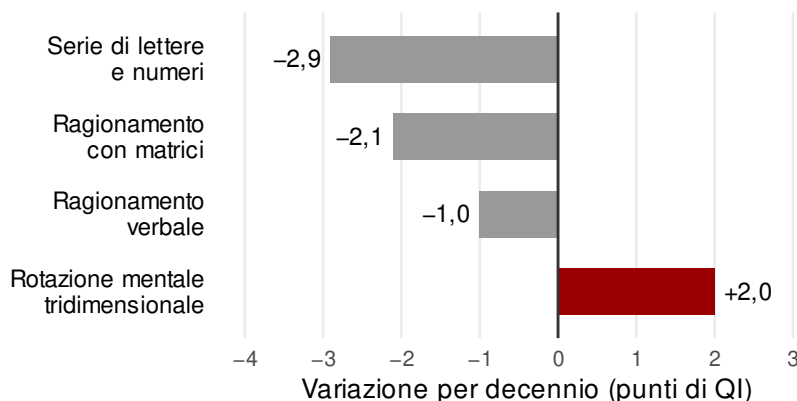


Figura 10.4: Il calo recente è diverso da un'abilità all'altra. Variazione media per decennio in quattro domini, in adulti statunitensi di 18-60 anni valutati online fra il 2006 e il 2018 (N = 394.378; rotazione mentale solo dal 2011); pendenze medie al netto di istruzione e genere, in deviazioni standard convertite in punti di QI (1 DS = 15 punti). Il calo nel ragionamento verbale è il più piccolo e il più incerto. Elaborazione da Dworak et al. (2023), tabelle 1 e 2.

La tentazione di attribuire tutto agli smartphone è forte. Va resistita, almeno nella forma monocausale. Le associazioni fra uso delle tecnologie digitali e benessere o cognizione negli adolescenti, nei grandi studi, sono in media piccole (Orben & Przybylski, 2019), e il calo dei punteggi è iniziato in alcuni paesi prima della loro diffusione. Lo stesso Flynn, negli ultimi scritti, prevedeva che il QI medio sarebbe diminuito se la società avesse ridotto le proprie richieste cognitive: meno lavori complessi, meno lettura, un'istruzione meno esigente (Flynn, 2020). È una previsione, non una spiegazione dimostrata. Ma ha il pregio di ricordare la lezione principale dell'effetto Flynn: l'intelligenza media di una popolazione dipende da ciò che quella popolazione chiede alle proprie menti.

In breve

- Per gran parte del Novecento i punteggi ai test di intelligenza sono cresciuti di circa tre punti di QI per decennio in molti

paesi, di più nelle prove di intelligenza fluida.

- Le cause sono ambientali: salute e nutrizione, scolarizzazione, ambienti più complessi e nuove abitudini di pensiero astratto, oltre a bias di misura.
- I guadagni sono disomogenei e non rispettano l'invarianza di misura: è probabile che siano aumentate soprattutto abilità specifiche, ma la plasticità dimostrata è reale.
- In diversi paesi occidentali l'effetto si è fermato o invertito; il calo avviene dentro le famiglie ed è quindi ambientale, non genetico.
- Usare test con norme recenti è indispensabile per non sovrastimare o sottostimare il funzionamento di una persona.

Per una presentazione divulgativa dello stesso tema si veda anche la pagina [L'effetto Flynn](#).

Riferimenti bibliografici

11 Cervello e intelligenza

Dove si trova l'intelligenza nel cervello? La domanda è naturale e quasi del tutto sbagliata. Quando una persona risolve un problema di ragionamento si attivano molte aree, comprese quelle visive che elaborano le figure. Ma questo dice come funziona il cervello in generale, non perché alcune persone risolvano il problema meglio di altre. Le aree visive, per esempio, si attivano in tutti, e non distinguono chi va bene da chi va male.

La domanda che interessa lo studio delle differenze individuali è un'altra: quali caratteristiche del cervello, nella struttura, nelle connessioni, nel modo di attivarsi, distinguono persone con QI diverso? Le neuroscienze hanno accumulato evidenze che il fattore generale è legato all'anatomia e al funzionamento cerebrale. Ma questo campo è anche uno dei migliori esempi di come la ricerca possa illudersi, e poi correggersi. Per questo il capitolo parte dalle cautele.

11.1 Strumenti per guardare dentro la testa

La maggior parte degli studi usa la risonanza magnetica, in tre forme principali.

- La **risonanza strutturale** fornisce un'immagine tridimensionale statica del cervello: volumi di sostanza grigia, bianca e liquido, spessore della corteccia nelle diverse aree, superfici e circonvoluzioni.
- La **risonanza funzionale** (fMRI) misura le variazioni del flusso di sangue ossigenato nelle diverse regioni. Permette di stimare quanto un'area si attiva durante un compito e quali aree si attivano insieme, a riposo o durante un compito. Da questi pattern si ricavano le reti di **connettività funzionale**.

- La **risonanza di diffusione** misura come si muovono le molecole d'acqua nei tessuti. Nei fasci di fibre nervose l'acqua si muove soprattutto lungo la direzione delle fibre. Un indice come l'**anisotropia frazionaria** misura questa coerenza, da 0 a 1, ed è usato come indicatore dell'integrità della sostanza bianca: vicino a 0 nel liquido cerebrospinale, dove l'acqua si muove in tutte le direzioni; fra 0,2 e 0,4 nella sostanza grigia o nella sostanza bianca danneggiata; fra 0,5 e 0,9 nella sostanza bianca integra, con fibre ben mielinizzate. Modelli più recenti stimano anche la densità e l'orientamento delle ramificazioni dei neuroni nella corteccia.

11.2 Perché tanti risultati non hanno retto

I dati di neuroimmagine sono molto rumorosi. Per molto tempo sono stati raccolti su campioni piccoli, quasi sempre ben sotto i duecento partecipanti, e analizzati con procedure fragili: pubblicazione selettiva dei risultati positivi, mancata correzione per i molti confronti, fuga di informazione fra i dati usati per costruire un modello e quelli usati per verificarlo. Il risultato è una letteratura piena di scoperte che non si replicano, o che si replicano con effetti molto più piccoli di quanto si credesse. Tre esempi mostrano come.

Il primo. Nel 2009 uno studio su 19 persone trovò che un indice globale di efficienza delle reti cerebrali a riposo correlava con il QI attorno a $-0,5$ (van den Heuvel et al., 2009). Sembrava rivoluzionario. Uno studio recente su quasi seimila ragazzi, replicato su oltre ottocento adulti, non ha trovato alcuna associazione apprezzabile fra il fattore generale e gli indici globali di efficienza o di organizzazione «a piccolo mondo» delle reti; le associazioni affidabili riguardano le connessioni delle singole regioni, sono distribuite su molte reti e sono di entità modesta (Molloy et al., 2025). Oggi sappiamo che associazioni affidabili fra misure cerebrali e caratteristiche psicologiche richiedono in genere migliaia di partecipanti, perché gli effetti reali sono piccoli (Marek et al., 2022).

Il secondo è una gara. Nel 2019 una competizione internazionale chiese a decine di gruppi di ricerca di prevedere l'intelligenza fluida di quasi quat-

tromila ragazzi a partire dalla risonanza strutturale, al netto del volume totale del cervello (Pohl et al., 2019). Molti usarono reti neurali profonde alimentate con le immagini grezze, nella speranza che milioni di parametri estraessero un'informazione che nessuna misura semplice coglie. Le previsioni furono verificate su un altro campione, di oltre quattromila ragazzi, tenuto nascosto fino alla fine. Il metodo vincitore, che peraltro non usava reti profonde, vi raggiunse una correlazione fra QI previsto e osservato di appena 0,03. Studi precedenti, in condizioni simili, avevano dichiarato correlazioni superiori a 0,40 con l'apprendimento automatico classico e fino a 0,70 con le reti profonde.

Il terzo esempio spiega il secondo: l'**overfitting**. I dati di neuroimmagine contengono moltissime informazioni per ogni persona: migliaia di regioni, milioni di voxel. Se si usano tutte per prevedere il QI di poche centinaia di persone, un algoritmo abbastanza flessibile trova sempre una combinazione che «indovina» il QI di ciascuna. Ma ha imparato le particolarità di quel campione, non una regola generale. La Figura 11.1 mostra il fenomeno con una simulazione.

La soluzione è verificare il modello su dati non usati per costruirlo, la **validazione incrociata**. Ma anche questa può ingannare, se i dati di stima e di verifica non sono davvero indipendenti. Nella gara del 2019 i ragazzi erano raggruppati per centro di raccolta e per famiglia, con fratelli e perfino gemelli: parte dell'informazione finiva sia nei dati di stima sia in quelli di verifica interni. I modelli sembravano funzionare finché non hanno incontrato il campione nascosto. Serve un campione distinto e del tutto indipendente.

Sotto tutti questi problemi c'è la **potenza statistica**. Fra gli studi di risonanza funzionale più citati fino al 2012, la numerosità mediana era di circa dodici partecipanti; nel 2017–2018 era salita a poco più di venti (Szucs & Ioannidis, 2020). Già nel 2013 si stimava che la potenza tipica degli studi di neuroscienze fosse fra l'8 e il 31% (Button et al., 2013). Con potenze così basse, gli effetti veri spesso non vengono rilevati, e quelli che risultano significativi sono sistematicamente sovrastimati.

La lezione generale vale per tutto il capitolo: un singolo studio non è mai definitivo. Conviene fidarsi dei risultati replicati da gruppi indipendenti

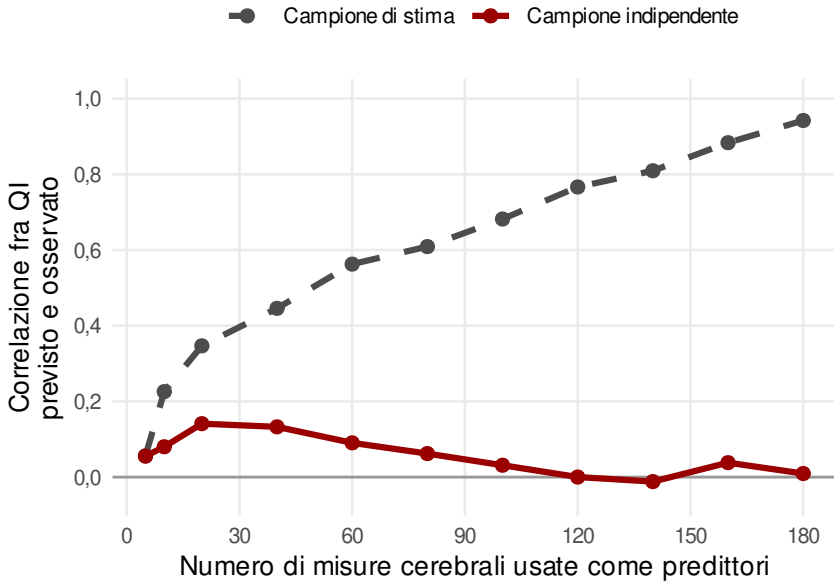


Figura 11.1: Simulazione dell’overfitting. Si prevede il QI da un numero crescente di misure cerebrali, ciascuna con un legame debolissimo con il QI, in un campione di 200 persone. La correlazione fra QI previsto e osservato cresce nel campione usato per stimare il modello, ma resta bassa e poi scende verso zero in un campione indipendente.

su campioni ampi. La buona notizia è che oggi questi campioni esistono, grazie a grandi consorzi internazionali.

11.3 Che cosa sappiamo davvero

11.3.1 Il volume cerebrale

Il risultato più solido è anche il più semplice: le persone con un cervello più grande tendono ad avere un QI più alto. In passato si citavano correlazioni fino a 0,40; le meta-analisi recenti e i grandi campioni le collocano attorno

a 0,20–0,30 (Cox et al., 2019; Pietschnig et al., 2015). L'associazione sembra legata soprattutto alla superficie totale e allo spessore della corteccia, più che a singole aree particolari.

È un effetto sicuro ma modesto. Con una correlazione di 0,25 il volume del cervello non permette di prevedere il QI di una persona: individui con lo stesso volume si distribuiscono su quasi tutta la scala (Figura 11.2). Un cervello più grande è un vantaggio statistico, non una garanzia.

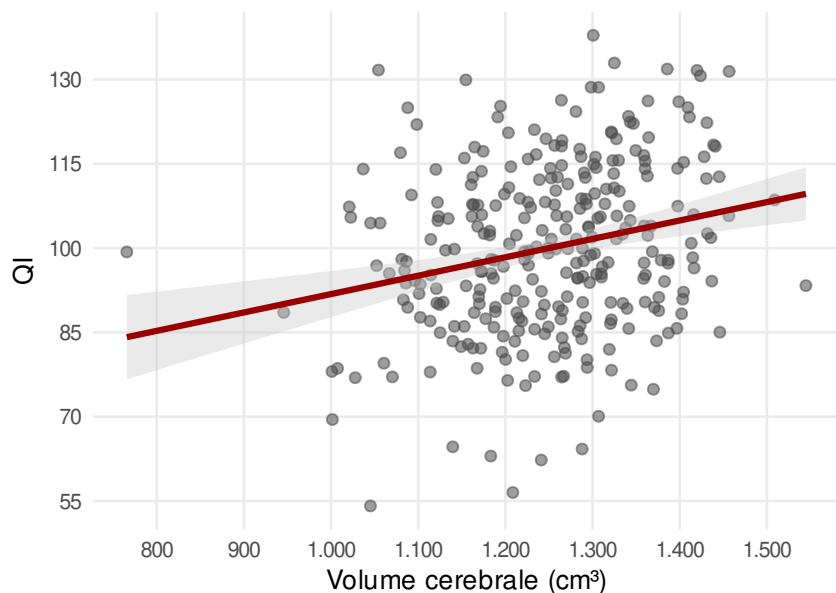


Figura 11.2: Volume cerebrale e QI: un effetto sicuro ma modesto. Dati simulati su valori realistici, 300 persone con correlazione 0,24 fra volume e QI. La retta di regressione sale, ma a parità di volume il QI varia su quasi tutta la scala: il volume non predice il QI del singolo.

C'è poi un'anomalia istruttiva. I cervelli maschili sono in media circa il 10% più grandi di quelli femminili, una differenza superiore a una deviazione standard (Ritchie et al., 2018). Con la correlazione appena vista ci si aspetterebbe un vantaggio maschile di qualche punto di QI. Le differenze medie fra i sessi nell'abilità generale, invece, sono trascurabili (Halpern

et al., 2007). Il dato non smentisce la correlazione fra volume e QI dentro ciascun sesso; suggerisce che, fra gruppi con cervelli organizzati in modo diverso, conti anche come il tessuto è distribuito e connesso, non solo quanto ce n'è.

11.3.2 Il cablaggio: la sostanza bianca

Non conta solo quanto tessuto c'è, ma come è connesso. La sostanza bianca è fatta dai fasci di assoni che collegano regioni distanti. La sua integrità, stimata con l'anisotropia frazionaria, correla con il fattore generale con valori attorno a 0,2 (Penke et al., 2012). L'associazione è più marcata in alcuni fasci: quelli che collegano lobi frontali e parietali e il corpo calloso, che collega i due emisferi. Il deterioramento della sostanza bianca nell'invecchiamento si accompagna al declino cognitivo, in particolare della velocità di elaborazione.

11.3.3 Reti efficienti

Uno studio su qualche centinaio di persone ha stimato la densità e l'organizzazione delle ramificazioni neuronali nella corteccia. Le persone con punteggi più alti avevano ramificazioni meno dense e più ordinate (Genç et al., 2018). Un'interpretazione è che lo sviluppo produca prima molte connessioni e poi ne elimini una parte, attraverso il processo di potatura sinaptica; un buon funzionamento cognitivo sarebbe associato a reti meno ridondanti e più efficienti. È un'ipotesi suggestiva, che richiede ulteriori conferme.

11.3.4 Attivazione efficiente

Alla fine degli anni Ottanta Richard Haier misurò il consumo di glucosio del cervello con la tomografia a emissione di positroni, mentre i partecipanti risolvevano le matrici di Raven. Chi otteneva punteggi più alti mostrava un metabolismo cerebrale **più basso**, non più alto (Haier et al., 1988). In un altro studio, persone che si allenavano per settimane a un videogioco

mostravano un'attività corticale ridotta man mano che diventavano esperte (Haier et al., 1992). Nacque l'ipotesi dell'**efficienza neurale**: un cervello più capace ottiene lo stesso risultato con meno dispendio (Neubauer & Fink, 2009).

L'ipotesi nella forma originale ha dato risultati incoerenti: l'efficienza emerge soprattutto in compiti di difficoltà moderata e in alcune regioni, e non sempre nella stessa direzione (Neubauer & Fink, 2009). Lo stesso Haier la considera ancora utile, a patto di definirla e misurarla meglio (Haier, 2017). È sopravvissuta, infatti, in forme più specifiche. Le persone con QI più alto sembrano disattivare di più la rete cerebrale attiva a riposo, mentre attivano di più le aree richieste dal compito (Basten et al., 2013). E le loro reti di connettività funzionale cambiano meno quando passano dal riposo a compiti diversi, come se fossero già predisposte al lavoro cognitivo; il risultato è stato replicato in campioni indipendenti (Thiele et al., 2022).

11.4 Una teoria di sintesi: P-FIT

Il quadro teorico più citato è la **teoria dell'integrazione parieto-frontale** (P-FIT), proposta da Jung e Haier sulla base di decine di studi di neuroimmagine (Jung & Haier, 2007). Secondo la teoria, le aree parietali integrano e astraggono informazioni sensoriali di diverso tipo, quelle frontali pianificano e decidono, e la qualità dell'intelligenza dipende soprattutto dalla comunicazione fra queste regioni, quindi dall'integrità dei fasci che le collegano.

Una meta-analisi successiva ha confermato in larga parte il ruolo delle regioni fronto-parietali, aggiungendo altre aree (Basten et al., 2015). La tendenza attuale è vedere l'intelligenza sempre meno legata a singole regioni e sempre più distribuita su molte reti (Deary et al., 2022). Il P-FIT funziona oggi più come un'impalcatura che come una teoria con previsioni precise. Va notato anche che riguarda soprattutto il fattore generale: mancano evidenze robuste e replicate sui correlati cerebrali delle abilità specifiche, al netto di g.

11.5 Che cosa resta

Le caratteristiche del cervello associate in modo replicabile a un fattore generale più alto sono quindi globali: un volume maggiore, una sostanza bianca più integra, specialmente nelle connessioni fronto-parietali e fra emisferi, reti più efficienti, un'attivazione più mirata durante i compiti. Tutte hanno effetti modesti. Nessuna, da sola o in combinazione, permette oggi di prevedere in modo adeguato il QI di un individuo. E molte sono probabilmente correlate fra loro, sicché i loro contributi non si sommano semplicemente.

Il bilancio è meno spettacolare di quanto molti titoli abbiano promesso. È però più solido. Il fattore generale ha correlati biologici reali, così come ha basi genetiche reali. Ma la distanza fra le varianti genetiche, le caratteristiche cerebrali e le prestazioni nei test resta in gran parte da esplorare.

In breve

- La domanda utile non è dove si attiva il cervello durante un test, ma quali sue caratteristiche distinguono persone con QI diverso.
- Molti risultati iniziali, ottenuti su campioni piccoli o con modelli sovradattati, non sono stati replicati; servono campioni di migliaia di persone.
- Il volume cerebrale correla con il QI attorno a 0,2–0,3; integrità della sostanza bianca ed efficienza delle reti mostrano associazioni simili o minori.
- La teoria P-FIT sottolinea l'integrazione fra regioni parietali e frontali, ma il quadro attuale è sempre più distribuito; nessun indicatore prevede il QI individuale.

Riferimenti bibliografici

12 Si può potenziare l'intelligenza?

Le differenze di intelligenza sono fortemente ereditabili e il QI è molto stabile dalla preadolescenza in poi. Eppure l'effetto Flynn mostra che le prestazioni cognitive di intere popolazioni possono crescere di molto. Ne nasce una domanda pratica: possiamo fare qualcosa, attivamente, per aumentare l'intelligenza di una persona?

La risposta breve è: sì, ma poco, e solo in certe circostanze. Guadagni ampi sono possibili soprattutto quando si interviene su bambini che vivono in condizioni di forte deprivazione. Per chi vive già in un ambiente favorevole, interventi come programmi prescolari aggiuntivi o giochi di allenamento cerebrale producono di solito effetti minimi o temporanei (Warne, 2020). La risposta lunga è più interessante, perché mostra come una letteratura scientifica possa produrre per anni risultati entusiasmanti che poi si dissolvono.

12.1 Effetti sicuri: la scuola

L'intervento più efficace che conosciamo è anche il più banale: andare a scuola. Più anni di istruzione si associano a QI più alti, e almeno in parte la relazione è causale. Una meta-analisi ha raccolto tre tipi di studi che permettono di avvicinarsi a una stima causale: studi longitudinali che controllano il QI misurato prima; studi che confrontano bambini nati a pochi giorni di distanza ma finiti in anni scolastici diversi a causa della data limite di iscrizione; studi su riforme che hanno allungato l'obbligo scolastico. La stima è di un aumento fra uno e cinque punti di QI per ogni anno di scuola in più (Ritchie & Tucker-Drob, 2018).

Non tutto è chiaro. Non sappiamo se l'effetto si cumuli in modo lineare per molti anni, né se riguardi l'abilità generale o soprattutto abilità specifi-

che. Uno studio longitudinale suggerisce la seconda ipotesi (Ritchie et al., 2015). Sembra che l'unica area priva di effetti apprezzabili sia la velocità di elaborazione. Ma una conclusione pratica è solida: la scuola non serve solo a trasmettere conoscenze. Aumenta le capacità con cui le si acquisisce.

12.2 Effetti sicuri: deprivazione e cura

Il secondo insieme di evidenze riguarda l'altro estremo: che cosa succede quando l'ambiente manca di cure di base. Una meta-analisi di 75 studi su quasi quattromila bambini cresciuti in istituti in 19 paesi ha trovato un QI medio attorno a 84, contro circa 104 dei coetanei cresciuti in famiglia. Il ritardo era maggiore per i bambini istituzionalizzati prima dei 12 mesi e nei paesi meno sviluppati (van IJzendoorn et al., 2008).

Uscire da queste condizioni verso un ambiente emotivamente e cognitivamente più ricco produce recuperi notevoli. Già negli anni Trenta Harold Skeels documentò aumenti di decine di punti in bambini trasferiti da un orfanotrofio a contesti più stimolanti (Skeels, 1966). L'adozione, in generale, è associata a un recupero sostanziale rispetto ai bambini rimasti nelle condizioni d'origine (van IJzendoorn et al., 2005). In uno studio svedese su fratelli separati, quelli adottati in famiglie con livello di istruzione più elevato avevano in media circa quattro punti di QI in più dei fratelli rimasti con i genitori biologici (Kendler et al., 2015).

Sono risultati importanti. Mostrano che ambienti molto diversi producono intelligenze diverse. Ma non sono facilmente traducibili in un intervento: non si può proporre l'adozione come trattamento.

12.3 Effetti probabili: interventi precoci in contesti svantaggiati

Si può allora portare il buon ambiente al bambino? Negli Stati Uniti, a partire dagli anni Sessanta, diversi progetti hanno offerto a bambini di famiglie molto povere servizi prescolari di alta qualità, spesso insieme a soste-

gno alle madri, cure mediche e pasti regolari. Alcuni avevano un disegno sperimentale con assegnazione casuale.

Il caso più spettacolare è il Milwaukee Project, avviato nel 1966 con famiglie afroamericane molto povere e madri con QI inferiore a 75: venti bambini assegnati all'intervento, venti al controllo. I primi frequentavano ogni giorno, fino ai sei anni, un servizio per l'infanzia di alta qualità, con un adulto ogni uno-tre bambini; le madri ricevevano per un anno e mezzo visite domestiche di alcune ore, più volte alla settimana, su cura dei figli, igiene, alimentazione e gestione della casa. A sei anni la differenza fra i due gruppi era di 21-32 punti di QI, a seconda del test (Jensen, 1989). Sono numeri da prendere con prudenza: i gruppi erano minuscoli, il resoconto completo arrivò solo vent'anni dopo, in un libro, e il primo direttore del progetto, Rick Heber, fu condannato nel 1981 per essersi appropriato dei fondi di un altro progetto (Jensen, 1989). Altri progetti randomizzati ottennero guadagni più contenuti ma chiari: circa 7,5 punti a cinque anni nel Carolina Abecedarian Project (Campbell et al., 2002), circa 12 nel Perry Preschool Project (Schweinhart et al., 2005).

Poi è arrivato il **fadeout** (Figura 12.1). Nel Milwaukee Project il vantaggio era sceso a 18 punti a dieci anni e a 10 punti a quattordici. I risultati scolastici raccontano la stessa storia con un dettaglio in più: in prima elementare i bambini trattati erano sopra i controlli in lettura e matematica, ma già lontani dalla media della popolazione; in quarta erano scesi entrambi i gruppi attorno al 10° percentile o poco sopra, e la distanza si era ridotta in lettura e quasi annullata in matematica (Jensen, 1989). Gli altri progetti seguirono lo stesso percorso: nell'Abecedarian restavano circa 5 punti a otto anni (Campbell et al., 2002), nel Perry Preschool il vantaggio era sparito entro la fine della scuola primaria (Schweinhart et al., 2005). Le meta-analisi degli studi randomizzati su interventi prescolari confermano lo schema: guadagni iniziali di alcuni punti, persi in gran parte entro pochi anni dalla fine dell'intervento (Protzko, 2015).

Questi studi lasciano alcuni punti fermi.

- Gli interventi funzionano meglio se precoci, in età prescolare.
- L'effetto sul QI svanisce quando l'intervento finisce. Per mantenerlo, l'ambiente migliore dovrebbe proseguire.

12.3 Effetti probabili: interventi precoci in contesti svantaggiati

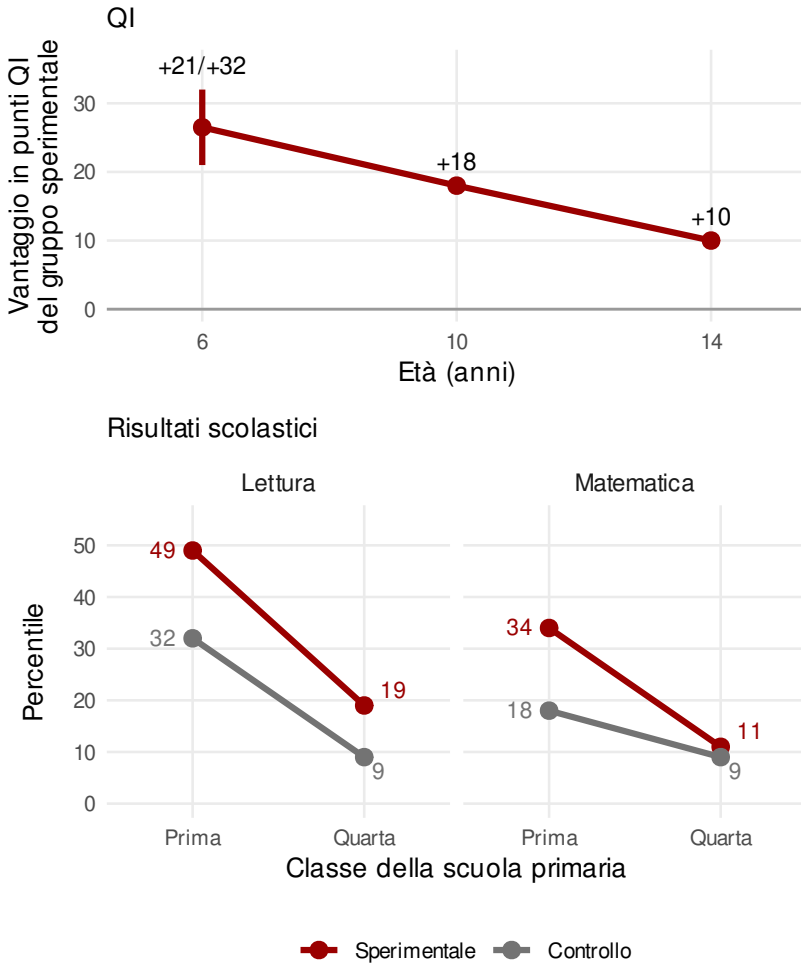


Figura 12.1: Il fadeout nel Milwaukee Project. Primo pannello: vantaggio di QI del gruppo sperimentale sul gruppo di controllo a 6, 10 e 14 anni (a 6 anni 21 punti con la WPPSI e 32 con la Stanford-Binet; il punto indica il centro dell'intervallo). Secondo pannello: percentile medio nelle prove di lettura e di matematica (Metropolitan Achievement Tests) alla fine della prima e della quarta elementare. Elaborazione dai dati riportati in Jensen (1989).

12 *Si può potenziare l'intelligenza?*

- Gli studi più recenti trovano effetti più piccoli di quelli storici. Una possibile ragione è che i contesti svantaggiati di oggi sono meno svantaggiati di allora: vaccinazioni, alimentazione e controlli medici, che i vecchi programmi includevano, sono ormai più accessibili a tutti (Duncan & Magnuson, 2013).
- Alcuni programmi hanno avuto effetti di lungo periodo su esiti diversi dal QI: più anni di istruzione, reddito più alto, meno reati. Per questo sono stati considerati economicamente vantaggiosi anche se il guadagno cognitivo era svanito (Heckman et al., 2010).

12.4 Effetti improbabili: la ricerca del Santo Graal

Da più di trent'anni ricercatori e aziende cercano un modo per aumentare l'intelligenza con attività comuni, senza cambiare drasticamente la vita di nessuno: ascoltare musica, allenarsi con giochi al computer, giocare a scacchi o ai videogiochi, prendere una pillola. Tre storie illustrano come è andata.

12.4.1 L'effetto Mozart

Nel 1993 uno studio su 36 studenti universitari riportò che dieci minuti di ascolto di una sonata di Mozart miglioravano la prestazione in alcune prove di ragionamento spaziale (Rauscher et al., 1993). L'effetto, se reale, era piccolo e durava pochi minuti. Ma la notizia fece il giro del mondo. Negli anni successivi lo stato della Georgia finanziò la distribuzione di dischi di musica classica a ogni neonato, e la Florida raccomandò musica classica nei servizi per l'infanzia.

Le repliche fallirono o diedero risultati contraddittori, spesso giustificati con dettagli sulla scelta dei brani anziché con l'ammissione che l'effetto era fragile. Una meta-analisi di circa 40 studi e oltre 3.000 partecipanti ha concluso che non esiste alcun effetto specifico di Mozart né della musica: si osserva solo un modesto miglioramento immediato dopo una stimolazione piacevole, compatibile con un aumento dell'attivazione e del buonumore.

re, insieme a forti segni di pubblicazione selettiva (Pietschnig et al., 2010; Thompson et al., 2001).

E chi fa musica per anni? I musicisti esperti ottengono in media punteggi un po' più alti in diverse prove cognitive (Talamini et al., 2017). Un recente studio condotto in decine di laboratori di molti paesi, con circa 600 musicisti e altrettanti controlli, ha trovato vantaggi in molte prove, dalla memoria al vocabolario alle matrici (Grassi & Talamini, 2025). Una rianalisi con variabili latenti suggerisce che il quadro corrisponda soprattutto a un vantaggio di quasi mezza deviazione standard nel fattore generale, più un forte vantaggio specifico nella memoria per le melodie (Toffalini & Feraco, 2026). Non è chiaro se sia la musica a causare il vantaggio o il vantaggio a favorire chi sceglie di diventare musicista. È chiaro però che non è un trattamento: è uno stile di vita. Le meta-analisi sugli interventi musicali di breve durata nei bambini non trovano benefici cognitivi o scolastici apprezzabili, una volta considerata la qualità degli studi (Sala & Gobet, 2020).

12.4.2 I training di memoria di lavoro

La seconda storia è più seria, perché aveva una teoria. Molti psicologi cognitivi hanno pensato che g dipendesse da un processo di base, come la memoria di lavoro o il controllo esecutivo. Se l'intelligenza fluida richiede di mantenere e manipolare regole in memoria di lavoro, allenando la memoria di lavoro si dovrebbe migliorare l'intelligenza fluida.

Nel 2008 uno studio molto citato riportò che alcune settimane di allenamento con un compito di memoria di lavoro, il *dual n-back*, miglioravano i punteggi in una prova di matrici (Jaeggi et al., 2008). Altri studi riportarono miglioramenti nei sintomi di ADHD dopo un allenamento simile (Klingberg et al., 2005). Seguirono un'ondata di ricerche e un'industria dell'allenamento cerebrale che ha promesso di migliorare lo studio, il lavoro, la salute mentale e di contrastare il declino cognitivo (Simons et al., 2016). Nel frattempo si sono diffusi anche i videogiochi d'azione, presentati come allenamento per attenzione e velocità.

Le meta-analisi hanno ridimensionato tutto. Gli allenamenti migliorano molto la prestazione nel compito allenato e, in misura minore, in compiti simili: è il **trasferimento vicino**. Ma il **trasferimento lontano**, verso abilità diverse, è minimo o nullo (Melby-Lervåg et al., 2016; Sala & Gobet, 2017). Lo stesso vale per scacchi, musica e videogiochi (Sala et al., 2018). Già nel 2010 un esperimento online con oltre diecimila partecipanti aveva trovato miglioramenti solo nei compiti allenati (Owen et al., 2010).

Perché la letteratura iniziale era così ottimista? Per una combinazione di problemi metodologici ricorrenti.

- **Pubblicazione selettiva.** Le riviste preferivano nettamente gli effetti «statisticamente significativi» e tendevano a respingere gli studi nulli.
- **Flessibilità analitica.** Molte misure, molti partecipanti esclusi o inclusi a discrezione, molte analisi tentate e solo alcune riportate: le possibilità di trovare qualcosa di «significativo» per caso erano altissime.
- **Nessuna preregistrazione.** Erano di fatto studi clinici randomizzati, ma quasi mai dichiaravano in anticipo ipotesi, misure e analisi.
- **Controlli passivi.** Mancava quasi sempre l'equivalente psicologico del placebo.

L'ultimo punto merita un confronto esplicito (Tabella 12.1). Il gruppo allenato non differisce dal gruppo di controllo soltanto per il training.

Tabella 12.1: Che cosa distingue il gruppo allenato da un controllo passivo e da un controllo attivo.

	Gruppo allenato	Controllo passivo	Controllo attivo
Attenzione ricevuta	molta	poca	pari a quella del gruppo allenato
Aspettative di migliorare	alte	nessuna	alte: crede di essere in trattamento

	Gruppo allenato	Controllo passivo	Controllo attivo
Familiarità con le prove	ripetuta esposizione a test e contesto	nessuna	la stessa
Impegno richiesto e abbandoni	alto; i meno motivati abbandonano	basso; pochi abbandoni	lo stesso

Con un controllo passivo, il gruppo allenato ha il massimo stimolo ad aumentare il punteggio nei test, anche se l'abilità «vera» non è cambiata, e perde per strada i partecipanti meno motivati. Negli studi clinici sui farmaci il placebo è obbligatorio. In psicologia un **controllo attivo** è difficile da costruire, ma si può fare: deve credere di essere in trattamento, essere ugualmente motivato e sottoposto agli stessi test ripetuti.

La pubblicazione selettiva, invece, lascia una traccia riconoscibile. Gli studi piccoli, che hanno stime imprecise, entrano in letteratura solo quando trovano effetti grandi, perché solo così raggiungono la significatività. Gli studi grandi, più precisi, riportano effetti piccoli, spesso vicini a zero. Il risultato è un'asimmetria: l'effetto stimato cresce al diminuire della precisione. Ma la si può sfruttare per correggere la stima. Se gli studi poco precisi riportano effetti grandi e quelli molto precisi effetti piccoli, quale effetto dovrebbe riportare un ipotetico studio infinitamente preciso? Basta prolungare la tendenza fino a un errore standard nullo (Figura 12.2).

Corretta la pubblicazione selettiva e considerati solo gli studi con controlli attivi, il trasferimento lontano di qualsiasi tipo di allenamento cognitivo diventa nullo (Gobet & Sala, 2023). La Figura 12.3 mostra il confronto per dieci gruppi di meta-analisi: memoria di lavoro in bambini, adulti e anziani, videogiochi, musica, scacchi, esercizi fisici combinati con giochi. Prima della correzione gli effetti medi sono piccoli ma tutti positivi, circa 0,12; dopo, oscillano attorno a zero.

Le promesse commerciali, intanto, hanno avuto conseguenze anche legali. Nel 2016 la Federal Trade Commission statunitense ha multato una nota

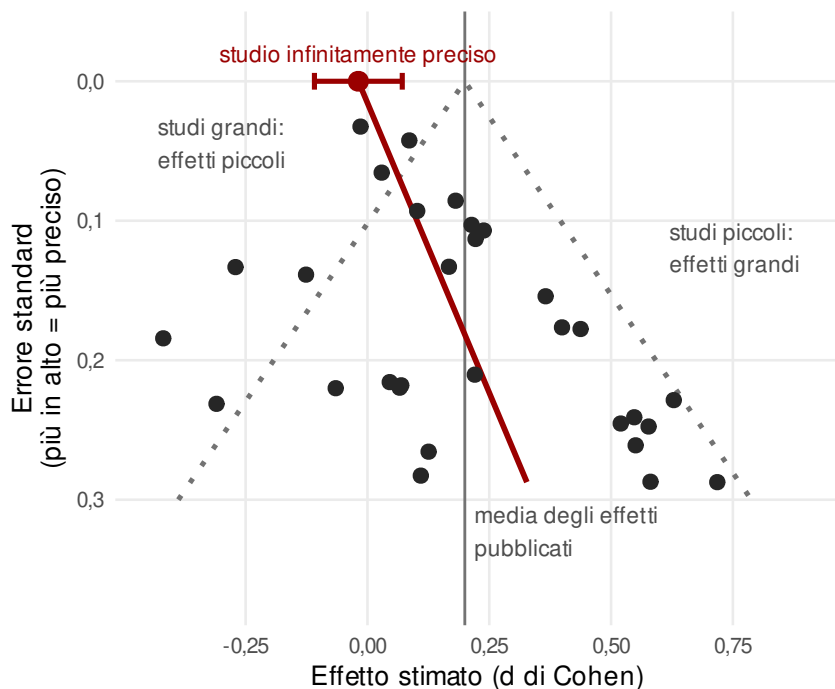


Figura 12.2: Pubblicazione selettiva e sua correzione, in una simulazione realistica. Gli studi pubblicati (30) stimano un effetto vero pari a zero; quelli con un risultato positivo significativo vengono sempre pubblicati, gli altri raramente. La media degli effetti pubblicati è 0,20 (linea verticale, al centro dell'imbuto tratteggiato). Gli studi meno precisi, in basso, riportano gli effetti più grandi. La retta rossa prolunga questa tendenza fino a un errore standard nullo: lo studio ipotetico infinitamente preciso stimerebbe un effetto vicino a zero (punto rosso, con intervallo di confidenza al 95%).

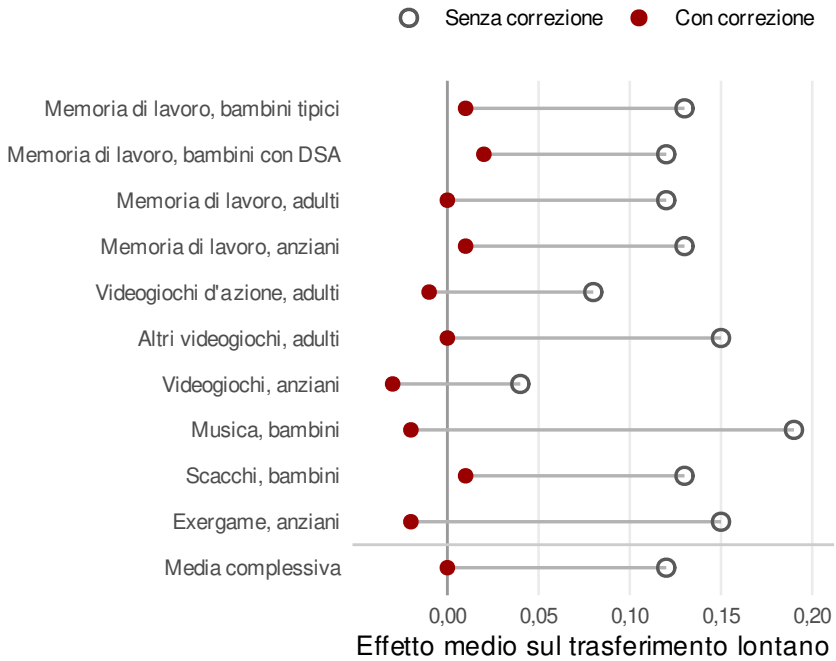


Figura 12.3: Trasferimento lontano dei training cognitivi prima e dopo la correzione. Per ogni gruppo di meta-analisi, l'effetto medio (dimensione dell'effetto standardizzata) con tutti gli studi, senza correzione (cerchi grigi), e con i soli studi a controllo attivo, corretto per la pubblicazione selettiva (punti rossi). L'ultima riga è la media complessiva. Elaborazione dalle tabelle 1 e 2 di Gobet e Sala (2023).

azienda di giochi di allenamento cerebrale per pubblicità ingannevole: prometteva di migliorare il rendimento scolastico e lavorativo e di proteggere dal declino cognitivo, senza prove adeguate (Federal Trade Commission, 2016).

12.4.3 Pillole e stimolazione

Nicotina e psicostimolanti, come quelli usati per l'ADHD, possono migliorare in acuto alcune prestazioni, soprattutto in attenzione e memoria di lavoro. Negli adulti sani gli effetti sono modesti e non modificano stabilmente l'abilità generale (Ilieva et al., 2015). Per chi ha un ADHD questi farmaci possono essere molto utili; come potenziatori dell'intelligenza nelle persone sane, no. Considerazioni simili valgono per le tecniche di stimolazione cerebrale non invasiva: alcune promesse, pochi effetti replicati su campioni ampi, di solito su funzioni specifiche e non sull'abilità generale. Oggi non esiste una pillola dell'intelligenza. Che non possa esistere in futuro è un'altra affermazione, e nessuno è in grado di dimostrarla.

12.5 Perché allenare una parte non migliora il tutto

Il fallimento dei training cognitivi è una delusione, ma era prevedibile. Russell Warne lo fa discendere dal modo in cui le abilità cognitive sono legate fra loro: i modelli dell'intelligenza assumono che sia g a sostenere la prestazione nei compiti cognitivi, non il contrario (Warne, 2020). Nel modello gerarchico le frecce vanno dall'alto in basso: dal fattore generale alle abilità ampie, dalle abilità ampie alle singole prove (Figura 12.4). A che cosa serve allora allenare un'area di funzionamento, come la memoria di lavoro, o un singolo compito? L'effetto non può risalire la corrente.

Chi propone i training cognitivi, secondo Warne, commette due errori. Non considera che g va inteso come causa delle prestazioni particolari, e quindi non può essere migliorato con esercizi su abilità particolari. E confonde l'intelligenza fluida con g : migliorare una prova di ragionamento, ammesso che accada, non equivale a migliorare il fattore generale.

L'argomento regge anche senza leggere g come un'entità che agisce dentro ogni persona. Se le prove correlano perché dipendono dalle stesse cause comuni, migliorare una prova senza toccare quelle cause lascia intatto tutto il resto. Allenarsi al *dual n-back* rende più bravi nel *dual n-back*: cambia strategie e familiarità con un compito, non la biologia o l'esperienza accumulata che sostengono anche il vocabolario e le matrici. Ed è esattamente ciò che

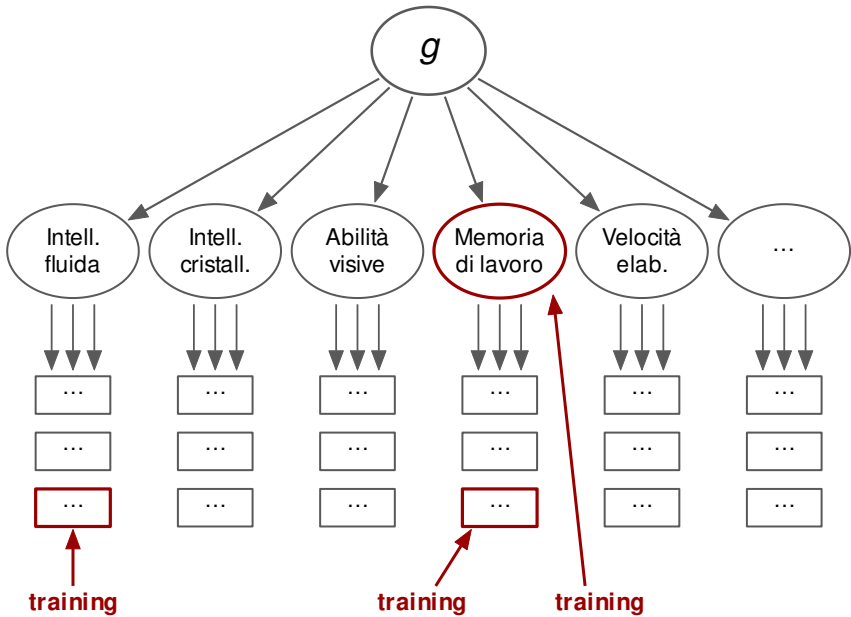


Figura 12.4: Perché l’allenamento di un’abilità o di un compito non migliora g . Nel modello gerarchico le frecce vanno dall’alto in basso: g sostiene le abilità ampie, che sostengono le singole prove (i riquadri). Un training (in rosso) agisce su un’abilità o su un compito, ma il suo effetto non risale verso g e quindi non si estende alle altre abilità. Elaborazione dall’argomento di Warne (2020).

i dati mostrano. Chi ha promesso il contrario ha confuso la prestazione in un compito con la capacità che quel compito dovrebbe misurare.

I risultati positivi, del resto, raccontano la stessa storia in forma opposta. Ciò che aumenta davvero le capacità cognitive sono cambiamenti ampi e duraturi dell’ambiente: anni di scuola, uscita da una deprivazione grave, una vita intera dedicata a un’attività complessa. Agiscono su molte cause insieme; quanto del guadagno riguardi g , e quanto abilità specifiche, resta discusso. Costano molto più di un’applicazione sul telefono, e funzionano.

12 Si può potenziare l'intelligenza?

In breve

- La scolarizzazione aumenta causalmente i punteggi cognitivi, con stime fra uno e cinque punti di QI per anno.
- Deprivazione grave e cura carente abbassano il QI; l'uscita verso ambienti migliori, come l'adozione, produce recuperi sostanziali.
- Gli interventi prescolari in contesti svantaggiati producono guadagni iniziali che svaniscono in pochi anni, pur con alcuni benefici di lungo periodo su altri esiti.
- Effetto Mozart, training di memoria di lavoro e videogiochi non producono trasferimento lontano, una volta corretti controlli passivi e pubblicazione selettiva.
- Nel modello gerarchico le frecce vanno da g alle abilità e alle prove, non il contrario: allenare un compito o un'abilità migliora quel compito e poco altro, perché non tocca ciò che sostiene tutte le prove.

Riferimenti bibliografici

Parte V

Strumenti II · Correlazione e qualità della misura

13 Correlazione e differenze individuali

Strumenti. Le due parti «Strumenti» raccolgono le basi tecniche del libro. I loro capitoli si possono leggere da soli, in qualsiasi momento, oppure saltare: i capitoli concettuali restano comprensibili e vi rimandano, con brevi note in verde, quando un passaggio richiede uno strumento in più.

Questo capitolo tratta che cosa misura una correlazione, come se ne giudica la grandezza, quando è attendibile, perché non basta a stabilire una causa e che cosa sono mediazione e moderazione. Presuppone soltanto i punti z del capitolo sulle **distribuzioni**. Le correlazioni sono già comparse a ogni passo, dalla positive manifold all'ereditabilità; qui se ne danno le regole di lettura, che serviranno subito per la scuola, la salute e il lavoro.

Chi ha un QI più alto tende a ottenere voti scolastici migliori. Eppure tutti conoscono qualcuno brillante con una pagella mediocre e qualcuno con risultati eccellenti senza particolari doti di ragionamento. Le due osservazioni non si contraddicono. Una tendenza può essere solida e ammettere moltissime eccezioni. La correlazione è lo strumento che permette di dire quanto.

Quasi tutto ciò che sappiamo sulle differenze individuali si esprime con correlazioni: fra prove cognitive, fra fratelli, fra un punteggio e un esito scolastico o lavorativo. Conviene quindi capire bene che cosa riassume questo numero e, soprattutto, che cosa lascia fuori.

13.1 Che cos'è una correlazione

La correlazione è una misura statistica della **direzione** e dell'**intensità** della relazione fra due variabili quantitative. L'idea intuitiva è semplice: se una persona sta sopra la media in una variabile, tende a stare sopra la media anche nell'altra, sotto, oppure è del tutto indifferente?

13.2 Un esempio: QI e voti scolastici

Le domande quotidiane hanno la stessa forma. Se una persona studia più della media, tende a prendere un voto più alto? Se oggi fa più caldo del solito, si vendono più gelati? Se un genitore ha un QI sopra la media, i figli tendono ad averlo sopra la media? E se oggi si vendono più gelati del solito, ci saranno più omicidi in città? L'ultima domanda sembra assurda, e ci torneremo.

Il **coefficiente di correlazione di Pearson**, indicato con r , riassume la risposta in un numero compreso fra -1 e $+1$; lo zero indica l'assenza di relazione lineare. Davanti a una correlazione conviene porsi sempre tre domande. È plausibile che le due variabili siano associate davvero, cioè che la correlazione sia diversa da zero? Se una variabile è sopra la media, l'altra tende a stare sopra o sotto? È la **direzione**, data dal segno. Quanto sono rare le eccezioni alla regola? È l'**intensità**, data dal valore assoluto: quanto i punti si dispongono vicino a una retta.

Il coefficiente non dipende dall'unità di misura. Esprimere i voti in decimi o in trentesimi, o il ragionamento in punti grezzi o in punti QI, non cambia r . Conta solo quanto stretta è la relazione fra le posizioni delle persone nelle due distribuzioni.

13.2 Un esempio: QI e voti scolastici

Nelle sintesi della letteratura la correlazione fra QI e risultati scolastici è dell'ordine di 0,5 (Roth et al., 2015). La Figura 13.1 mostra che aspetto ha una correlazione di questa grandezza. Ogni punto è una persona. La tendenza è evidente: chi sta a destra della media nel QI sta più spesso sopra la media nei voti. Ma è una tendenza statistica, mai deterministica. Ci sono persone con buoni voti nonostante un QI modesto e persone con voti scarsi nonostante un QI elevato, per i motivi più svariati. Le eccezioni non smentiscono la correlazione: sono già incluse nel valore 0,5.

Che cosa significa, concretamente, $r = 0,50$? Un errore frequente è leggerlo come una percentuale di persone: non vuol dire che metà delle persone segue la regola e metà no. Il modo più semplice di intuirlo passa per le deviazioni standard. Se una persona è una deviazione standard sopra la media nel QI, cioè a 115, ci si aspetta che sia in media mezza deviazione

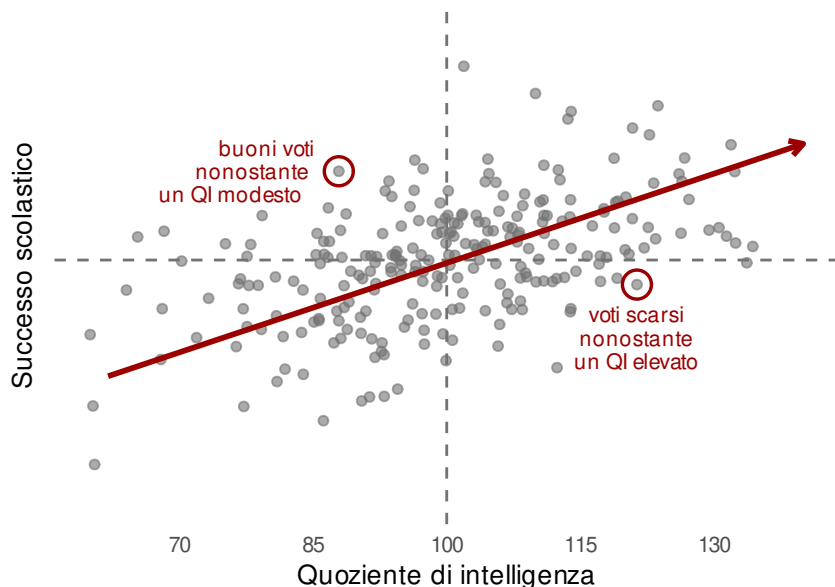


Figura 13.1: Una correlazione di 0,50 fra QI e successo scolastico. Ogni punto è una persona; le linee tratteggiate indicano le medie. La tendenza è forte ma non deterministica: sono evidenziati due casi che vanno contro la tendenza. Dati simulati.

standard sopra la media nei risultati scolastici (Figura 13.2). La correlazione dice di quanto, in media, la posizione in una variabile si trasferisce all'altra. Con i punti z del capitolo sulle distribuzioni l'idea prende una forma precisa. Se la relazione fra due misure è lineare, il valore atteso di una persona nella seconda misura, espresso in punti z , è r volte il suo punto z nella prima:

$$\widehat{z_Y} = r z_X$$

Con $r = 0,5$, chi si trova a $z = +1$ nel QI ha come previsione $z = +0,5$ nei risultati scolastici. La previsione si avvicina alla media: è il fenomeno della **regressione verso la media**, che si presenta ogni volta che due misure non sono perfettamente correlate. Non è una forza che spinge le persone verso

13.3 Quanto è grande una correlazione

la media, ma una conseguenza aritmetica della correlazione imperfetta, e vale in entrambe le direzioni: anche chi ha voti eccellenti avrà in media un QI alto, ma meno estremo dei suoi voti. Lo stesso accade fra genitori e figli: i figli di genitori con un QI molto alto tendono a stare sopra la media, ma meno dei genitori.

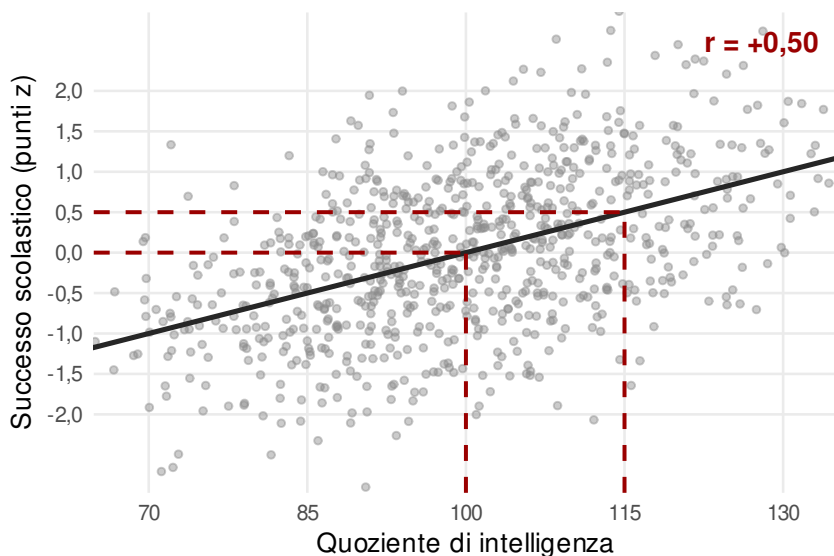


Figura 13.2: Che cosa significa $r = +0,50$: chi è una deviazione standard sopra la media nel QI (115) è in media mezza deviazione standard sopra la media nel successo scolastico ($z = +0,50$). La retta indica il valore atteso. Dati simulati.

13.3 Quanto è grande una correlazione

Esistono etichette di comodo per la grandezza di r (Figura 13.3). Una correlazione nulla descrive due cose che non c'entrano niente l'una con l'altra. Attorno a 0,20 si trova il caso più tipico di effetto vero in psicologia: una rassegna di oltre settecento correlazioni tratte da meta-analisi sulle differenze individuali ha trovato proprio un valore mediano di questo ordine

(Gignac & Szodorai, 2016). Una correlazione di 0,50, come quella fra QI e voti, è forte: variabili evidentemente molto legate. Una correlazione di 0,80, in psicologia, indica spesso la stessa cosa misurata in due modi diversi, anche se non basta da sola a stabilire che due prove misurino lo stesso costrutto.

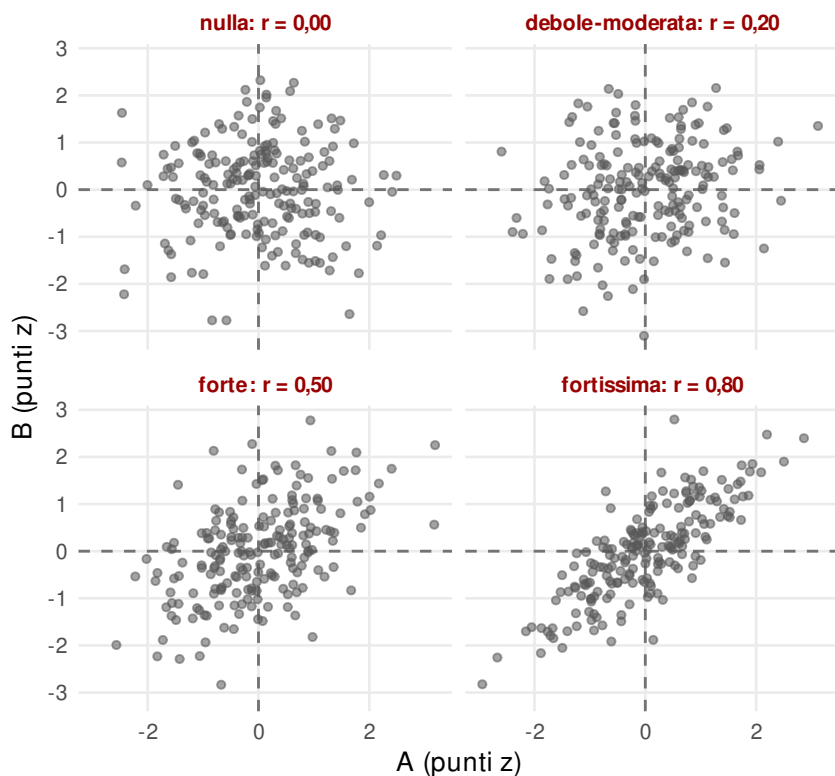


Figura 13.3: Correlazioni di grandezza crescente, con le etichette di comodo usate nel testo. Le variabili sono in punti z; le linee tratteggiate indicano le medie. Dati simulati con correlazione esatta.

Le etichette vanno usate con prudenza, perché la grandezza di un effetto dipende dal contesto. Anche un r piccolo può contare: un fattore che

13.4 Casi estremi e relazioni non lineari

sposta di poco la probabilità di un esito, applicato a milioni di persone o cumulato per anni, può avere conseguenze consistenti (Funder & Ozer, 2019). Il giudizio sulla grandezza va quindi legato alla domanda: spiegare un meccanismo, prevedere un singolo caso o stimare un effetto su una popolazione sono obiettivi diversi.

Punto metodologico: la «varianza spiegata»

Nei modelli di regressione lineare e nell'ANOVA si parla spesso di percentuale di varianza spiegata, indicata con R^2 o con indici equivalenti come η^2 , come dimensione dell'effetto. R^2 è di fatto una correlazione al quadrato: con un solo predittore coincide con il quadrato di r ; con più predittori è il quadrato della correlazione fra il criterio e la loro migliore combinazione lineare. Con $r = 0,50$ si ottiene $R^2 = 0,25$: si dice allora che il QI «spiega» il 25% della varianza dei voti. «Spiegare» ha qui un significato statistico: indica quanto la retta riduce l'errore di previsione, non quale parte delle cause dei voti sia stata individuata.

13.4 Casi estremi e relazioni non lineari

Il coefficiente di Pearson ha due punti deboli che conviene conoscere.

Il primo sono i **casi estremi**. Pearson pesa le distanze dalla media, e un singolo punto molto lontano dagli altri può spostare r in modo sensibile. Nella Figura 13.4 una nuvola senza alcuna relazione diventa una correlazione di 0,22 per colpa di un solo caso. Un rimedio è il coefficiente **rho di Spearman**, che lavora sui ranghi: ogni valore è sostituito dalla sua posizione nel campione. Il caso estremo diventa così semplicemente il primo della classifica, non un punto fuori dalla distribuzione, e la correlazione torna vicina a zero. Non è però una scorciatoia automatica: un valore estremo può essere un errore di inserimento da correggere oppure il dato più interessante del campione. Prima di scegliere il coefficiente, bisogna guardare i dati (Schober et al., 2018).

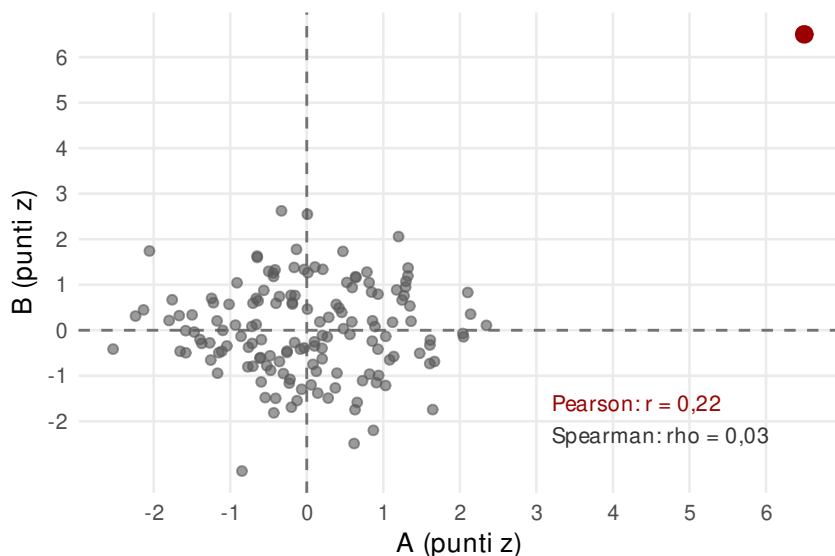


Figura 13.4: Un singolo caso estremo, in alto a destra, fa salire la correlazione di Pearson da zero a circa 0,22; la rho di Spearman, calcolata sui ranghi, resta vicina a zero. Dati simulati.

Il secondo punto debole è la **forma** della relazione. Pearson misura l'associazione **lineare**. Relazioni non lineari sono comuni quando si studia lo sviluppo: molte abilità crescono rapidamente nell'infanzia, rallentano nell'adolescenza, restano stabili e poi declinano. La Figura 13.5 mostra punteggi simulati in una prova di ragionamento con un massimo di 60 punti. Fra 6 e 18 anni una retta descrive abbastanza bene la crescita, anche se una curva flessibile fa meglio. Sull'intero arco di vita, invece, la correlazione lineare è praticamente nulla, benché il punteggio dipenda chiaramente dall'età. Un r nullo dice «nessuna tendenza lineare», non «nessuna relazione». Curve flessibili, come le interpolazioni locali, descrivono meglio queste traiettorie, al prezzo di essere meno facili da riassumere in un numero. Una parabola, cioè una relazione quadratica, è un compromesso spesso deludente: prevede discretamente in media e sbaglia di molto proprio dove la curva cambia forma.

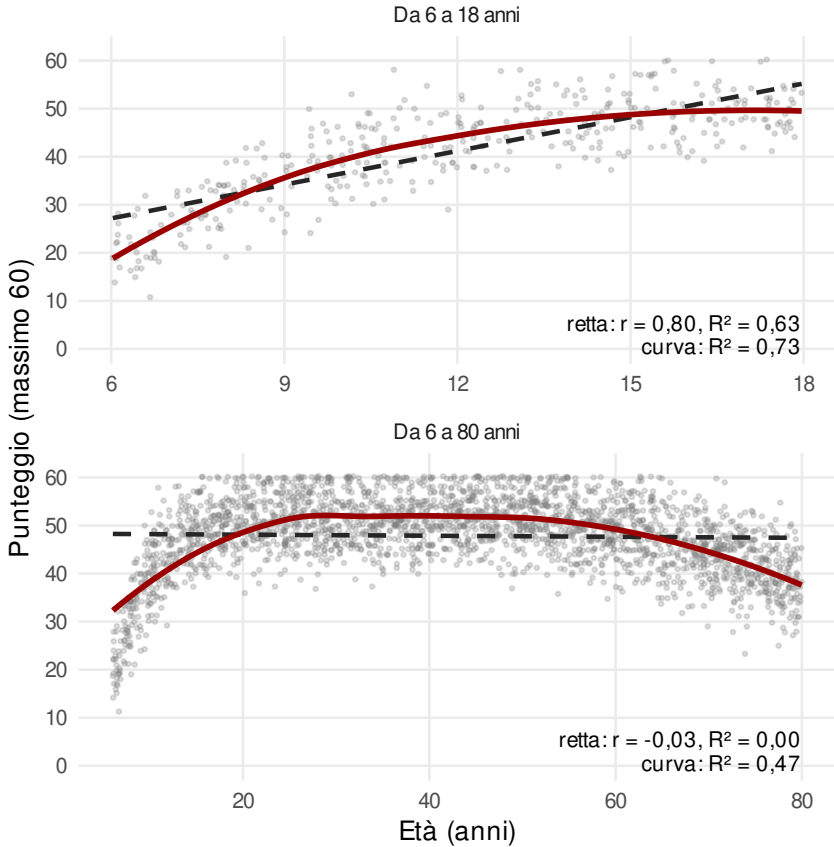


Figura 13.5: Punteggi simulati in una prova di ragionamento in funzione dell'età. Nel primo pannello, fra 6 e 18 anni, la retta (tratteggiata) descrive abbastanza bene la crescita; nel secondo, sull'intero arco di vita, la correlazione lineare è quasi nulla, mentre una curva flessibile (interpolazione locale, in rosso) coglie la relazione.

13.5 Correlano davvero?

Se le cose correlano davvero, non lo fanno «e basta»: c'è un motivo. Ed è sulle correlazioni che si regge una parte enorme della ricerca non sperimentale. Prima di cercare il motivo, però, bisogna accertare che la correlazione osservata nel campione sia attendibile, cioè presumibilmente diversa da zero nella popolazione da cui il campione proviene.

Una correlazione calcolata su un campione è una stima. Se lo stesso studio fosse ripetuto su altre persone, il valore cambierebbe, e quanto cambierebbe dipende soprattutto dalla numerosità. Nella Figura 13.6 la popolazione, in grigio, ha una correlazione vera di 0,20 fra QI e apertura mentale. Un campione di 30 persone, estratto a caso, dà 0,28; uno di 200 dà 0,20. Il campione piccolo ha trovato la correlazione *più forte*, eppure è quello che non permette di concludere nulla: con 30 persone si possono ottenere valori vicini a zero, negativi, oppure doppi del valore vero, e un 0,28 è del tutto compatibile con una popolazione in cui la correlazione è zero. Con 200 persone la stima è molto più stabile, e anche un valore più piccolo si distingue dallo zero.

Un metodo fondamentale, anche se non l'unico, per decidere è il **test di significatività**: chiede se il valore osservato sia compatibile con una popolazione in cui la correlazione è zero. Il valore *p* misura questa compatibilità. Non misura la grandezza dell'effetto né la probabilità che l'ipotesi sia vera (Wasserstein & Lazar, 2016). Una correlazione di 0,28 può essere «non significativa» con 30 persone e una di 0,20 «significativa» con 200: a cambiare è la precisione della stima, non la grandezza dell'effetto. Per questo conviene riportare sempre anche un intervallo di confidenza, che mostra l'ampiezza dell'incertezza.

13.6 Spiegare perché correlano

Se due variabili correlano davvero, le possibilità principali sono quattro: A causa B; B causa A; le due si influenzano a vicenda, magari rafforzandosi nel tempo; oppure esistono cause comuni.

13.6 Spiegare perché correlano

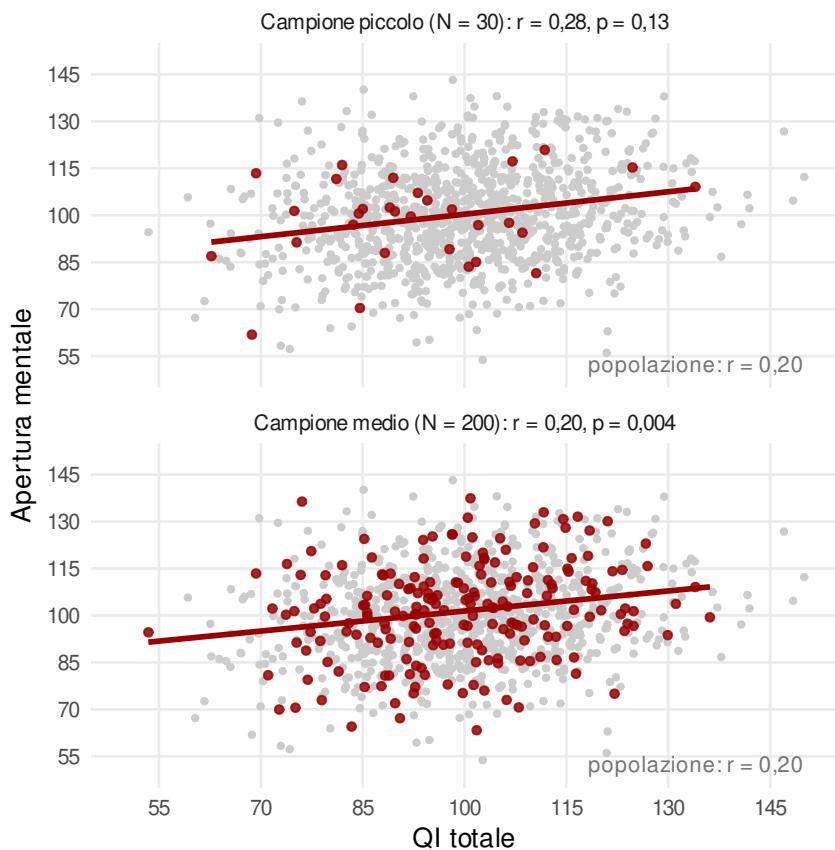


Figura 13.6: Lo stesso effetto visto con due campioni. In grigio una popolazione simulata di mille persone in cui la correlazione vera fra QI e apertura mentale è 0,20; in rosso le persone estratte a caso in un campione piccolo e in uno medio, con la retta stimata su di esse. Il campione piccolo stima una correlazione più forte, ma non permette di distinguere da zero; quello medio stima una correlazione più debole, ma statisticamente attendibile.

Un caso che riguarda proprio l'intelligenza è la correlazione fra QI e istruzione (Figura 13.7). L'intelligenza favorisce la prosecuzione degli studi? L'istruzione aumenta le abilità cognitive? Le due cose si alimentano a vicenda nel tempo? Oppure condizioni familiari come lo status socioeconomico influenzano entrambe? Per rispondere non basta la correlazione. Servono studi longitudinali, che mostrino se una variazione individuale nell'una anticipa una variazione nell'altra, e analisi che verifichino se la correlazione si riduce confrontando persone con condizioni ambientali simili. La risposta che emerge da diversi studi è che sono all'opera sia una causalità reciproca, compreso un effetto causale dell'istruzione sui punteggi, sia cause comuni come lo status socioeconomico della famiglia (Ritchie & Tucker-Drob, 2018). Anche gli esperimenti naturali, come i cambiamenti di legge sull'obbligo scolastico, sono stati decisivi.

L'esempio dei gelati è più netto. Nei giorni in cui si vendono più gelati avvengono anche più omicidi e più annegamenti. Nessuno pensa che i gelati rendano violenti o facciano annegare: il caldo porta le persone fuori casa, in spiaggia e in strada. La temperatura è una **causa comune**. In rete si trovano intere raccolte di correlazioni assurde, come quella fra numero di pirati e riscaldamento globale, ottenute confrontando serie temporali che crescono o calano per motivi indipendenti. Fanno sorridere, ma ricordano un fatto serio: una correlazione da sola non indica la direzione causale né esclude terze variabili (Rohrer, 2018). Correlazione non significa causalità, anche se spesso, dietro, qualche causalità c'è. E la sequenza temporale aiuta senza risolvere tutto: se A precede B, B non può aver causato A, ma resta possibile che una terza variabile abbia preceduto entrambe.

13.7 Mediazione e moderazione

Due concetti aiutano a precisare le ipotesi causali. Si parla di **mediazione**, o causazione indiretta, quando A influenza B attraverso una o più variabili intermedie. Un consumo elevato di alcol peggiora il rendimento agli esami? Se sì, probabilmente non in modo diretto, ma attraverso più percorsi: effetti sul sonno e sulla memoria, e situazioni sociali che sottraggono tempo allo studio (Figura 13.8). Una mediazione stimata con dati raccolti in

13.7 Mediazione e moderazione

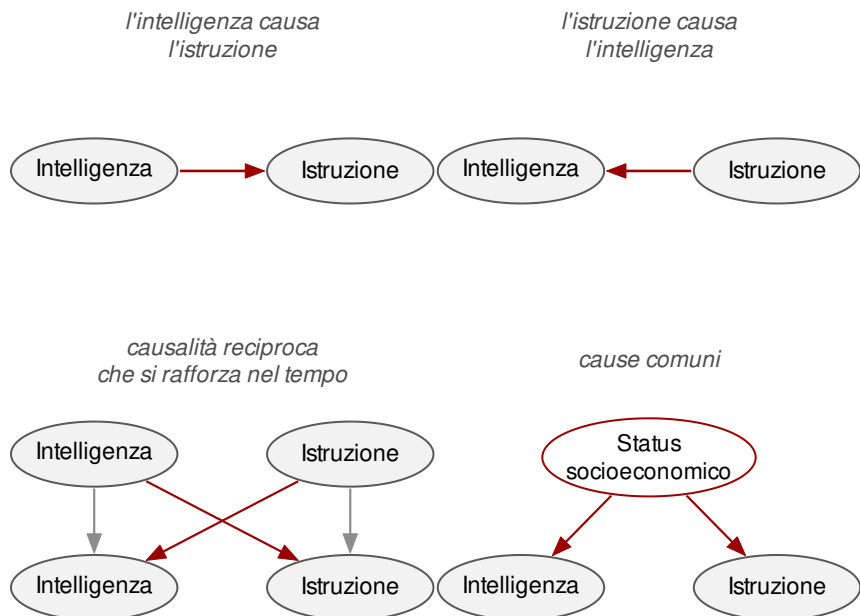


Figura 13.7: Quattro spiegazioni possibili della correlazione fra intelligenza e istruzione. Le frecce rappresentano ipotesi causali; le evidenze disponibili sostengono in parte sia la causalità reciproca sia le cause comuni.

un unico momento, però, non dimostra il processo nel tempo (Maxwell & Cole, 2007).

Si parla di **moderazione**, o causazione modificata, quando la relazione fra A e B cambia in funzione di una terza variabile (Baron & Kenny, 1986). La presenza di osservatori tende a migliorare la prestazione in compiti già automatizzati e a peggiorarla in compiti nuovi (Zajonc, 1965): il tipo di compito modera l'effetto del pubblico. Nel campo dell'intelligenza un esempio discusso riguarda lo status socioeconomico, che in alcuni studi statunitensi sembra modificare quanto le differenze genetiche si traducono in differenze di punteggio (Turkheimer et al., 2003). Entrambi i casi sono nella Figura 13.9. I limiti di questo risultato si sono visti nel capitolo sull'ereditabilità.

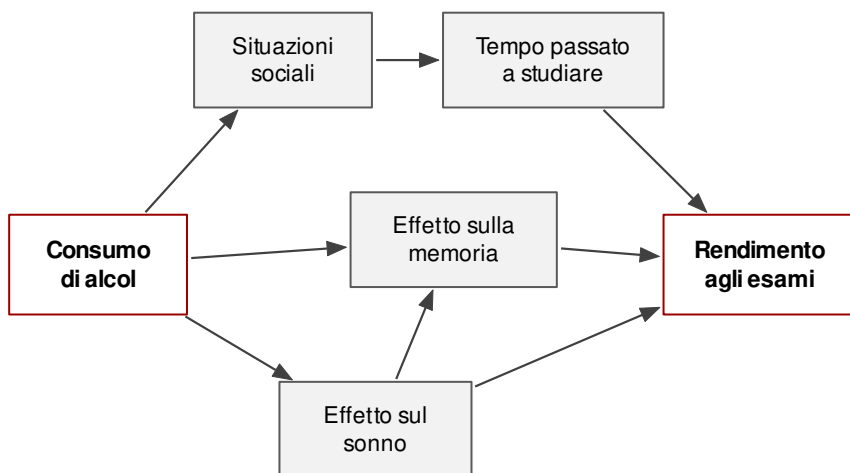


Figura 13.8: Mediazione: il consumo di alcol potrebbe influire sul rendimento agli esami attraverso più percorsi indiretti, cioè gli effetti su sonno e memoria e le situazioni sociali che riducono il tempo di studio. Schema ipotetico.

13.8 Quattro controlli prima di fidarsi

Una correlazione è uno dei risultati più semplici della ricerca e uno dei più facili da interpretare male. Prima di trarne conclusioni conviene controllare quattro cose: il segno, che dice la direzione; l'intensità, da leggere nel contesto della domanda; la forma dei dati, che può rendere il coefficiente fuorviante; il disegno dello studio, che decide quali spiegazioni restano in gioco. Saltare l'ultimo passaggio è l'errore più comune. Ed è quello che produce i titoli di giornale peggiori.

In breve

- La correlazione di Pearson riassume direzione e intensità dell'associazione lineare fra due misure, da -1 a $+1$, indipendentemente dalle unità di misura.
- $r = 0,50$ significa che chi è una deviazione standard sopra la

13.8 Quattro controlli prima di fidarsi

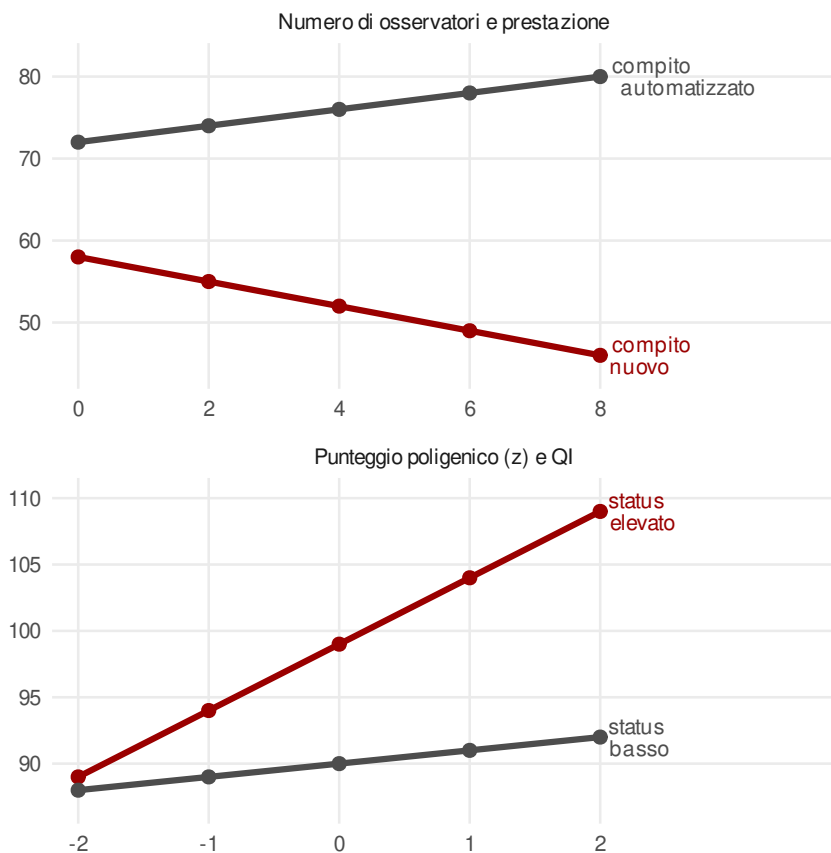


Figura 13.9: Due esempi di moderazione. Nel primo pannello, al crescere del numero di osservatori la prestazione migliora in un compito già automatizzato e peggiora in un compito nuovo. Nel secondo, il legame fra punteggio poligenico e QI è più forte in condizioni socioeconomiche elevate che in condizioni svantaggiate. Valori illustrativi.

media in una variabile è, in media, mezza deviazione standard sopra la media nell'altra; in psicologia valori attorno a 0,20 sono tipici e 0,50 è un'associazione forte.

- Con una correlazione imperfetta la previsione è sempre meno estrema del dato di partenza: è la regressione verso la media.
- Casi estremi e relazioni non lineari possono rendere r fuorviante: i dati vanno sempre guardati.
- Con campioni piccoli la stima è molto imprecisa; il valore p non misura né la grandezza dell'effetto né la probabilità che un'ipotesi sia vera.
- Correlazione non significa causalità: direzione inversa, causalità reciproca e cause comuni vanno valutate con disegni adeguati; mediazione e moderazione precisano come una causa agisce.

Riferimenti bibliografici

14 Quando una misura è buona

Strumenti. Le due parti «Strumenti» raccolgono le basi tecniche del libro. I loro capitoli si possono leggere da soli, in qualsiasi momento, oppure saltare: i capitoli concettuali restano comprensibili e vi rimandano, con brevi note in verde, quando un passaggio richiede uno strumento in più.

Questo capitolo tratta i criteri con cui si giudica un test: validità, attendibilità come stabilità e come consistenza interna, errore di misura, invarianza. Presuppone la **correlazione** e il modello a un fattore del capitolo sul **fattore g**. Serve per interpretare un singolo punteggio, per i correttivi statistici del capitolo sul lavoro e per la valutazione nel neurosviluppo.

Una bilancia da cucina che sbaglia di quindici grammi è inutile per dosare il lievito. Una bilancia pesapersone che sbaglia di tre chili è inutile per seguire una dieta. Entrambe misurano il peso, e nessuno lo mette in dubbio. Eppure nessuno le userebbe per quegli scopi. Misurare la cosa giusta non basta: serve misurarla abbastanza bene per l'uso che se ne fa. In psicologia lo stesso ragionamento vale per ogni test, con una complicazione: ciò che vogliamo misurare non si vede.

Le abilità cognitive sono variabili latenti. Le stimiamo a partire da risposte a compiti, e il punteggio che ne ricaviamo è il risultato di una catena di inferenze. Come facciamo a sapere che quel numero va bene? La psicomетria ha sviluppato una serie di criteri. Questo capitolo li presenta e li applica ai test di intelligenza, che nel panorama della misura psicologica hanno indici di qualità fra i più alti. Proprio per questo è utile vedere dove anche loro incontrano dei limiti.

14.1 Quattro domande su un test

I criteri principali possono essere riassunti in quattro domande.

- **Validità.** Il test misura davvero ciò che dovrebbe misurare, e permette le interpretazioni e le previsioni per cui viene usato?
- **Attendibilità**, o affidabilità. Il punteggio è preciso e stabile? Se ripetessi la misura, otterrei un risultato simile?
- **Sensibilità.** Lo strumento distingue bene i diversi livelli di abilità? Se è usato a scopo diagnostico, separa bene chi ha una condizione da chi non ce l'ha?
- **Robustezza.** Il test resta valido e attendibile anche in condizioni non ideali: una popolazione diversa da quella di riferimento, una condizione del neurosviluppo, una somministrazione non standard, una ripetizione ravvicinata?

Le batterie di intelligenza ben costruite superano bene le prime tre prove. La quarta dipende dalla situazione: nessun test è robusto in ogni circostanza, ma esistono contromisure note. Vediamole una per una.

14.2 Validità: che cosa misura il test?

Nella formulazione moderna la validità non è una proprietà del test, ma del sostegno che evidenze e teoria offrono a una specifica interpretazione dei punteggi per un certo uso (American Educational Research Association et al., 2014). Si accumula attraverso diversi tipi di evidenza.

La **validità di contenuto** chiede se le prove coprono in modo appropriato il dominio da misurare. Per l'intelligenza la questione ha una storia curiosa. Spearman osservò che quasi qualunque compito cognitivo, purché impegnativo, fornisce informazioni sul fattore generale. Chiamò questo fenomeno **indifferenza dell'indicatore** (Spearman, 1927). Costruire un test che misuri il fattore generale è quindi, in un certo senso, facile: basta combinare prove abbastanza diverse e bilanciarle in modo sensato (Warne, 2020). Le batterie moderne preferiscono comunque una base teorica esplicita, di solito il modello gerarchico visto nel capitolo sui modelli gerarchici.

La **validità strutturale** chiede se le correlazioni fra item, subtest e indici seguono un modello teorico plausibile. Si verifica con l'analisi fattoriale confermativa, valutando quanto il modello si adatta ai dati.

14.3 Attendibilità come stabilità

La **validità convergente** chiede se il test correla con altri test dello stesso costruito. Per l'intelligenza è eccellente: i QI ottenuti con strumenti diversi correlano in genere almeno 0,70, spesso molto di più.

La **validità predittiva** chiede se il punteggio prevede esiti che, in base alla teoria, dovrebbe prevedere: rendimento scolastico, formazione e prestazione lavorativa, salute. È il punto di forza dei test di intelligenza, e vi dedicheremo un'intera parte del libro.

La **validità discriminante** chiede se il test non misuri anche cose che non dovrebbe misurare. Per l'intelligenza è un criterio delicato: un'abilità generale ha molte ragioni teoriche per correlare con molte altre variabili, e una correlazione con lo status socioeconomico, per esempio, non è di per sé un segno di contaminazione.

Infine l'**invarianza di misura** chiede se il punteggio si interpreta allo stesso modo in gruppi diversi. Possiamo leggere il QI totale di un bambino con ADHD o con autismo come quello di un bambino a sviluppo tipico? E quello di un bambino cresciuto in un'altra lingua? Non è una proprietà fissa del test, ma della combinazione fra test, gruppo e interpretazione. Torneremo su questo punto nel capitolo sulla valutazione nel neurosviluppo.

14.3 Attendibilità come stabilità

Un modo diretto di valutare la precisione è somministrare lo stesso test due volte e correlare i punteggi. È l'**attendibilità test-retest**. Le differenze fra le due misure possono avere due origini: un errore casuale, legato a oscillazioni dell'attenzione, all'ora del giorno o alla stanchezza; oppure un cambiamento reale del funzionamento. Per il QI i cambiamenti reali a breve termine sono modesti. Esiste anche un effetto pratica, perché la seconda volta le prove sono meno nuove, ma è abbastanza prevedibile da poter essere considerato. In genere una stabilità sotto 0,70 si considera problematica, se si possono escludere cambiamenti reali.

Nelle batterie per adulti il QI totale, risomministrato dopo poche settimane, ha correlazioni test-retest attorno a 0,95: nel campione di verifica della WAIS-IV, circa 300 adulti ritestati dopo tre settimane, è 0,96 (Wechsler,

2008). Sembra quasi perfetto. Ma che cosa significa per un singolo caso? La Figura 14.1 lo mostra con una simulazione. Anche con una stabilità di 0,96 la differenza tipica fra due somministrazioni è di circa quattro punti, e in un caso su venti supera gli otto. Con una stabilità di 0,70 la differenza tipica sale a circa dodici punti e in un caso su venti supera i venti.

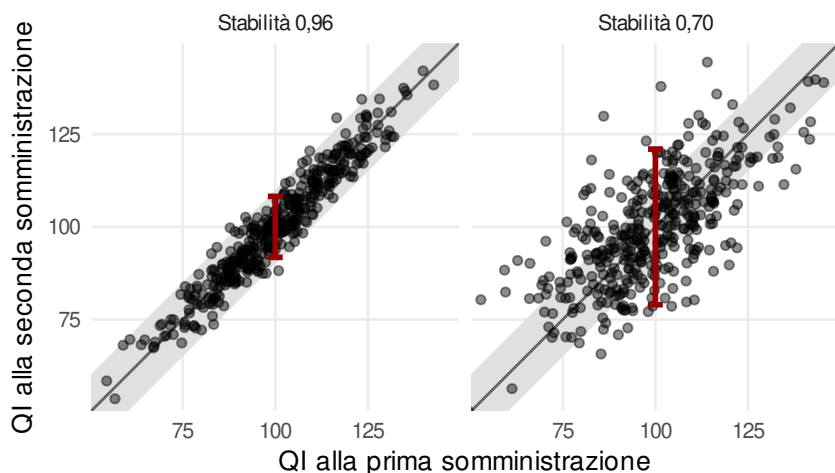


Figura 14.1: Punteggi di QI simulati in due somministrazioni dello stesso test, con stabilità test-retest di 0,96, come per il QI totale ripetuto dopo poche settimane, e di 0,70, vicina a quella degli indici su intervalli di qualche anno in età evolutiva. La linea diagonale indica punteggi identici; la fascia grigia comprende le differenze entro ± 10 punti. La barra rossa mostra dove cade il 95% dei punteggi alla seconda somministrazione per chi aveva 100 alla prima.

In età evolutiva e su intervalli lunghi la stabilità è inferiore. In uno studio su 344 studenti valutati due volte con la WISC-IV a quasi tre anni di distanza, la stabilità del QI totale era 0,82, quella degli indici fra 0,65 e 0,76. Un quarto degli studenti aveva QI totali che differivano di dieci punti o più (Watkins & Smith, 2013). Una parte della variazione è certamente reale: i bambini cambiano, e una valutazione può dover essere aggiornata. Ma la conclusione pratica non cambia: un singolo punteggio, soprattutto di un indice, non è una fotografia definitiva.

14.4 Attendibilità come consistenza interna

Nel lungo periodo, però, la stabilità dell'ordinamento fra le persone è notevole. Nella coorte scozzese testata a 11 anni nel 1932 e ritestata quasi settant'anni dopo, la correlazione fra i punteggi alle due età era dell'ordine di 0,6–0,7 (Deary et al., 2000, 2004). Poche caratteristiche psicologiche mostrano una continuità simile lungo l'intera vita adulta.

14.4 Attendibilità come consistenza interna

Il secondo modo di stimare l'attendibilità non richiede due somministrazioni. Guarda alla coerenza fra le parti di un test: quanto gli item, o i subtest, misurano la stessa cosa.

Il ragionamento parte dalla **teoria classica dei test**. Il punteggio osservato X è la somma di un punteggio vero T , che rappresenta il costrutto latente, e di un errore E :

$$X = T + E$$

Nel caso dell'intelligenza il punteggio vero è il fattore generale, e l'errore comprende sia l'errore casuale sia le componenti specifiche s di ogni prova. L'attendibilità è la quota della varianza di X dovuta a T . Come si stima, se T non si vede? Sotto certe assunzioni, la si ricava dalla forza delle correlazioni fra gli elementi e dal loro numero. Correlazioni medie più alte indicano che ogni elemento contiene meno errore; più elementi permettono agli errori casuali di compensarsi. Un punteggio molto attendibile si ottiene quindi con elementi fortemente correlati, oppure con molti elementi.

L'indice più noto è l'**alfa di Cronbach** (α), che lega proprio queste due grandezze, la correlazione media fra gli elementi $\bar{r}$ e il loro numero k :

$$\alpha = \frac{k \cdot \bar{r}}{1 + (k - 1) \cdot \bar{r}}$$

L'alfa assume che tutti gli elementi rappresentino il costrutto nella stessa misura e con la stessa metrica, un'ipotesi spesso irrealistica. Oggi si preferiscono indici basati su un modello fattoriale, come l'**omega di McDonald**

(ω) (McNeish, 2018). Anche qui si cita spesso una soglia minima di 0,70. La Figura 14.2 mostra come cambia l'alfa al variare di $\bar{r}$ e di k .

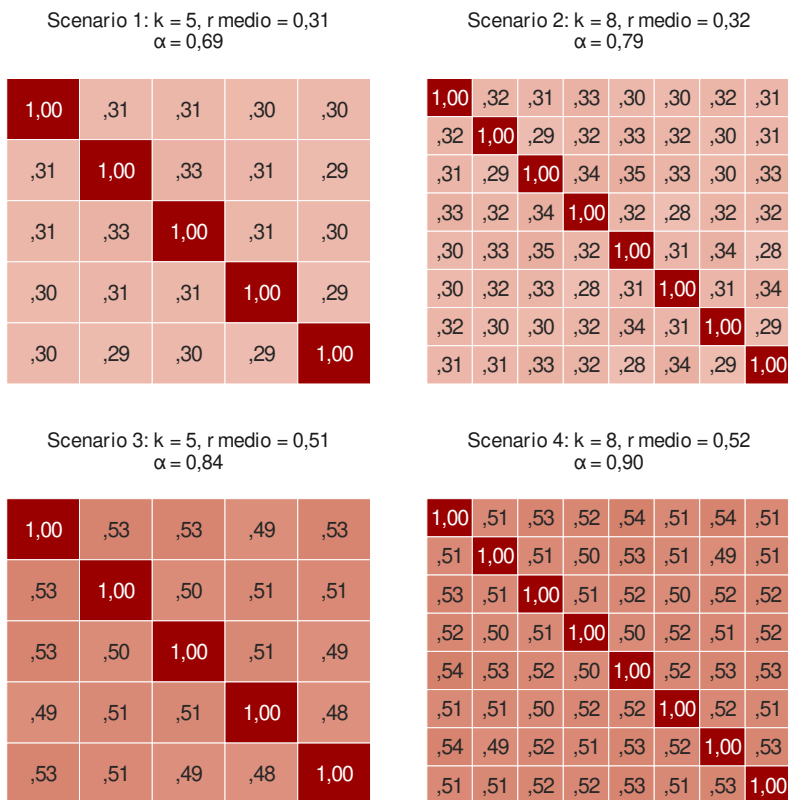


Figura 14.2: Quattro scenari di consistenza interna. Ogni pannello mostra la matrice di correlazioni fra gli elementi di una scala: con correlazione media 0,30 o 0,50 e con 5 o 8 elementi. L'alfa di Cronbach cresce sia con la correlazione media sia con il numero di elementi. Dati simulati.

Che cosa significa un'attendibilità di 0,70? Che il punteggio osservato condivide il 70% della varianza con il punteggio vero del costrutto latente. Vale per la somma semplice dei punteggi, nel caso dell'alfa, e per il punteggio

14.4 Attendibilità come consistenza interna

fattoriale calcolato dal modello, nel caso dell'omega. La correlazione fra punteggio osservato e punteggio vero è quindi la radice quadrata: circa 0,84 (Figura 14.3). Le batterie di intelligenza raggiungono per il QI totale valori superiori a 0,95. Per gli indici di dominio si scende di solito fra 0,80 e 0,90, per i singoli subtest a volte sotto; nel complesso, gli indici di intelligenza si collocano fra circa 0,70 e 0,98, valori comunque eccellenti nel panorama dei test psicologici.

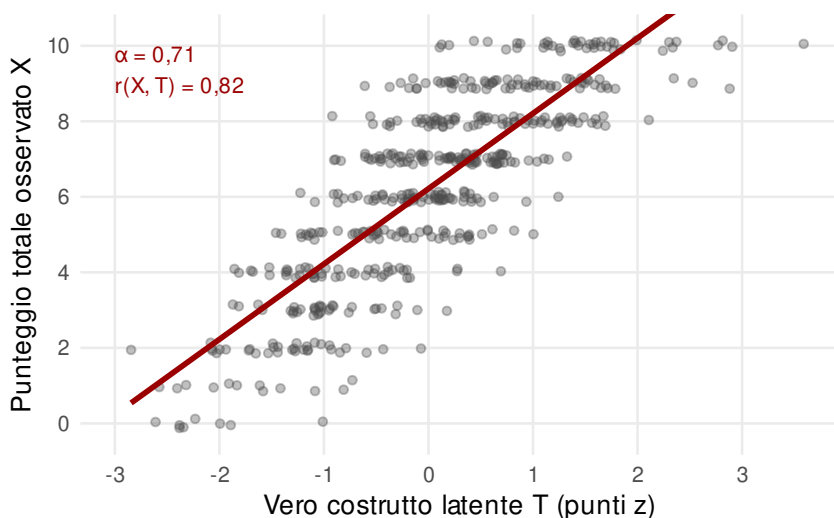


Figura 14.3: La relazione fra il vero costruito latente (T, in punti z) e il punteggio totale osservato (X) in una scala di dieci item con α di circa 0,70: la correlazione fra i due, 0,82, è vicina a $\sqrt{0,70} \approx 0,84$. Dati simulati.

Dall'attendibilità si ricava l'**errore standard di misura**, cioè la deviazione standard degli errori attorno al punteggio vero. Sulla scala del QI vale $15 \times \sqrt{(1 - \text{attendibilità})}$. Con un'attendibilità di 0,97 è di circa 2,6 punti; con 0,85 sale a quasi 6. Per questo un referto serio riporta il punteggio con un intervallo di confidenza, non come un numero esatto.

14.5 Il problema degli indici

Fin qui il quadro è rassicurante. Il problema nasce quando dal QI totale si passa agli indici: comprensione verbale, ragionamento visuo-percettivo, memoria di lavoro, velocità di elaborazione e così via. Nella pratica clinica si interpretano spesso profili e discrepanze fra indici. Sono abbastanza attendibili per farlo?

L'attendibilità di un indice, calcolata come se fosse una scala a sé, è buona. Ma un indice di memoria di lavoro riflette in buona parte il fattore generale, perché tutte le abilità cognitive sono correlate. La domanda giusta è un'altra: quanta parte del punteggio di un indice è specifica di quel dominio, al netto del fattore generale? E questa parte è attendibile?

Un esempio sui dati normativi italiani della WISC-IV, 2.200 bambini e ragazzi, rende il problema concreto. È un'analisi originale, condotta per questo libro sui dieci subtest principali; i risultati concordano con quelli pubblicati sullo stesso campione con tutti i quindici subtest (Kush & Canivez, 2019). Se si applica ai dieci subtest principali un modello a un solo fattore, tutti risultano legati a g , con saturazioni da 0,28 a 0,75, e la consistenza interna del totale è buona: $\omega = 0,80$, $\alpha = 0,82$ (Figura 14.4). Ma questo non è il modello giusto, perché l'intelligenza non è strettamente unitaria: prove dello stesso dominio correlano fra loro più di quanto g da solo spieghi.

Per rispondere si usano modelli gerarchici o **bifattoriali**, che separano la varianza comune a tutte le prove da quella propria di ciascun dominio. Da essi si ricava l'**omega gerarchico** (ω_h), che stima quanta varianza del punteggio è dovuta a un solo fattore (Rodriguez et al., 2016) (Figura 14.5). I risultati sui campioni normativi delle principali batterie sono convergenti. Il fattore generale ha un omega gerarchico elevato, attorno a 0,75–0,80. I domini specifici, al netto di g , scendono spesso sotto 0,30 e talvolta vicino a zero (Gignac & Kretzschmar, 2017; Watkins, 2006). Le rianalisi delle edizioni più recenti, come la WISC-V, mostrano lo stesso schema (Canivez et al., 2017), e così i dati normativi italiani. Nel campione di taratura italiano della WISC-IV, con tutti i quindici subtest, il fattore generale ha un omega gerarchico di 0,80, mentre quello dei fattori di dominio non va oltre 0,20, con l'eccezione della velocità di elaborazione, che arriva a 0,52 (Kush & Canivez, 2019); lo stesso modello è stato applicato anche al confronto con

14.5 Il problema degli indici

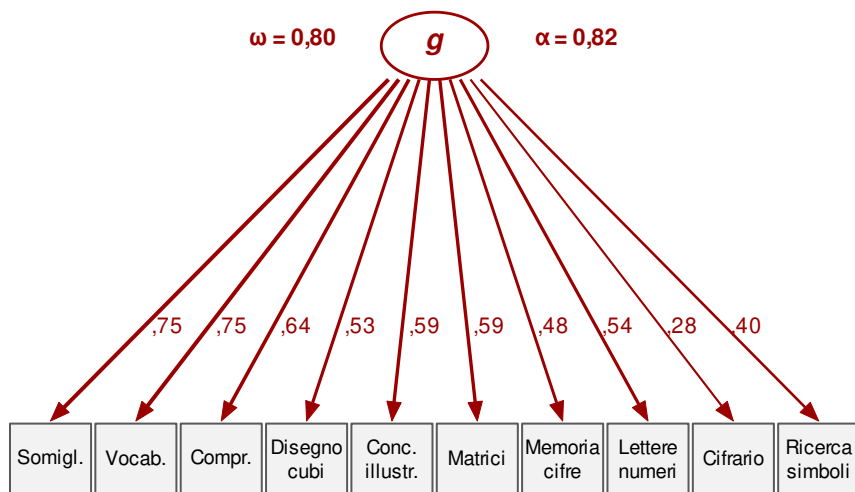


Figura 14.4: Un modello a un solo fattore per i dieci subtest principali della WISC-IV, con le saturazioni di ciascun subtest su *g* e gli indici di consistenza interna del totale. Saturazioni e coefficienti da un'analisi originale dei dati normativi italiani ($N = 2.200$).

ampi campioni di bambini con disturbi dell'apprendimento (Giofrè et al., 2019). Nel campione di taratura italiano dell'IDS-2, su 1.281 bambini e ragazzi, il fattore generale ha un omega gerarchico di 0,77, mentre quello dei fattori specifici varia fra 0,08 e 0,51 (Toffalini & Cornoldi, 2026). In tutti questi casi il fattore generale risulta attendibile e interpretabile, i fattori di dominio quasi mai abbastanza da essere letti da soli.

Le conclusioni tratte da questi dati sono di due tipi. Alcuni ricercatori sostengono che sia semplicemente inappropriato interpretare gli indici di dominio e che ci si debba limitare al QI totale (Gignac & Kretzschmar, 2017; Watkins, 2006). Una posizione più equilibrata ammette l'interpretazione degli indici, ma solo se motivata dal caso specifico e sostenuta da evidenze convergenti: per esempio quando si valuta una condizione per cui un certo profilo è atteso sulla base della letteratura (Giofrè et al., 2017; Toffalini et al., 2017). In entrambi i casi la lettura disinvoltata di ogni discrepanza fra indici, senza tener conto dell'errore di misura, non regge.

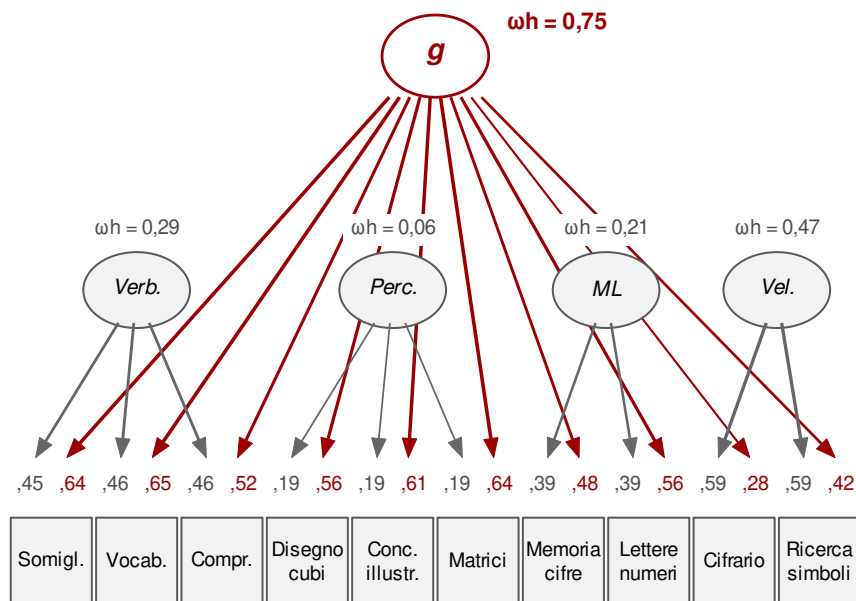


Figura 14.5: Lo stesso esempio con un modello bifattoriale: ogni subtest dipende dal fattore generale g (freccie rosse) e da un fattore di dominio (freccie grigie), che cattura solo ciò che resta al netto di g . L'omega gerarchico (ω_h) di g resta alto; quello dei domini è basso. Saturazioni e coefficienti da un'analisi originale dei dati normativi italiani della WISC-IV ($N = 2.200$).

14.6 Sensibilità e robustezza

Un test sensibile distingue bene i livelli di abilità in tutto l'intervallo che interessa. Le batterie di intelligenza coprono per progetto un ampio intervallo, di solito da QI attorno a 40 fino a 160. Gli effetti tetto, in cui molte persone raggiungono il punteggio massimo, sono rari e gestiti con versioni diverse dello stesso strumento: forme per bambini, forme avanzate per adulti con abilità elevate, prove non verbali per chi ha difficoltà linguistiche marcate.

La robustezza è più difficile da garantire. Una persona con una disabilità

motoria non può essere valutata con prove a tempo che richiedono manipolazione; una persona con ADHD può essere penalizzata in modo sproporzionato dalle prove di memoria di lavoro e velocità; una persona che non padroneggia la lingua italiana sarà svantaggiata nelle prove verbali. Per questi casi esistono contromisure: indici che escludono le componenti più problematiche, come l'indice di abilità generale al posto del QI totale; batterie non verbali; sostituzioni di subtest e procedure di stima parziale previste dai manuali. Le contromisure funzionano se si sa quando servono. Usarle richiede di conoscere la persona, non solo lo strumento.

14.7 Una misura buona, non infallibile

I test di intelligenza ben costruiti sono fra gli strumenti più validi e attendibili della psicologia. Questo giustifica la fiducia nel QI totale come stima del funzionamento cognitivo generale. Non giustifica la fiducia indiscriminata in ogni numero che compare in un referto. Il QI ha un margine di errore che va dichiarato. Gli indici di dominio portano un'informazione specifica modesta e vanno interpretati con argomenti. La robustezza dipende da chi è la persona valutata.

In breve

- La qualità di un test si giudica per validità, attendibilità, sensibilità e robustezza; la validità riguarda un'interpretazione per un uso, non il test in sé.
- Le batterie di intelligenza hanno validità convergente e predittiva elevate e una stabilità notevole anche su decenni.
- Anche un'attendibilità di 0,95 lascia differenze di alcuni punti fra due somministrazioni; i punteggi vanno riportati con un intervallo di confidenza.
- Gli indici di dominio, al netto del fattore generale, sono poco attendibili: la loro interpretazione va motivata e corroborata.

Riferimenti bibliografici

Parte VI

Scuola, salute e lavoro

15 Intelligenza, scuola e condizioni sociali

I test di intelligenza sono nati per prevedere. Binet voleva sapere quali bambini avrebbero avuto difficoltà a trarre profitto dalla scuola. I test dell'esercito dovevano indicare chi avrebbe sostenuto ruoli complessi. Wechsler cercava di stimare la capacità di vivere in autonomia. Questa parte del libro verifica quanto quelle previsioni funzionino davvero, partendo dall'ambito per cui i test sono stati inventati: la scuola.

Una premessa vale per i tre capitoli. Le correlazioni fra QI ed esiti rilevanti vanno da circa 0,20 a 0,50, e superano quella soglia solo nella scuola. Per la psicologia sono valori da medi a molto alti, e si indeboliscono man mano che l'esito si allontana dalle aule: forti con i voti, moderate con la prestazione lavorativa, modeste con il reddito. Ma anche una correlazione di 0,50 lascia la maggior parte della varianza non spiegata. Il QI pesa, a volte molto. Non è un destino.

15.1 Una previsione robusta

Il QI prevede il rendimento scolastico in qualsiasi materia. In uno studio su oltre 70.000 studenti inglesi, i punteggi a una batteria cognitiva somministrata a 11 anni sono stati confrontati con i risultati degli esami nazionali sostenuti a 16. Le correlazioni con le singole materie andavano da circa 0,4 nelle discipline artistiche a quasi 0,8 in matematica. Fra il fattore cognitivo generale e un fattore generale di rendimento negli esami la correlazione era circa 0,8 (Deary et al., 2007). Una meta-analisi su studi di tutto il mondo colloca la correlazione fra intelligenza e voti scolastici attorno a 0,54 (Roth et al., 2015).

Perché una relazione così forte? Il QI riflette capacità che la scuola richiede continuamente: memorizzare, astrarre, cogliere regole implicite, adattarsi

a situazioni poco chiare, generalizzare ciò che si è imparato. Il QI, inoltre, misura queste capacità in modo più preciso di quanto possano fare osservatori esperti. Il giudizio degli insegnanti sulle abilità cognitive degli studenti correla con i test in misura apprezzabile, ma lontana dalla perfezione (Machts et al., 2016).

15.2 Una correlazione non è un'identità

Data una relazione così stretta, alcuni ricercatori usano direttamente voti, punteggi nelle indagini come PISA o anni di istruzione come indicatori dell'intelligenza. I punteggi PISA, in effetti, riflettono in larga parte un fattore generale (Pokropek et al., 2022). Ma la scorciatoia è pericolosa. Il QI, misurato bene, è una stima del funzionamento cognitivo attuale, non di un potenziale innato: anch'esso risente della scolarizzazione, come si è visto nel capitolo sul potenziamento. Ma ne risente molto meno del rendimento scolastico, che dipende direttamente dalla qualità delle scuole, dai programmi, dalla motivazione, dalle condizioni familiari, dal modo in cui una prova viene presentata.

Un caso italiano mostra il rischio. Lo psicologo Richard Lynn sostenne che gli italiani del Sud avessero un QI medio di circa dieci punti inferiore a quello degli italiani del Nord, e propose una spiegazione genetica. A sostegno citava correlazioni con reddito, mortalità infantile, statura e livelli di istruzione (Lynn, 2010). Ma i suoi «QI regionali» non erano QI: erano punteggi PISA di quindicenni in lettura, matematica e scienze.

Le repliche hanno messo in luce molti problemi (Cornoldi et al., 2010). Il divario Nord-Sud nei risultati scolastici diventa evidente nella scuola secondaria, dove il sistema educativo del Sud è storicamente più debole; se riflettesse l'intelligenza, dovrebbe essere visibile già alla primaria. Gli studenti del Sud ottengono buoni risultati nelle prove somministrate individualmente e peggiori in quelle collettive, segno di un diverso coinvolgimento nel compito. Fra un'edizione e l'altra di PISA i punteggi del Sud cambiavano in modo troppo rapido per riflettere differenze stabili di capacità. E i dati non coincidevano con quelli delle prove nazionali INVALSI. La lezione è generale: si può inferire l'intelligenza dagli apprendimenti

solo dove le condizioni di apprendimento sono uniformi. Fra regioni con sistemi scolastici e contesti diversi non lo sono.

15.3 Intelligenza e istruzione: chi causa chi?

Il QI misurato nell'infanzia e il livello di istruzione raggiunto da adulti sono fortemente correlati, con valori oltre 0,50 (Strenze, 2007). Le spiegazioni possibili sono le solite tre.

***Strumenti** → Le spiegazioni alternative di una correlazione, con questo stesso esempio: Correlazione e differenze individuali.*

L'intelligenza favorisce l'istruzione. Chi apprende più facilmente riceve più rinforzi positivi, sviluppa più fiducia nelle proprie capacità, sperimenta meno fallimenti, viene incoraggiato a proseguire. Nelle coorti britanniche nate nel Novecento, un QI di una deviazione standard più alto a 10–11 anni si associava a odds quasi triplicate di ottenere una laurea entro i 26 anni («Childhood cognitive ability and adult academic attainment», 2010). È un esempio di **effetto Matteo**, dal versetto evangelico secondo cui a chi ha sarà dato ancora: piccoli vantaggi iniziali si accumulano e si amplificano (Stanovich, 1986).

L'istruzione aumenta l'intelligenza. Come si è visto nel capitolo sul potenziamento, la scuola produce stimoli cognitivi, abitua all'astrazione, fornisce conoscenze e familiarità con le situazioni di valutazione. Le stime causali vanno da uno a cinque punti di QI per anno di istruzione, a seconda del disegno di studio (Ritchie & Tucker-Drob, 2018).

Cause comuni. Lo status socioeconomico della famiglia influenza sia lo sviluppo cognitivo sia le opportunità di studio. È certamente un fattore confondente. Nei paesi occidentali con istruzione, sanità e opportunità relativamente universali, tenerne conto riduce le stime, ma non le stravolge.

Le tre spiegazioni non si escludono. Le evidenze indicano che sono tutte e tre all'opera, e che si rinforzano a vicenda nel tempo.

La Figura 15.1 mette a confronto le due direzioni causali con le stime principali. Nelle coorti britanniche la correlazione fra QI a 10–11 anni ed età di uscita dagli studi si riduce, ma resta sostanziale, quando si tiene conto dello status socioeconomico della famiglia («Childhood cognitive ability and adult academic attainment», 2010): è il motivo per cui le cause comuni pesano, ma non spiegano tutto.

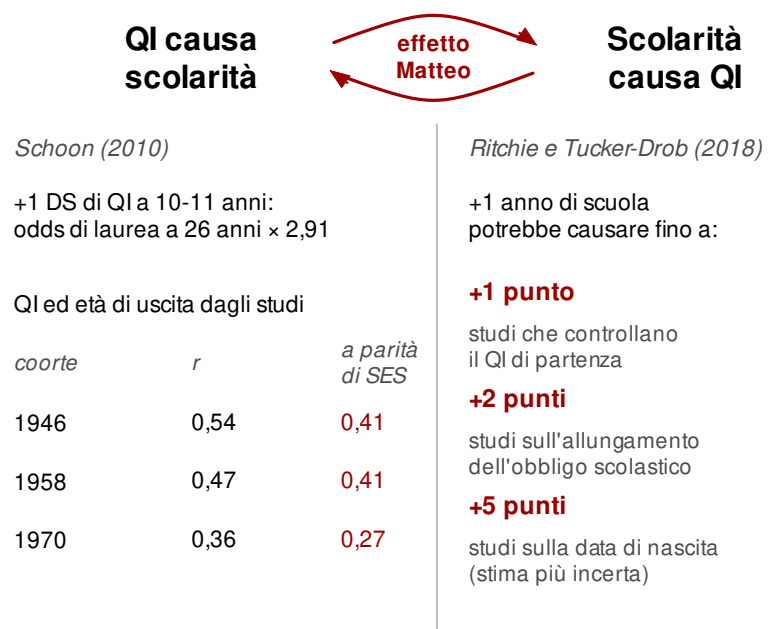


Figura 15.1: QI e livello di scolarità si influenzano nei due sensi e si rinforzano a vicenda nel tempo (effetto Matteo). Prima colonna: il QI nell’infanzia prevede la laurea e la durata degli studi, anche a parità di status socioeconomico (Schoon, 2010). Seconda colonna: stime dell’effetto causale di un anno di scuola sul QI, secondo il tipo di studio (Ritchie e Tucker-Drob, 2018).

15.4 Dall'istruzione allo status sociale

Se il QI prevede l'istruzione, prevede anche lo status lavorativo e il reddito? Una meta-analisi di studi longitudinali, in cui l'intelligenza era misurata prima degli esiti, ha stimato correlazioni attorno a 0,56 con il livello di istruzione, 0,43 con il prestigio della professione e 0,20 con il reddito (Strenze, 2007). L'intelligenza prevede quindi il successo socioeconomico, ma il legame si indebolisce man mano che ci si allontana dalla scuola.

Il mediatore principale è l'istruzione. Nei paesi industrializzati ogni anno di istruzione in più si associa in media a un reddito più alto di circa il 10% (Psacharopoulos & Patrinos, 2018). Il percorso tipico è quindi quello della Figura 15.2: l'intelligenza favorisce il successo scolastico, che porta a un livello di istruzione più alto, che a sua volta apre posizioni lavorative migliori e quindi un reddito maggiore. L'istruzione, come si è visto, retroagisce anche sull'intelligenza. L'effetto diretto dell'intelligenza sullo status e sul reddito, a parità di istruzione, esiste ma è più piccolo.

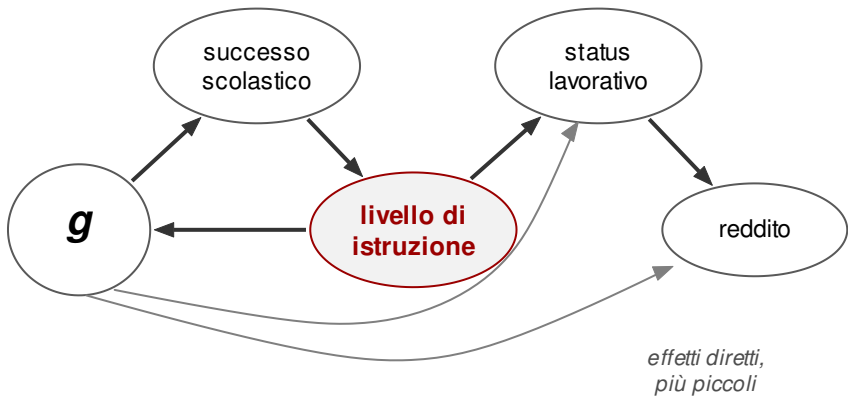


Figura 15.2: L'istruzione come mediatore fra livello intellettuale generale e reddito. Le frecce spesse indicano il percorso principale, che passa per successo scolastico, livello di istruzione e status lavorativo; l'istruzione a sua volta aumenta g. Le frecce sottili indicano gli effetti diretti di g su status lavorativo e reddito, a parità di istruzione, più piccoli.

Uno studio sui partecipanti della UK Biobank permette di vedere la differenza fra effetto totale ed effetto diretto (Figura 15.3). Usando varianti genetiche come strumento per stimare effetti causali, una deviazione standard di intelligenza in più aumentava di circa 11–12 punti percentuali la probabilità di avere un reddito familiare sopra le soglie considerate; a parità di livello di istruzione, l'aumento si riduceva a circa 5 punti (N. M. Davies et al., 2019). Le stime di questo tipo poggiano su assunzioni forti, ma il messaggio è coerente con gli studi longitudinali: l'intelligenza crea soprattutto le condizioni perché il reddito sia maggiore, e la condizione più importante è l'istruzione.

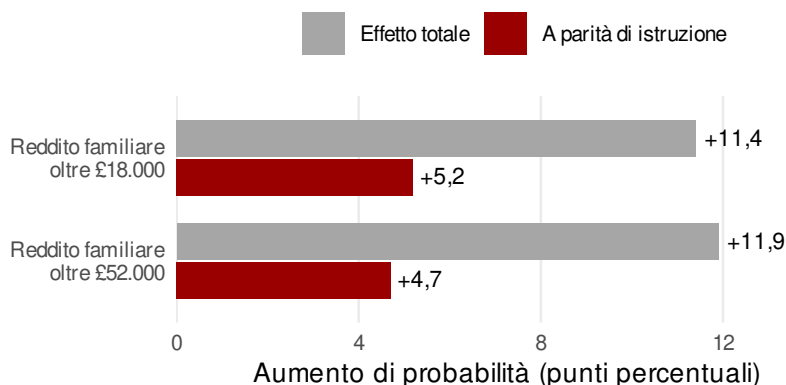


Figura 15.3: Aumento della probabilità di superare due soglie di reddito familiare per una deviazione standard di intelligenza in più: effetto totale ed effetto che resta a parità di livello di istruzione. Stime di randomizzazione mendeliana nella UK Biobank. Elaborazione da Davies et al. (2019).

La forza di questi legami non è fissa. Nelle coorti britanniche più recenti la correlazione fra QI nell'infanzia e durata degli studi si è ridotta: con l'espansione dell'istruzione, proseguire gli studi è diventato meno selettivo («Childhood cognitive ability and adult academic attainment», 2010). In Norvegia, nei maschi nati fra il 1950 e il 1991, la correlazione fra il punteggio al test della leva e gli anni di istruzione a 30 anni è scesa da circa 0,6 a meno di 0,5, e quella con il valore economico medio del titolo raggiunto è calata ancora di più (Van Hootegem et al., 2023). È un cambiamento che

solleva domande scomode: quando un titolo di studio diventa la norma, smette di distinguere, e la selezione si sposta altrove, per esempio sul tipo di scuola o di università frequentata.

15.5 In cima alla distribuzione

Finora si è parlato di correlazioni calcolate su tutta la distribuzione. Ma la relazione vale anche in alto, e qui circola un mito tenace: oltre una certa soglia, diciamo un QI di 120, l'intelligenza non farebbe più differenza, e conterebbero solo impegno, creatività e carattere.

I dati più convincenti sul punto vengono dallo *Study of Mathematically Precocious Youth*, che dagli anni Settanta segue negli Stati Uniti migliaia di ragazzi selezionati a 13 anni perché si trovavano nell'1% più alto in prove di ragionamento. Anche dentro quell'1%, chi stava nel quarto superiore aveva da adulto probabilità sensibilmente maggiori di ottenere un dottorato, pubblicare in ambito scientifico o registrare un brevetto rispetto a chi stava nel quarto inferiore (Lubinski & Benbow, 2006). Fra i ragazzi identificati nell'uno su diecimila più alto, molti sono arrivati a posizioni di grande rilievo nelle università, nelle aziende e nelle professioni (Kell et al., 2013). La soglia oltre la quale «non conta più» non si trova. Si trova, semmai, una differenza di probabilità che continua ad agire anche fra persone già eccezionali.

Questo non trasforma la **plusdotazione** in una diagnosi. Non esiste un confine naturale fra alto potenziale e normalità, così come non esiste fra disabilità intellettiva e funzionamento limite; le soglie usate, spesso un QI di 130, sono convenzioni di rarità. E dire che «ogni bambino è dotato a modo suo», per quanto gentile, svuota la parola del suo significato: se tutti sono rari, nessuno lo è (Warne, 2020).

La domanda pratica è che cosa farne a scuola. Un secolo di ricerca indica che l'accelerazione, cioè anticipare l'ingresso a scuola o saltare un anno o una materia, e alcune forme di raggruppamento per livello hanno effetti positivi sul rendimento degli studenti più dotati (Steenbergen-Hu et al., 2016). Non sono privilegi. Sono il riconoscimento che un programma tarato sulla media è troppo lento per chi sta molto sopra, esattamente come

è troppo veloce per chi sta molto sotto. Resta infine il caso dei bambini con un'abilità generale alta e un disturbo dell'apprendimento: due rarità che si mascherano a vicenda, come si vedrà nel capitolo sulla valutazione nel neurosviluppo.

15.6 Che cosa significa, e che cosa no

L'intelligenza è il miglior singolo predittore del rendimento scolastico che si conosca, e attraverso l'istruzione contribuisce alla posizione sociale. Questo non significa che la scuola sia una formalità o che i risultati siano già scritti. Significa, piuttosto, che le differenze cognitive si traducono in differenze sociali attraverso istituzioni concrete: classi, programmi, selezioni, titoli. Istituzioni diverse possono amplificare o attenuare quella traduzione. E chi usa i risultati scolastici come se fossero un test di intelligenza dimentica proprio queste istituzioni.

In breve

- Il QI prevede il rendimento scolastico in tutte le materie, con correlazioni attorno a 0,5 e fino a 0,8 a livello di fattori latenti.
- Rendimento scolastico e intelligenza non sono la stessa cosa: usare i risultati scolastici come proxy del QI fra contesti diversi porta a conclusioni sbagliate, come nel caso del divario Nord-Sud in Italia.
- Intelligenza e istruzione si influenzano a vicenda e condividono cause comuni, come lo status socioeconomico familiare.
- L'intelligenza prevede status lavorativo e reddito soprattutto attraverso l'istruzione; il legame si indebolisce allontanandosi dalla scuola.
- Anche ai livelli più alti l'abilità continua a fare differenza; per gli studenti più dotati, accelerazione e raggruppamento per livello migliorano il rendimento.

Riferimenti bibliografici

16 Salute e longevità

Un test somministrato a 11 anni prevede la probabilità di essere vivi a 76. Detta così, sembra una provocazione. È invece uno dei risultati più solidi della ricerca sull'intelligenza, e ha dato origine a un campo di studi con un nome proprio: l'**epidemiologia cognitiva**, che studia le associazioni prospettiche fra abilità cognitive ed esiti di salute (Deary et al., 2021).

Prospettiche significa che l'intelligenza viene misurata prima, spesso nell'infanzia, e gli esiti di salute molti anni dopo. Questo esclude una spiegazione ovvia: che siano le malattie dell'età adulta ad abbassare il QI. Non esclude però molte altre spiegazioni, ed è proprio sulla loro natura che conviene ragionare.

16.1 La sorpresa scozzese

Nel 1932 il Consiglio scozzese per la ricerca educativa somministrò un test di intelligenza a quasi tutti i bambini delle scuole scozzesi nati nel 1921. Più di sessant'anni dopo, un gruppo di ricercatori cercò di rintracciare i 2.792 bambini dell'area di Aberdeen per sapere se fossero ancora vivi. Chi era morto prima del 1997 aveva ottenuto a 11 anni un punteggio medio di circa quattro punti inferiore a quello di chi era ancora in vita (Whalley & Deary, 2001). Fra le donne, circa il 70% del quarto con i punteggi più alti era vivo a 76 anni, contro circa il 45% del quarto con i punteggi più bassi. Fra gli uomini la differenza andava nella stessa direzione ma era più contenuta, anche per le morti nella Seconda guerra mondiale, che avevano colpito di più chi aveva punteggi alti.

Lo stesso schema si è ripetuto in molte coorti: coscritti in Svezia, Danimarca, Norvegia e Israele, campioni statunitensi e britannici. Una meta-analisi di 16 studi, con oltre un milione di partecipanti seguiti per periodi fra 17 e

69 anni, ha stimato che un punteggio cognitivo più alto di una deviazione standard si associa a un rischio di morte inferiore del 24% (Calvin et al., 2011). L'associazione era simile per uomini e donne e non mostrava segni evidenti di pubblicazione selettiva.

16.2 Di che cosa si muore meno

Il secondo grande studio scozzese, del 1947, ha permesso di guardare alle cause. Più di 65.000 persone nate nel 1936 e testate a 11 anni sono state seguite per 68 anni. Un QI infantile più alto di una deviazione standard si associava a un rischio minore di morte per quasi tutte le cause principali (Figura 16.1): malattie respiratorie, malattie coronariche, ictus, tumori legati al fumo, malattie dell'apparato digerente, demenza, incidenti (Calvin et al., 2017). Il legame era quasi lineare lungo tutta la distribuzione: non riguardava solo chi aveva punteggi molto bassi.

Un rapporto di rischio di 0,75 significa che, a ogni momento del periodo osservato, chi aveva un QI infantile più alto di quindici punti aveva un rischio di morire per quella causa pari a circa tre quarti di quello di chi stava nella media. Non è una piccola differenza. È però una differenza di probabilità fra gruppi, non una previsione sul singolo: moltissime persone con QI alto muoiono giovani, moltissime con QI basso vivono a lungo.

16.3 Perché? Tre percorsi

L'associazione fra intelligenza e longevità è complessa, piena di mediatori e in parte dovuta a cause comuni. Si possono distinguere tre percorsi principali, non alternativi fra loro (Figura 16.2).

Il percorso socioeconomico. Un QI più alto favorisce più istruzione, che porta a lavori più sicuri, meno logoranti e meglio pagati, che a loro volta danno accesso ad abitazioni più salubri, quartieri meno inquinati e cure migliori. I dati sono istruttivi. Nella meta-analisi citata, tenere conto dello status socioeconomico della famiglia d'origine non modificava quasi l'associazione: le condizioni dell'infanzia non sembrano la causa comune di

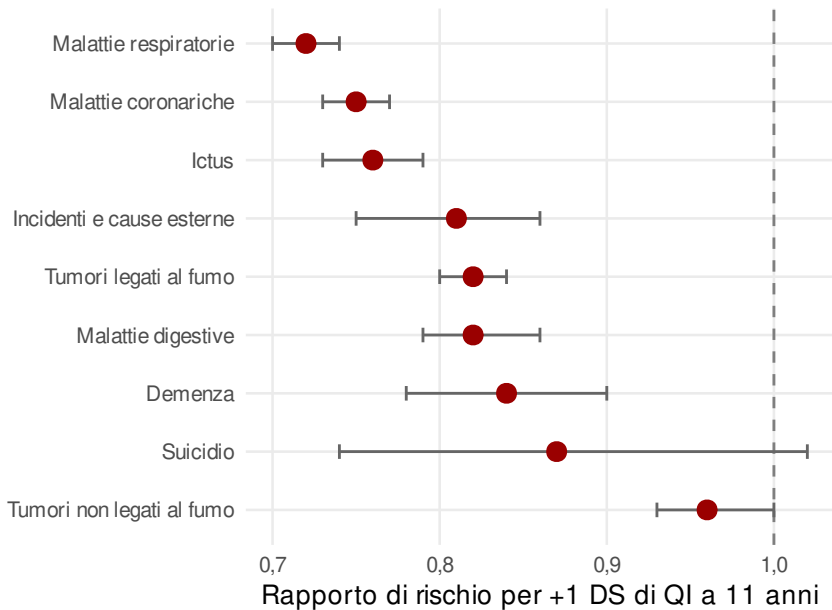


Figura 16.1: Rapporti di rischio di morte associati a un QI a 11 anni più alto di una deviazione standard, per causa, nella coorte scozzese nata nel 1936 (indagine del 1947), seguita fino a 79 anni. Valori inferiori a 1 indicano un rischio minore; le barre orizzontali sono intervalli di confidenza al 95%. Elaborazione da Calvin et al. (2017).

QI e longevità. Tenere conto della posizione sociale da adulti e del livello di istruzione la riduceva invece di circa un terzo e di circa la metà (Calvin et al., 2011). È compatibile con l'idea che istruzione e lavoro siano in parte la strada attraverso cui l'intelligenza si traduce in salute. Ma una riduzione statistica non distingue un mediatore da un fattore confondente, e il resto dell'associazione chiede comunque altre spiegazioni.

Il percorso della *health literacy*. Mantenersi in salute è anche un compito cognitivo. Da un lato richiede di capire, accettare e mettere in pratica procedure mediche e istruzioni sui farmaci. Dall'altro richiede di attenersi a stili di vita salutari, cioè valutare rischi, pianificare, rinunciare a grati-

ficazioni immediate. Le persone con punteggi più alti tendono a fumare meno, a bere meno alcol e meno bevande zuccherate, a fare più attività fisica, ad avere una dieta migliore. Linda Gottfredson ha proposto che l'intelligenza sia una delle cause fondamentali delle disuguaglianze di salute, proprio perché la gestione della propria salute è, di fatto, una lunga serie di problemi da risolvere (Gottfredson, 2004).

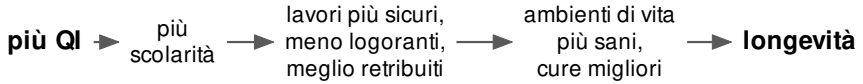
Il percorso dell'integrità del sistema e della genetica. Un'ipotesi diversa è che punteggi cognitivi alti e buona salute siano entrambi segni di un organismo ben costruito e ben funzionante. Un sistema nervoso integro, efficiente e ben sviluppato sarebbe un indicatore di un organismo resistente in generale (Deary, 2012b). Una parte di quella qualità di base ha origine genetica: analisi su gemelli e su dati genetici molecolari suggeriscono che l'associazione fra intelligenza e durata della vita sia in buona parte dovuta a fattori genetici comuni (Arden et al., 2016), un esempio della pleiotropia vista nel capitolo sulla genetica molecolare. Un'altra parte passa per la salute fisica nell'infanzia, che influenza lo sviluppo del sistema nervoso e si prolunga nella salute adulta. In questo percorso l'intelligenza non causa la longevità: QI e longevità sono correlati perché dipendono dalle stesse condizioni di partenza.

16.4 Invecchiamento e demenza

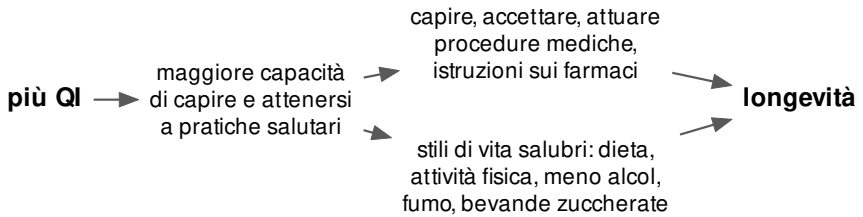
Un capitolo a parte riguarda l'invecchiamento cognitivo. Il QI infantile prevede anche il rischio di morte per demenza, come si è visto nella coorte scozzese. Una parte di questa associazione è probabilmente legata al concetto di **riserva**. A parità di alterazioni cerebrali dovute all'età o alla malattia, alcune persone mantengono un funzionamento migliore di altre. Si distingue fra una riserva cerebrale, legata a caratteristiche strutturali come il volume del cervello, e una riserva cognitiva, legata al modo in cui i compiti vengono svolti e alle strategie disponibili. Studi epidemiologici indicano che istruzione, complessità del lavoro e attività intellettuali nel tempo libero si associano a un rischio di demenza più basso (Y. Stern, 2012).

Il quadro ha una conseguenza pratica per la valutazione. Una persona con un livello cognitivo di partenza elevato può mostrare punteggi ancora nel-

Tramite fattori socioeconomici



Tramite health literacy



Integrità del sistema e genetica

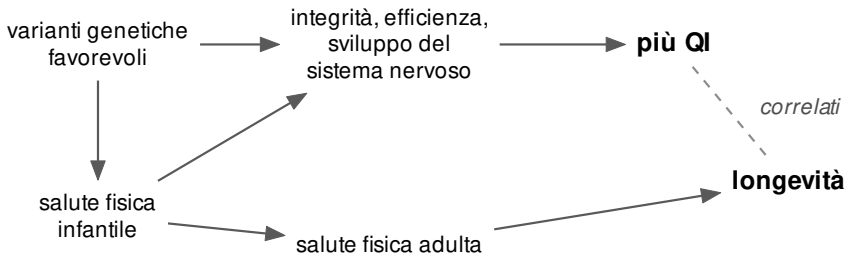


Figura 16.2: Tre percorsi, non alternativi, che possono legare intelligenza e longevità. Nei primi due l'intelligenza contribuisce causalmente alla longevità, attraverso fattori socioeconomici o attraverso la capacità di capire e seguire indicazioni di salute (health literacy). Nel terzo, varianti genetiche e salute fisica infantile influenzano sia lo sviluppo del sistema nervoso, e quindi il QI, sia la salute adulta: QI e longevità sono correlati (linea tratteggiata) senza che l'uno causi l'altra.

la media quando il declino è già in corso; una con un livello di partenza basso può sembrare compromessa senza esserlo. Interpretare un punteggio nell'età anziana richiede quindi di stimare, quando possibile, il livello precedente, per esempio attraverso istruzione, professione o prove di conoscenze cristallizzate, che risentono meno dell'invecchiamento.

16.5 Che cosa se ne può concludere

L'associazione fra intelligenza e salute è robusta, replicata e di dimensioni non trascurabili. Le sue implicazioni pratiche sono però diverse da quelle che si potrebbero immaginare.

Non suggerisce di usare il QI per fare previsioni sanitarie individuali: la previsione sarebbe troppo imprecisa. Suggestisce invece che la complessità delle informazioni e delle procedure sanitarie è un fattore di disuguaglianza. Istruzioni più semplici, percorsi di cura meno faticosi, sistemi che non richiedono di districarsi da soli fra burocrazie e prescrizioni possono ridurre lo svantaggio di chi ha più difficoltà a risolvere problemi, qualunque ne sia la causa. Come per la scuola, il percorso dal punteggio all'esito passa per istituzioni costruite dagli esseri umani, e quindi modificabili.

Resta infine una distinzione da tenere sempre presente. Queste associazioni descrivono differenze fra individui all'interno di una popolazione. Non si possono trasferire, senza altre evidenze, alle differenze medie fra gruppi o fra paesi, nelle quali la salute dipende da molte altre condizioni.

In breve

- Il QI misurato nell'infanzia prevede la longevità: in diverse coorti, una deviazione standard in più si associa a circa un quarto di rischio di morte in meno.
- L'associazione riguarda quasi tutte le principali cause di morte ed è approssimativamente lineare.
- I percorsi ipotizzati passano per i fattori socioeconomici, per la *health literacy* e per cause comuni come l'integrità del sistema e la genetica, e probabilmente agiscono insieme.

- Si tratta di differenze di probabilità fra gruppi, non di previsioni individuali; la complessità delle cure è essa stessa una fonte di disuguaglianza.

Riferimenti bibliografici

17 Intelligenza e lavoro

Il capitolo su scuola e condizioni sociali ha mostrato che l'intelligenza prevede la posizione lavorativa e il reddito, soprattutto attraverso l'istruzione. Questo capitolo cambia prospettiva. Mettiamoci dal lato di chi deve assumere: un QI più alto è una caratteristica da cercare? E se sì, quanto conta rispetto ad altre?

Nella selezione del personale i test di intelligenza raramente si chiamano così. Prove di «problem solving», di «attitudine» o di «conoscenze generali» sono spesso prove di intelligenza fluida o cristallizzata con un altro nome. Anche il titolo di studio richiesto in un annuncio funziona in parte come filtro cognitivo indiretto, che evita di misurare il QI esplicitamente (Warne, 2020). Le batterie complete, invece, sono poco pratiche e in alcuni paesi pongono problemi legali.

Perché usarle? Perché sono economiche: rapide da somministrare, facili da correggere, standardizzate. Perché colgono la capacità di imparare compiti nuovi, più che ciò che si sa già fare. E perché offrono un criterio basato su dati. A condizione, naturalmente, che chi le usa sappia valutarne validità, attendibilità e adeguatezza al contesto.

17.1 Un caso estremo: l'esercito statunitense

L'esercito degli Stati Uniti è un'organizzazione con più di un milione di persone e centinaia di specializzazioni: analisti, informatici, meccanici, autisti, cuochi, medici, interpreti, contabili. La grande maggioranza non ha un ruolo direttamente legato al combattimento, e molti ruoli richiedono di imparare procedure tecniche complesse. Per questo, oltre ai filtri medici, fisici, legali e amministrativi, si valutano capacità cognitive e conoscenze pregresse. Le reclute svolgono una batteria di attitudini, l'*Armed Ser-*

vices Vocational Aptitude Battery (ASVAB), che serve a orientarle fra le specializzazioni. Quattro delle sue prove (conoscenza lessicale, comprensione di brani, ragionamento aritmetico e conoscenze matematiche) formano l'*Armed Forces Qualification Test* (AFQT), il cui punteggio totale, espresso in percentile, colloca ogni candidato in una categoria. Su queste categorie il Dipartimento della Difesa fissa regole precise (Figura 17.1).

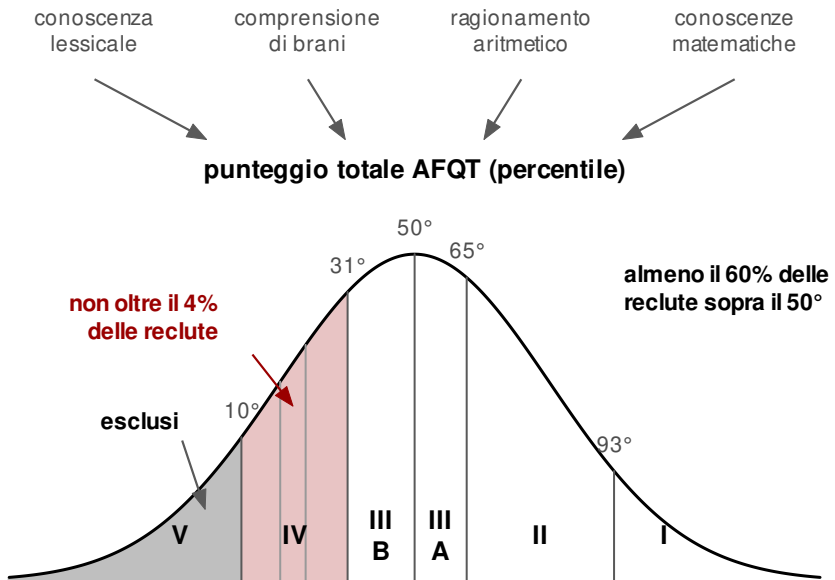


Figura 17.1: Categorie del punteggio totale AFQT, espresso in percentile, e regole di reclutamento del Dipartimento della Difesa statunitense: chi rientra nella categoria V è escluso, le reclute di categoria IV (sottocategorie A, B e C, fra il 10° e il 30° percentile) non devono superare il 4%, e almeno il 60% delle reclute deve trovarsi sopra il 50° percentile.

Fra il 1966 e il 1971, durante la guerra del Vietnam, il segretario alla Difesa Robert McNamara abbassò quelle soglie con un programma chiamato «Project 100,000». Circa 354.000 uomini che in precedenza sarebbero stati esclusi furono arruolati: per lo più reclute di categoria IV, fra il 10° e il 30° percentile, cioè grosso modo con un QI fra 80 e 92, e in parte con punteggi

ancora più bassi, ammessi in deroga. La motivazione ufficiale era offrire disciplina, formazione e opportunità di integrazione a giovani di contesti svantaggiati (Warne, 2020).

Il programma fu un fallimento. Le nuove reclute dovevano ripetere l'addestramento molto più spesso, venivano congedate con maggiore frequenza e, secondo le ricostruzioni disponibili, morivano in combattimento a un tasso sensibilmente più alto. Anche dopo il ritorno alla vita civile, i loro esiti in termini di occupazione, reddito e stabilità familiare risultavano peggiori di quelli di coetanei con capacità simili che non avevano prestato servizio (Gregory, 2015). L'errore non stava nella mancanza di risorse per l'addestramento, ma nell'idea che chiunque possa essere formato per qualunque compito con sufficiente impegno. Il caso è diventato un esempio di come ignorare le differenze cognitive possa danneggiare proprio le persone che si vorrebbe aiutare.

17.2 Le stime classiche

Nella vita civile, il riferimento per decenni sono state le meta-analisi di Frank Schmidt e John Hunter. Secondo le loro stime, l'abilità mentale generale correlava circa 0,6 con il successo nella formazione professionale e circa 0,5 con la prestazione lavorativa: il miglior singolo predittore conosciuto (Schmidt & Hunter, 2004; Schmidt & Hunter, 1998).

Quei valori però non erano correlazioni osservate. Le correlazioni grezze erano spesso meno della metà. Erano state corrette per due fattori.

Il primo è l'**attendibilità**. Un punteggio di QI è la somma di g e di un errore casuale. Se un test rapido è composto per metà dall'uno e per metà dall'altro, la sua correlazione con la produttività lavorativa sarà più bassa di quella di g , perché l'errore non correla con nulla (Figura 17.2). Conoscendo o ipotizzando l'attendibilità della misura, si può prendere la correlazione osservata e «disattenuarla», risalendo a quella sottostante. Lo stesso vale per l'errore di misura nella valutazione della prestazione.

Strumenti → *Attendibilità ed errore di misura sono trattati in Quando una misura è buona; la variabilità campionaria delle correlazioni in Correlazione e differenze individuali.*

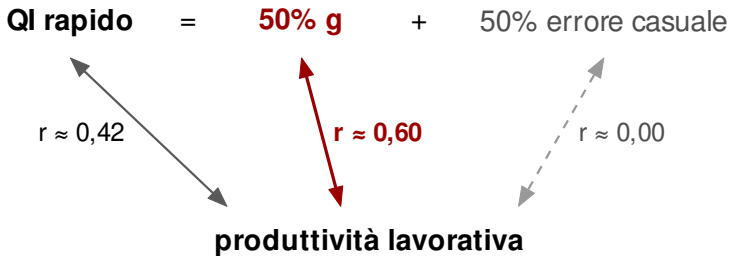


Figura 17.2: Attendibilità e disattenuazione, in un esempio. Un test rapido di QI è fatto per metà di g e per metà di errore casuale. Se g correla 0,60 con la produttività lavorativa e l'errore non correla, la correlazione osservata del test è circa 0,42 ($0,60 \times \sqrt{0,5}$). Conoscendo l'attendibilità, si può risalire dallo 0,42 osservato allo 0,60 sottostante.

Il secondo è la **restrizione del range**. Le correlazioni vengono calcolate su persone che svolgono effettivamente quel lavoro, cioè già selezionate, anche in base alle loro capacità: chi aveva punteggi molto bassi non è stato assunto, chi li aveva molto alti spesso fa un lavoro diverso, e chi rende poco viene allontanato. In un gruppo così omogeneo la correlazione si riduce, come mostra la Figura 17.3.

In linea di principio, entrambe le correzioni sono legittime. Il problema è come sono state applicate. Revisioni recenti hanno mostrato che la correzione per restrizione del range era spesso basata su assunzioni arbitrarie e veniva applicata anche quando la selezione era stata modesta, con un effetto sistematico di sovrastima. Rifacendo i calcoli con correzioni più prudenti, la validità dei test cognitivi per prevedere la prestazione lavorativa scende attorno a 0,3 (Sackett et al., 2022). Resta un predittore utile, ma non più nettamente superiore a tutti gli altri.

La revisione non ha chiuso la questione. Altri specialisti della selezione sostengono che rinunciare alla correzione porti all'errore opposto, cioè a sot-

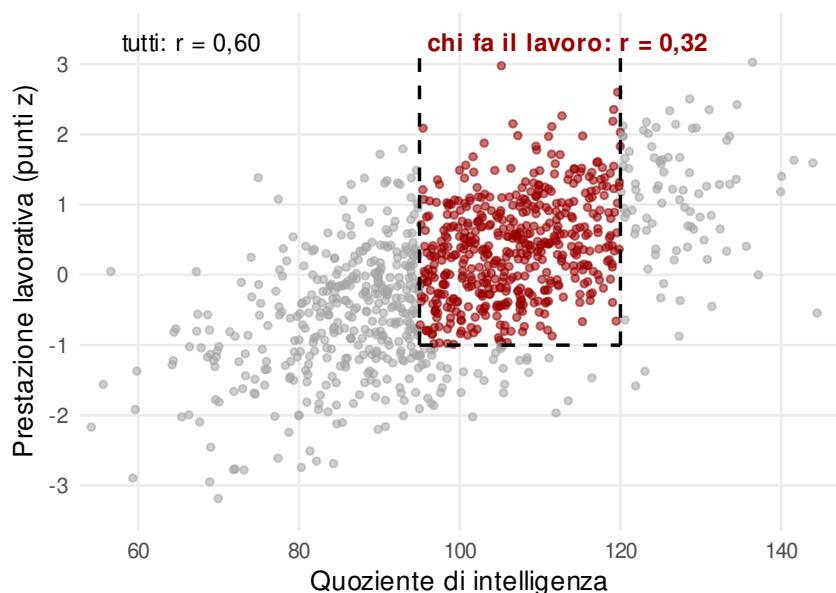


Figura 17.3: Restrizione del range in dati simulati. In tutta la popolazione la correlazione fra QI e prestazione lavorativa è 0,60; fra le sole persone che svolgono effettivamente il lavoro (riquadro: QI fra 95 e 120 e prestazione non troppo bassa) scende a circa 0,3.

tostimare sistematicamente la validità (Oh et al., 2023). Una meta-analisi dei soli studi di questo secolo ha trovato una correlazione osservata media di 0,16 e, dopo correzioni moderate, di 0,22 (Sackett et al., 2024). Il punto di consenso è più modesto, ma solido: l'abilità cognitiva prevede la prestazione lavorativa, con una validità almeno moderata, e lo 0,5 delle sintesi classiche va trattato come un limite superiore piuttosto che come un dato acquisito.

17.3 Dipende dal compito, e dal criterio

Le stime variano molto con il tipo di lavoro e con ciò che si vuole prevedere.

La complessità del lavoro. L'intelligenza prevede la prestazione in modo più forte nei lavori complessi, che richiedono di analizzare informazioni, decidere e coordinare, e in modo più debole in quelli semplici e ripetitivi. Durante la formazione, invece, la sua capacità predittiva è elevata quasi per qualsiasi lavoro: imparare un compito nuovo richiede sempre capacità cognitive (Schmidt & Hunter, 2004). La Figura 17.4 mostra le due tendenze con le stime classiche, che come si è visto vanno lette come limiti superiori.

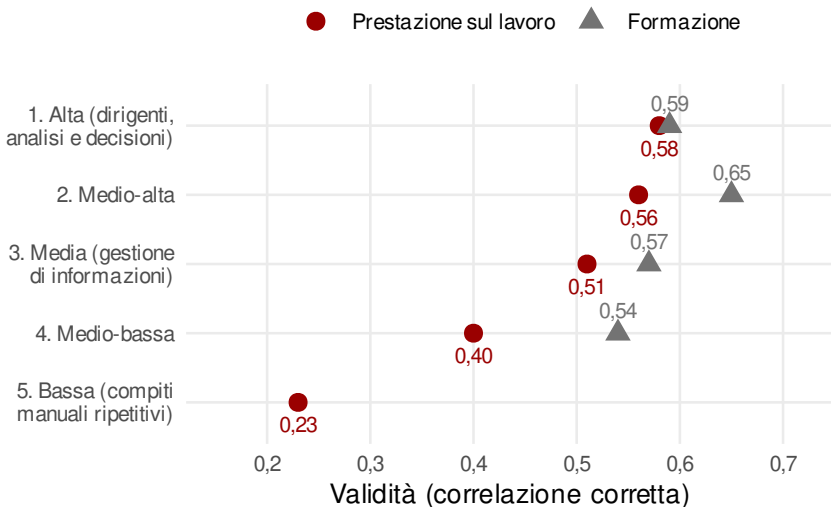


Figura 17.4: Validità dell'abilità cognitiva generale per livello di complessità del lavoro, nella prestazione sul lavoro e nella formazione. Stime classiche corrette per attendibilità e restrizione del range, da considerare limiti superiori; per i lavori di complessità minima manca la stima sulla formazione. Elaborazione da Schmidt e Hunter (2004), tabella 2.

La stratificazione fra professioni. Le professioni differiscono molto nel QI medio di chi le svolge. Uno studio recente su centinaia di professioni ha trovato una differenza superiore a una deviazione standard e mezza fra gli estremi, ma ha mostrato anche che le professioni si differenziano, in misura minore, per tratti di personalità e interessi (Wolfram, 2023). Le gerarchie occupazionali non sono costruite solo sull'intelligenza.

Il criterio di prestazione. Le stime più recenti, riassunte nella Tabella 17.1, mostrano quanto conti ciò che si sceglie di prevedere. La correlazione è più alta con il rendimento nella formazione e con misure oggettive della prestazione nei compiti centrali del ruolo, come le prove pratiche standardizzate usate nei ruoli militari (Cucina et al., 2024). È più bassa con i giudizi complessivi dei superiori, che dipendono anche da relazioni e impressioni (Sackett et al., 2024). Si estende anche oltre i compiti del ruolo: è presente con la collaborazione spontanea verso i colleghi e, in senso negativo e più debole, con assenteismo e comportamenti scorretti (Nye et al., 2022).

Tabella 17.1: Correlazioni fra QI o *g* e diversi indicatori della prestazione lavorativa nelle stime recenti. Dove sono indicati due valori, rappresentano l'intervallo fra le stime di sintesi diverse. Sintesi da Cucina et al. (2024), Nye et al. (2022) e Sackett et al. (2024).

Indicatore	Esempi	Correlazione con QI / <i>g</i>
Prestazione durante la formazione	voto all'esame finale, teorico o pratico, di un corso tecnico	da 0,49 a 0,60
Prestazione oggettiva nei compiti centrali del ruolo	accuratezza o errori nelle procedure tecniche; rapidità e soddisfazione nel servizio clienti; prove pratiche standardizzate	da 0,23 a 0,55
Giudizio complessivo «esterno»	rendimento giudicato dal superiore o da valutatori incaricati	da 0,21 a 0,22

Indicatore	Esempi	Correlazione con QI / g
Comportamenti di cittadinanza organizzativa	aiuto spontaneo ai colleghi, cooperazione, ruoli assunti senza obbligo contrattuale	0,28
Comportamenti di ritiro	assenteismo, ritardi, intenzione di lasciare il posto	-0,11
Comportamenti devianti o contrari all'azienda	furti, violazione dei regolamenti, sabotaggio	-0,21

17.4 Oltre g

Il quadro che emerge dalle ricerche più recenti è meno centrato sul solo fattore generale.

Le **abilità cognitive specifiche** possono aggiungere informazione. Una meta-analisi ha trovato che conoscenze generali, velocità di elaborazione, abilità quantitative e abilità visive prevedono la prestazione in misura paragonabile o talvolta superiore al ragionamento fluido e all'abilità generale, specialmente quando corrispondono alle richieste della mansione (Nye et al., 2022). È un risultato in parziale tensione con la prevalenza di g vista nei capitoli sui modelli, e la discussione è aperta. Nel lavoro, a differenza che in un test, conta anche ciò che si è già imparato.

Il consenso attuale tende quindi a spostare l'attenzione da un singolo predittore a una costellazione di abilità e tratti desiderabili (Tabella 17.2). La personalità, in particolare, aggiunge un contributo indipendente: la coscienziosità, cioè affidabilità, organizzazione e perseveranza, prevede la prestazione in quasi tutti i tipi di lavoro (Barrick & Mount, 1991; Wilmot & Ones, 2019).

Tabella 17.2: Oltre *g*: abilità e tratti che contribuiscono a prevedere la prestazione lavorativa. Sintesi da Nye et al. (2022), Wilmot & Ones (2019), Y. Lee et al. (2019), Sackett et al. (2023) e Pletzer & Abrahams (2025).

Predittore	Che cosa aggiunge
<i>g</i> / QI	Utile in ogni caso come indicatore generale (r da 0,23 a 0,55 secondo il criterio)
Abilità cognitive specifiche	Soprattutto conoscenze generali (intelligenza cristallizzata), velocità e in parte abilità visive aggiungono predittività al QI; conviene puntare alle abilità più coinvolte nella mansione
Coscienziosità	Predittore generale della prestazione in qualsiasi tipo di lavoro ($r \approx 0,30$)
Stabilità emotiva	Utile in aggiunta alla coscienziosità quando conta il lavoro di gruppo
Estroversione, amicalità	Utili nei lavori a contatto con il pubblico o con compiti di guida (r da 0,23 a 0,38)
Onestà, umiltà	Prevedono soprattutto, in negativo, ritiro e comportamenti devianti ($r \approx -0,44$), meno la prestazione oggettiva

Molto popolari sono oggi le cosiddette **soft skills**: pensiero analitico, flessibilità, adattabilità, creatività, curiosità, capacità di apprendere. Il problema è che questa etichetta raccoglie costrutti che si sovrappongono in gran parte a intelligenza e personalità, con la differenza di essere presentati come intrinsecamente allenabili. Sono un esempio del rischio discusso negli approfondimenti finali: dare nomi nuovi a costrutti già noti.

17.5 Una valutazione equilibrata

L'intelligenza è un buon predittore della capacità di imparare un lavoro e di svolgerlo, soprattutto quando il lavoro è complesso. Le stime classiche erano probabilmente gonfiate da correzioni troppo generose, e oggi la si considera un predittore importante fra altri, non l'unico che conta. Per chi seleziona, la conclusione pratica è combinare misure: abilità generale, abilità specifiche rilevanti per la mansione, tratti di personalità. Per chi viene selezionato, la conclusione è un'altra: un test cognitivo può dire qualcosa di vero su quanto facilmente si impara, ma dice poco su quanto si sarà affidabili, collaborativi o motivati.

In breve

- I test cognitivi sono molto usati nella selezione del personale, spesso con altri nomi, perché sono rapidi, standardizzati e prevedono la capacità di apprendere compiti nuovi.
- Il caso del Project 100,000 mostra i costi di ignorare le differenze cognitive nell'assegnazione a ruoli complessi e rischiosi.
- Le stime classiche di validità, attorno a 0,5, erano gonfiate da correzioni eccessive per restrizione del range; stime più prudenti si collocano fra 0,2 e 0,3.
- La capacità predittiva è maggiore nei lavori complessi e durante la formazione; abilità specifiche e coscienziosità aggiungono informazione utile.

Riferimenti bibliografici

Parte VII

Intelligenza e condizioni del neurosviluppo

18 Dalle categorie alle dimensioni

Una diagnosi sembra una risposta: c'è o non c'è. Nel neurosviluppo, però, la risposta è quasi sempre una questione di grado. Essere molto distanti dalla media in qualche aspetto del funzionamento cognitivo, emotivo, comportamentale o sociale può creare difficoltà di adattamento al contesto in cui si vive. Ma quella distanza raramente si lascia chiudere negli inquadramenti netti che la medicina usa per le malattie.

Questo capitolo e il prossimo riguardano le **condizioni del neurosviluppo**: disabilità intellettiva, autismo, ADHD, disturbi specifici dell'apprendimento. Il loro legame con l'intelligenza è stretto, per due ragioni. La prima è pratica: la valutazione dell'intelligenza fa parte di quasi ogni percorso diagnostico. La seconda è teorica, ed è il tema di questo capitolo: le condizioni del neurosviluppo hanno la stessa struttura delle abilità cognitive, fatta di dimensioni continue e correlate. Capirlo cambia il modo di pensare la diagnosi.

Le definizioni diagnostiche riportate qui servono a orientarsi e vanno sempre verificate nelle versioni aggiornate dei manuali. Non sono una guida clinica.

18.1 Un caso illustrativo

Una bambina di nove anni viene valutata per difficoltà scolastiche marcate. La lettura è lenta e poco accurata, l'ortografia fragile, la matematica fonte di ansia. Genitori e insegnanti riferiscono distrazione frequente. Per il resto ragiona, ricorda e si relaziona come i coetanei. Il caso è costruito a scopo illustrativo, ma è molto tipico.

La valutazione cognitiva restituisce un QI totale di 84. Il dato, letto da solo, indicherebbe un funzionamento generale sotto la media. Ma gli indici

raccontano altro: comprensione verbale e ragionamento sono nella media, memoria di lavoro e velocità di elaborazione molto sotto (Figura 18.1). Nei subtest la caduta è ancora più netta: la memoria di cifre è quasi al minimo della scala, mentre il ragionamento con matrici è sopra la media. L'indice di abilità generale, che esclude memoria di lavoro e velocità, è 100. Le prove di lettura e scrittura sono molto deboli, quelle di matematica deboli soprattutto nel calcolo scritto. Il questionario sull'ansia per la matematica dà valori elevati. I questionari sull'attenzione compilati da genitori e insegnanti indicano difficoltà rilevanti, non confermate in modo netto dalle prove attentive.

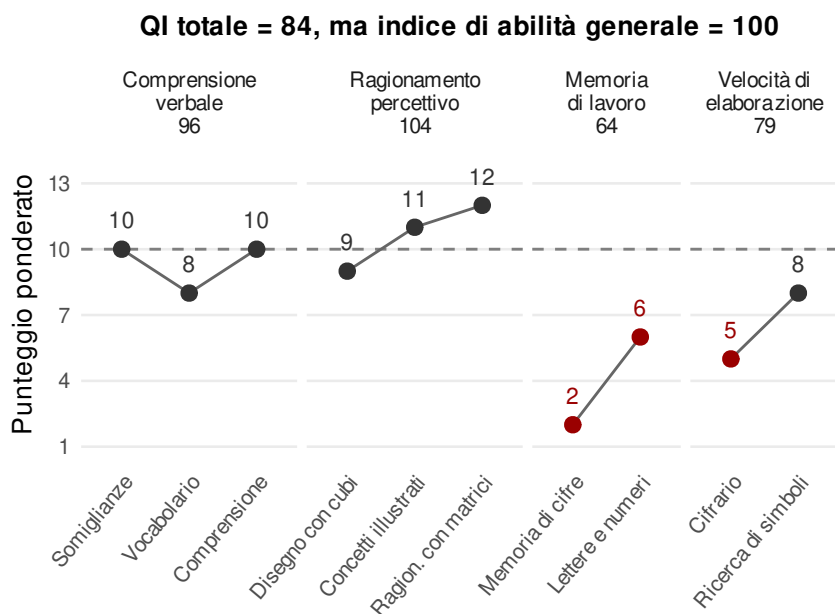


Figura 18.1: Il profilo del caso illustrativo: punteggi ponderati nei dieci subtest principali di una scala Wechsler (media 10, deviazione standard 3), raggruppati nei quattro indici, con il punteggio di ciascun indice in scala QI. In rosso i subtest più di una deviazione standard sotto la media: tutti in memoria di lavoro e velocità di elaborazione. Il QI totale (84) risente di queste cadute; l'indice di abilità generale (100), che le esclude, no. Dati costruiti a scopo illustrativo.

La conclusione sarà probabilmente una diagnosi di disturbo misto delle abilità scolastiche, con una segnalazione di difficoltà attentive. Ma il caso solleva domande che nessuna etichetta risolve. Dislessia e discalculia sono davvero condizioni distinte, se si presentano quasi sempre insieme? Le difficoltà in matematica sono cognitive o emotive? L'ADHD c'è o non c'è, se i sintomi riferiti sono chiari e le prove meno? Il quadro è misto, spurio, ed è del tutto reale.

18.2 Le categorie dei manuali

La pratica clinica inquadra queste difficoltà in categorie prefissate, quasi si trattasse di sindromi mediche. I manuali di riferimento, il DSM-5-TR e l'ICD-11, definiscono i disturbi del neurosviluppo come condizioni che esordiscono nel periodo dello sviluppo e comportano difficoltà significative nell'acquisizione o nell'esecuzione di funzioni intellettive, motorie, linguistiche o sociali (Association, 2022; World Health Organization, 2019). La classificazione si basa sul quadro osservabile, non sulle cause.

Le categorie principali sono queste:

- **disabilità intellettiva:** limitazioni nel funzionamento intellettivo generale e nel comportamento adattivo nella vita quotidiana;
- **disturbo dello spettro dell'autismo:** difficoltà persistenti nella comunicazione e interazione sociale, insieme a comportamenti, interessi e attività ristretti e ripetitivi;
- **ADHD:** disattenzione e/o iperattività-impulsività persistenti, presenti in più contesti e tali da compromettere il funzionamento;
- **disturbi specifici dell'apprendimento:** difficoltà persistenti in lettura, scrittura o calcolo, non spiegate da disabilità intellettiva, deficit sensoriali o scolarizzazione insufficiente. In Italia la legge 170 del 2010 li riconosce come dislessia, disgrafia, disortografia e discalculia;
- disturbi della comunicazione e del movimento.

Nella disabilità intellettiva si distingue da sempre fra forme **sindromiche**, con cause genetiche identificate e segni fisici riconoscibili, come la sindrome di Down o dell'X fragile, e forme **non sindromiche**, senza una causa

unica individuabile. Le seconde rappresentano in buona parte la coda inferiore della distribuzione normale delle abilità, prodotta dagli stessi innumerevoli fattori che generano le differenze nella popolazione generale. Questa distinzione è la chiave per capire tutte le altre condizioni.

18.3 L'illusione del deficit centrale

Per decenni la ricerca ha cercato per ogni disturbo un **deficit centrale**: un modulo cognitivo difettoso in uno sviluppo altrimenti normale. Un disturbo, un deficit. L'intelligenza contava solo come inquadramento generale o come criterio di esclusione. Oggi questo approccio è sostanzialmente abbandonato (Astle & Fletcher-Watson, 2020). I candidati non sono mancati.

Per l'**autismo** si sono proposti un deficit nella capacità di rappresentarsi gli stati mentali altrui, la cosiddetta teoria della mente; un malfunzionamento del sistema dei neuroni specchio; una tendenza a elaborare i dettagli perdendo la visione d'insieme; un orientamento estremo verso la sistematizzazione a scapito dell'empatia. Le differenze medie osservate sono reali, e nei compiti classici di teoria della mente possono essere ampie; ma variano molto da persona a persona e non bastano a definire la condizione. Le prime evidenze sui neuroni specchio non sono state confermate (Hamilton, 2013). Una meta-analisi non ha trovato né una superiorità nell'elaborazione locale né un deficit in quella globale (Van der Hallen et al., 2015). Anche l'idea di un deficit sociale tutto interno alla persona autistica è stata messa in discussione. Secondo l'ipotesi della **doppia empatia**, le difficoltà nascono soprattutto nell'incontro fra modi diversi di comunicare (Milton, 2012). Un esperimento ha trovato che catene di persone autistiche si trasmettevano un'informazione con la stessa efficacia di catene di persone non autistiche, mentre le catene miste funzionavano peggio (Crompton et al., 2020). È un risultato interessante, non ancora una teoria consolidata.

Per l'**ADHD** si sono proposti deficit di inibizione, di funzioni esecutive, di tolleranza dell'attesa. Le differenze medie esistono, di solito attorno a mezza deviazione standard, ma nessuna distingue in modo affidabile chi ha la diagnosi da chi non ce l'ha. E differenze di entità simile si trovano

anche in compiti non specifici: memoria a breve termine, velocità, perfino ragionamento astratto (Sj wall et al., 2013).

Per la **dislessia** il candidato pi  forte   il deficit fonologico, insieme alla lentezza nella denominazione rapida. Qui le differenze medie sono grandi, oltre una deviazione standard. Ma molti casi di dislessia non presentano quei deficit, e nessun marcatore singolo   necessario o sufficiente (Penington, 2006). Per la **discalculia** il presunto deficit del senso del numero mostra differenze medie modeste, mentre molti processi generali, come memoria di lavoro, velocit  e ragionamento, risultano ugualmente o pi  compromessi.

Il bilancio   netto. Nessun marcatore genetico, di neuroimmagine, neurofisiologico, cognitivo o neuropsicologico ha oggi una capacit  diagnostica adeguata per nessuna delle condizioni del neurosviluppo (Astle et al., 2022).

18.4 Soglie arbitrarie

C'  una ragione profonda per cui i deficit centrali non si trovano.   la stessa per cui le abilit  cognitive hanno una distribuzione continua. Ogni comportamento umano   il risultato di moltissimi fattori genetici, neurobiologici, ambientali ed esperienziali. Che si sommino o, in parte, si moltiplichino, il risultato di molti fattori   comunque una distribuzione continua. E in una distribuzione continua non esiste una soglia naturale.

La lettura ne   l'esempio classico. Gi  all'inizio degli anni Novanta uno studio longitudinale mostr  che l'abilit  di lettura nella popolazione si distribuisce lungo un unico continuo, senza un punto in cui finisca la distribuzione dei lettori tipici e cominci quella dei dislessici (Shaywitz et al., 1992). La dislessia, in questo senso, non   una categoria naturale:   il nome che diamo alla parte bassa della distribuzione di un'abilit  che la scuola richiede. Anche per l'ADHD, uno studio su gemelli ha trovato che l'ereditabilit  dei sintomi   la stessa lungo tutto il continuo, dai livelli pi  alti fino alla piena normalit : non esiste un nucleo genetico distinto del disturbo (Levy et al., 1997).

Ne segue che la prevalenza di questi disturbi dipende dalla soglia scelta (Figura 18.2). Le soglie riflettono convenzioni di rarità, come il 5° percentile o due deviazioni standard sotto la media, e considerazioni pragmatiche sul bisogno di supporto. Con la prima convenzione la «prevalenza» è il 5%, con la seconda circa il 2%. Sono decisioni ragionevoli, non scoperte. E quando le difficoltà sono molte, ciascuna appena sotto soglia, nessuna convenzione aiuta: la decisione resta difficile.

Strumenti → Percentili, deviazioni standard e rarità statistica, e perché un punteggio vicino al tetto o al pavimento di un test è poco informativo: *Distribuzioni, punti z e punteggi standardizzati.*

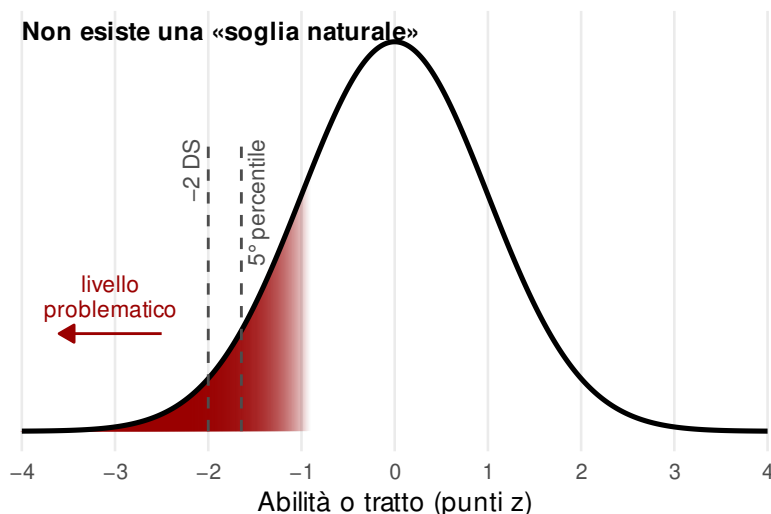


Figura 18.2: Non esiste una soglia naturale. L'abilità o il tratto, misurati con strumenti standardizzati, hanno una distribuzione continua: verso sinistra il livello diventa via via più problematico (sfumatura), senza alcun punto di discontinuità. Le linee tratteggiate indicano due soglie convenzionali di rarità: -2 deviazioni standard (circa il 2% della popolazione) e il 5° percentile ($z \approx -1,64$).

Questo porta a un rovesciamento della causalità che conviene rendere esplicito. Non si legge con fatica a causa della dislessia: si riceve la diagnosi di dislessia perché si legge con fatica, per ragioni non altrimenti

spiegate. Non si è iperattivi a causa dell'ADHD: si riceve l'etichetta di ADHD perché si è molto iperattivi. L'etichetta descrive un comportamento osservato, non ne individua la causa. E poiché il comportamento è a valle di innumerevoli concause, non si troveranno mai poche variabili fortemente associate, ma molte variabili associate debolmente.

18.5 Comorbidità ovunque

Il secondo problema per l'approccio categoriale è la **comorbidità**. Le condizioni del neurosviluppo si presentano insieme molto più spesso di quanto ci si aspetterebbe se fossero indipendenti, e un approccio fatto di categorie e deficit centrali fatica a spiegarlo (Figura 18.3). Fra le persone autistiche, circa una su quattro ha anche una disabilità intellettiva (Mutluer et al., 2022); prima del 2000 le stime superavano il 50%, quando la diagnosi di autismo era riservata ai quadri più gravi. Circa il 40% ha anche un ADHD (Rong et al., 2021). Anche nelle forme sindromiche di disabilità intellettiva l'ADHD è frequente: in un campione di bambini con sindrome di Down riguardava quasi il 44% (Ekstein et al., 2011). Un bambino con un disturbo del linguaggio in età prescolare ha un rischio elevato di ricevere poi una diagnosi di disturbo dell'apprendimento (Snowling et al., 2019). Nelle casistiche cliniche dei disturbi dell'apprendimento, i quadri misti sono quasi la metà (Toffalini et al., 2017). Anche nella popolazione generale, dove la maggior parte dei bambini con ADHD, dislessia o discalculia ha una sola diagnosi, chi ne ha una ha da due a tre volte più probabilità di averne una seconda (van Bergen et al., 2025).

Sotto le etichette troviamo quindi tratti **continui**, con soglie arbitrarie, e **correlati**, con comorbidità estese. È esattamente la struttura delle abilità cognitive: distribuzioni gaussiane e positive manifold. E, come per le abilità, la spiegazione più semplice della correlazione è l'esistenza di cause comuni.

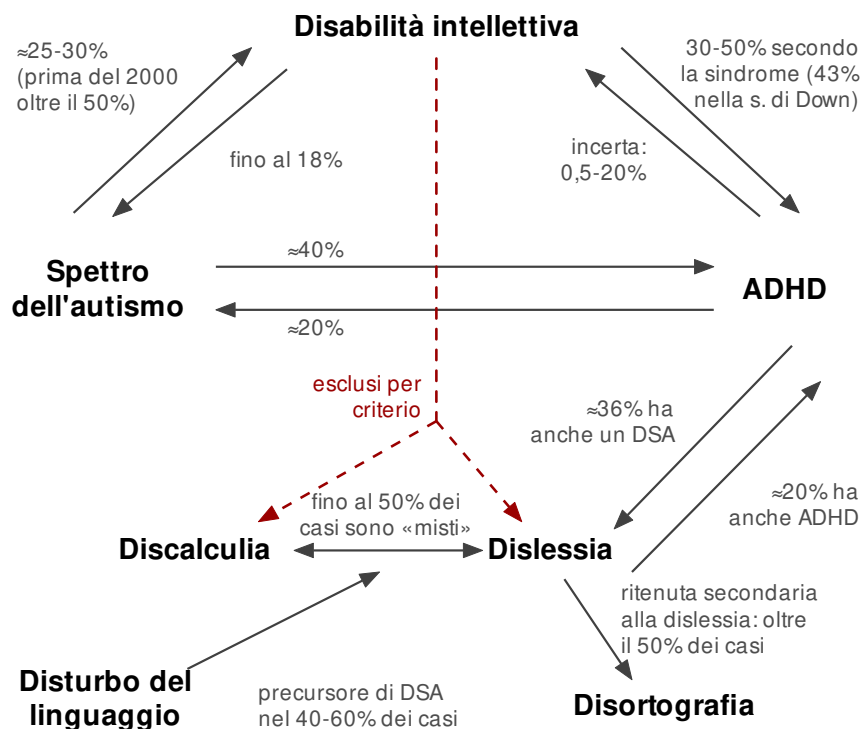


Figura 18.3: La comorbidità fra condizioni del neurosviluppo. Ogni freccia va da una condizione a un'altra e riporta la quota approssimativa di casi della prima che presentano anche la seconda; le stime variano molto con il campione e i criteri. Disabilità intellettiva e disturbi dell'apprendimento si escludono per definizione (in rosso). Adattamento da una sintesi dell'autore, basata su Ekstein et al. (2011), Mu-tluer et al. (2022), Rong et al. (2021), Snowling et al. (2019), Toffalini et al. (2017), van Bergen et al. (2025) e altri studi.

18.6 Verso modelli dimensionali

Nella psicopatologia dell'adulto questa consapevolezza ha portato a proposte alternative ai manuali categoriali. La più sviluppata è la **tassonomia gerarchica della psicopatologia** (HiTOP), che organizza sintomi e tratti in dimensioni a diversi livelli di generalità, fino a un possibile fattore generale p (Caspi et al., 2014; Kotov et al., 2017). Un modello del genere accoglie come naturali la comorbidità, la variabilità all'interno di uno stesso disturbo e i casi sottosoglia. Invece di chiedersi quale disturbo abbia una persona, si chiede quale sia il suo profilo su più dimensioni.

È stato proposto di includere le condizioni del neurosviluppo in questi modelli come un'ulteriore dimensione, in continuità con la psicopatologia adulta e in relazione con la struttura dell'intelligenza (Astle et al., 2022). I legami esistono. Nello studio che ha proposto il fattore p , punteggi più alti in p si associavano a un funzionamento cognitivo un po' più basso già nell'infanzia e a un declino maggiore in età adulta (Caspi et al., 2014). La direzione causale resta da chiarire: un livello cognitivo più alto potrebbe proteggere, la psicopatologia potrebbe danneggiare, ed entrambe le cose possono essere vere.

Il problema non riguarda solo la clinica. Anche la ricerca pensa ancora per categorie, e le categorie creano un dilemma: i campioni diagnostici «puri» non sono rappresentativi, e i campioni rappresentativi non sono puri. Profili cognitivi simili si trovano in condizioni diverse, e profili diversi nella stessa condizione. Studiare i profili e le dimensioni è più informativo che studiare le etichette.

18.7 Neurodivergenza e adattamento

Resta una domanda. Se tratti come quelli autistici o legati all'ADHD sono così diffusi ed ereditabili, perché l'evoluzione non li ha eliminati? Una risposta possibile è che non siano solo disfunzionali. La selezione naturale non produce fenotipi perfetti, ma compromessi. Alta attività, ricerca di novità, decisioni rapide ed esplorazione sono poco adatte alle aule scolastiche

moderne, ma potrebbero essere state vantaggiose in altri ambienti. Un'attenzione sistematica alle regolarità e ai dettagli può essere una risorsa in molti contesti. Sono ipotesi, non conclusioni; ma ricordano che la stessa caratteristica può essere un problema in un ambiente e una risorsa in un altro.

Per questo si è diffuso il termine **neurodivergenza**, che indica una differenza rilevante rispetto al funzionamento tipico senza presupporre che sia un deficit. Non sostituisce la parola disturbo: come si vedrà nel prossimo capitolo, le due descrivono livelli diversi dello stesso fenomeno.

In breve

- Le condizioni del neurosviluppo sono classificate dai manuali come categorie, ma i loro confini con la normalità e fra di loro non sono netti.
- La ricerca di un deficit cognitivo centrale per ciascun disturbo è fallita: nessun marcatore ha una capacità diagnostica adeguata.
- I tratti sottostanti sono continui, con soglie diagnostiche convenzionali, e correlati, con comorbidità estese: la stessa struttura delle abilità cognitive.
- I modelli dimensionali e gerarchici, come HiTOP, descrivono questa struttura meglio delle categorie e collegano neurosviluppo, psicopatologia e intelligenza.

Riferimenti bibliografici

19 Valutare l'intelligenza nel neurosviluppo

La valutazione dell'intelligenza è parte integrante, e spesso obbligatoria, di quasi ogni percorso diagnostico nel neurosviluppo. Serve a escludere o confermare una disabilità intellettiva, a descrivere punti di forza e di debolezza, a orientare gli interventi. Ma una batteria di intelligenza è stata costruita e tarata su una popolazione generale. Quando la si usa con persone che hanno un funzionamento diverso dalla media proprio negli aspetti che il test misura, sorgono problemi specifici.

Oggi non esiste una cornice teorica unica per queste condizioni, se non il termine ombrello di **neurodiversità** (Figura 19.1). Sotto l'ombrello stanno due concetti diversi. La *neurodivergenza*, introdotta alla fine del capitolo precedente, è una differenza rilevante del funzionamento neurologico e cognitivo rispetto a quello tipico. Il *disturbo* è un esito clinicamente significativo: disagio rilevante o compromissione del funzionamento in aree importanti della vita. Può associarsi a condizioni sindromiche, che si prestano a categorie nette, o, più spesso, a condizioni non sindromiche, dimensionali e dipendenti dal contesto, in cui la compromissione generale pesa di solito più di quella specifica. Una neurodivergenza può tradursi in disturbo in misura molto variabile; quando è eccezionale, può essere anche una risorsa: talento, creatività, capacità di concentrazione intensa, plusdotazione. Che risorsa e disturbo possano convivere nella stessa persona è una possibilità reale, ancora poco studiata. In tutti questi casi la valutazione dell'intelligenza, articolata in più componenti, è parte integrante della valutazione clinica e un tema centrale della ricerca.

Questo capitolo esamina come si valuta l'intelligenza nelle principali condizioni, quali profili medi emergono dalla ricerca e perché quei profili vanno interpretati con cautela. Un avvertimento preliminare: i profili medi descrivono gruppi. Dedurre una diagnosi dal profilo di un singolo individuo è un errore.

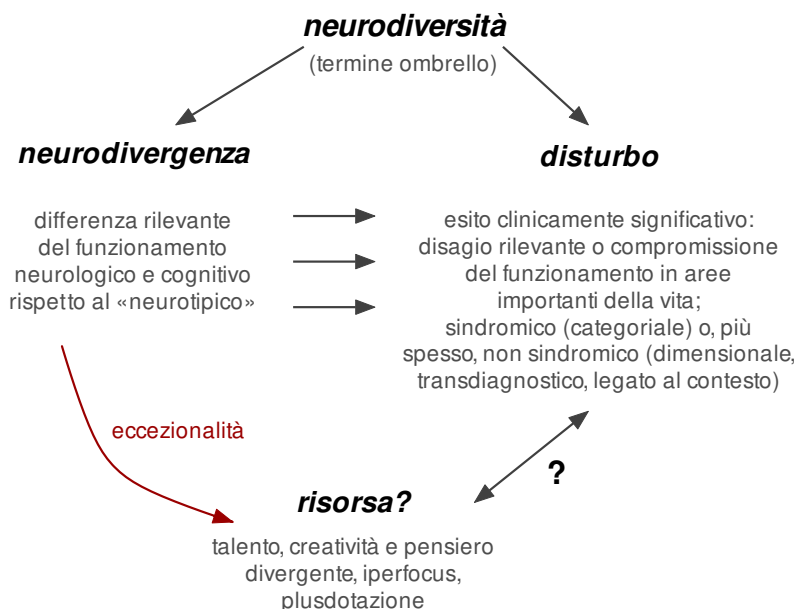


Figura 19.1: Neurodiversità, neurodivergenza e disturbo. La neurodivergenza può tradursi, in misura variabile, in un disturbo; quando è eccezionale può diventare una risorsa; il rapporto fra risorsa e disturbo resta una domanda aperta.

19.1 Disabilità intellettiva

La disabilità intellettiva richiede tre elementi: limitazioni nel funzionamento intellettivo, confermate dalla valutazione clinica e da test standardizzati individuali; limitazioni nel **comportamento adattivo**, cioè nella capacità di soddisfare le richieste di autonomia e responsabilità sociale attese per l'età e il contesto; esordio durante lo sviluppo (Association, 2022).

Tradizionalmente la gravità veniva stabilita in base al QI: lieve sotto 70, moderata sotto 50–55, grave sotto 35–40, profonda sotto 20–25. Gli approcci attuali la determinano soprattutto in base al funzionamento adattivo e al bisogno di supporto, valutati in tre aree: concettuale, che comprende

linguaggio, apprendimenti e memoria; sociale, che comprende comunicazione, relazioni, rispetto delle regole e vulnerabilità alla manipolazione; pratica, che comprende cura di sé, gestione della casa e del denaro, lavoro e vita autonoma. Il comportamento adattivo si misura con scale standardizzate, di solito compilate attraverso un'intervista a chi conosce bene la persona, ed espresse sulla stessa metrica del QI. Il QI resta un criterio di fatto necessario ma non sufficiente.

Le forme **sindromiche** tendono a mostrare profili cognitivi caratteristici, pur con grande variabilità individuale. La sindrome di Down, la causa genetica più comune di disabilità intellettiva, e la sindrome dell'X fragile, la più comune fra quelle ereditarie, offrono un contrasto istruttivo (Tabella 19.1): in linguaggio e abilità visuospatiali i loro profili sono quasi speculari (Grieco et al., 2015; Quintin et al., 2016). Nella sindrome di Williams il quadro somiglia a quello dell'X fragile, con abilità verbali relativamente preservate e difficoltà visuospatiali gravi (Martens et al., 2008). In queste sindromi si osserva inoltre una tendenza del QI a diminuire con l'età: le abilità crescono, ma più lentamente della norma, e la distanza dai coetanei si allarga.

Tabella 19.1: Profili cognitivi tendenziali in due forme sindromiche di disabilità intellettiva: punti di forza (+) e di debolezza (–) rispetto al livello generale della persona.

Area	Sindrome di Down	Sindrome dell'X fragile
Linguaggio	– difficoltà più gravi, soprattutto nell'espressione	+ punto di forza relativo nelle abilità verbali
Memoria a breve termine	– debolezza marcata, soprattutto verbale	– debolezza generale, insieme alle funzioni esecutive
Memoria a lungo termine		+ discreta
Abilità visuospatiali	+ meno compromesse	– difficoltà più gravi

Anche questi profili «tipici» sono tendenze di gruppo, da verificare sui casi concreti. Le forme **non sindromiche** sono estremamente eterogenee, sia nel fenotipo sia nelle cause genetiche, e non hanno profili caratteristici; il QI tende anche meno a diminuire con l'età. Ogni caso fa storia a sé, e il profilo va valutato sempre.

19.2 Autismo: un paradosso

Fra le persone con una diagnosi di autismo, una quota consistente, fra un quarto nelle sintesi internazionali e oltre un terzo nelle stime epidemiologiche statunitensi, ha anche una disabilità intellettiva (Maenner et al., 2023; Mutluer et al., 2022). Eppure le varianti genetiche comuni associate all'autismo sono, in media, associate a un'intelligenza leggermente **più alta**: la correlazione genetica fra i due tratti è debolmente positiva (Savage et al., 2018). Come si conciliano i due dati?

Una possibile spiegazione è un'eziologia mista. Una parte dei casi, spesso con disabilità intellettiva, sarebbe legata a varianti genetiche rare o mutazioni nuove di grande effetto, simili per meccanismo alle forme sindromiche (Satterstrom et al., 2020). Il resto rifletterebbe la coda di un tratto poligenico diffuso, in parte associato a varianti favorevoli alle abilità cognitive.

Un'altra ipotesi parla di **intelligenza sbilanciata** (Crespi, 2016): uno sviluppo disarmonico, con alcune abilità forti, talvolta eccezionali, e un funzionamento quotidiano deludente, soprattutto rispetto alle richieste sociali. Le abilità visuospaziali, il ragionamento fluido e alcune funzioni sensoriali risultano spesso buone; le conoscenze verbali e l'uso del linguaggio più deboli; l'attenzione molto focalizzata ma poco flessibile.

Nelle scale Wechsler le cadute più evidenti riguardano memoria di lavoro e velocità di elaborazione, che possono scendere di circa una deviazione standard. È però anche il profilo più comune di tutta la neurodivergenza, quindi poco specifico.

C'è infine un problema di campionamento. Una meta-analisi degli studi di ricerca sull'autismo ha stimato che circa il 94% dei partecipanti non aveva

una disabilità intellettiva (Russell et al., 2019). Le casistiche cliniche epidemiologiche ne contano molti di più. La ricerca descrive soprattutto la parte ad alto funzionamento dello spettro, oppure le casistiche cliniche ne descrivono soprattutto la parte a basso funzionamento? Con un tratto dimensionale e confini diagnostici in movimento, la domanda potrebbe non avere una risposta semplice.

19.3 ADHD

Per l'ADHD la situazione genetica è opposta a quella dell'autismo: la correlazione genetica con l'intelligenza è moderatamente **negativa** (Demontis et al., 2023). A livello dei punteggi, però, la differenza è meno marcata. La comorbidità con la disabilità intellettiva è contenuta. Una meta-analisi ha stimato per i gruppi con ADHD un QI totale medio inferiore di circa 0,6 deviazioni standard, quasi dieci punti (Frazier et al., 2004). Una parte della differenza dipende però dalle prove che chiedono proprio attenzione sostenuta e memoria di lavoro, cioè dalle stesse difficoltà che definiscono il disturbo.

Il profilo tipico alle scale Wechsler mostra una compromissione marcata della memoria di lavoro, di circa una deviazione standard, e una compromissione parziale della velocità di elaborazione. L'osservazione qualitativa dei protocolli rivela spesso una buona velocità accompagnata da molti errori. Se però si considerano tutti i tipi di compito cognitivo studiati nelle meta-analisi, il profilo medio appare più piatto di quanto suggerirebbe l'idea di un deficit esecutivo specifico.

19.4 Disturbi specifici dell'apprendimento

In teoria i disturbi specifici dell'apprendimento rappresentano il livello più circoscritto della compromissione cognitiva. Il **modello a cono** proposto da Cornoldi e Vecchi lo rende visibile (Figura 19.2): in cima stanno le compromissioni di alto livello, come la disabilità intellettiva, con effetti a cascata su qualsiasi compito; nel mezzo quelle di livello intermedio, come auti-

smo e ADHD, che colpiscono diversi domini; alla base quelle che riguardano processi molto specifici, come lettura, scrittura e calcolo (Cornoldi & Vecchi, 2003).

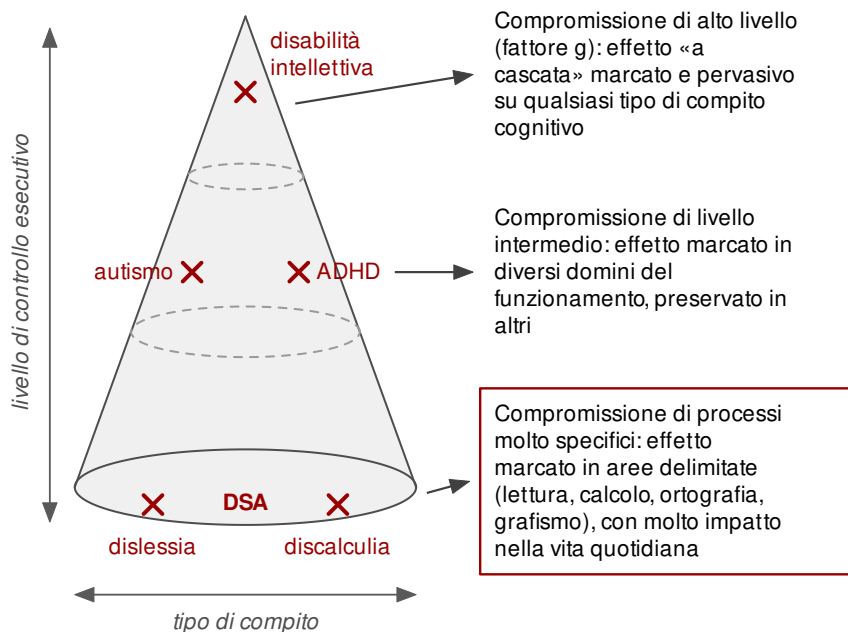


Figura 19.2: Il modello a cono delle difficoltà cognitive. In verticale il livello di controllo esecutivo richiesto dal compito, in orizzontale il tipo di compito: una compromissione di alto livello, come nella disabilità intellettiva, si propaga a tutto il cono; le compromissioni di livello intermedio colpiscono diversi domini; quelle di base, come nei disturbi specifici dell'apprendimento (DSA), riguardano processi molto specifici. Elaborazione da Cornoldi e Vecchi (2003).

La specificità, però, è imposta per definizione. La disabilità intellettiva è un criterio di esclusione esplicito; il funzionamento intellettivo limite, con QI fra 70 e 85, è un elemento di attenzione, e spesso si richiede che almeno alcune aree dell'intelligenza siano nella norma. In passato si richiedeva una discrepanza fra QI e abilità scolastica. Ma le abilità scolastiche sono correlate con il fattore generale sia a livello fenotipico sia genetico. Un disturbo

del calcolo è «specifico» non perché il calcolo sia indipendente dall'intelligenza, ma perché lo si definisce escludendo chi ha un'intelligenza bassa (Figura 19.3).

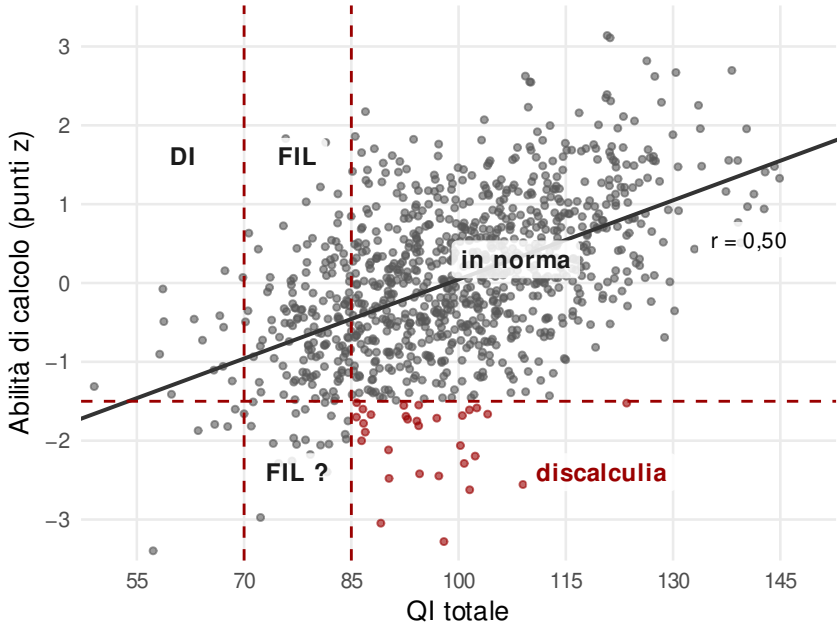


Figura 19.3: Fissiamo delle soglie su una popolazione «normale» e attribuiamo delle etichette. Dati simulati in cui l'abilità di calcolo correla 0,50 con il QI totale. Le soglie (linee tratteggiate) sono QI 70 per la disabilità intellettiva (DI), QI 85 per il funzionamento intellettivo limite (FIL) e -1,5 punti z nel calcolo. Solo chi ha un calcolo molto debole e un QI di almeno 85 riceve l'etichetta di discalculia (in rosso); con un QI fra 70 e 85 il caso resta dubbio.

L'incoerenza dei criteri di inclusione ed esclusione fra gli studi, inevitabile perché i criteri stessi sono arbitrari, produce risultati incoerenti. Studiare dimensioni continue in modo multivariato, senza imporre confini che in natura non esistono, è più informativo.

Nei dati clinici italiani, i profili medi dei bambini con diagnosi di disturbo dell'apprendimento mostrano un andamento ricorrente: comprensione

verbale e ragionamento percettivo nella norma o vicini, memoria di lavoro e velocità di elaborazione più bassi. La forma è simile nella dislessia, nella disortografia, nella discalculia e nei quadri misti; cambiano piuttosto l'entità delle cadute e la posizione complessiva (Giofrè et al., 2017; Toffalini et al., 2017). Un andamento analogo si osserva negli adulti con diagnosi di disturbo dell'apprendimento (Pizzigallo et al., 2023).

19.5 Il problema dell'invarianza di misura

Quando si usa una batteria in una popolazione diversa da quella su cui è stata costruita, bisogna chiedersi se i punteggi abbiano lo stesso significato. È la questione dell'**invarianza di misura**, che si articola in livelli successivi (Tabella 19.2), ciascuno riferito a un tipo di parametro del modello fattoriale (Figura 19.4) (Meredith, 1993; Putnick & Bornstein, 2016).

Tabella 19.2: Livelli di invarianza di misura.

Livello	Domanda	Situazione tipica nei gruppi clinici
Configurale	La struttura dei fattori è la stessa?	Di solito garantita dalla costruzione della batteria
Metrica	I legami fra subtest, indici e fattore generale sono uguali?	Spesso violata, con conseguente incertezza interpretativa
Scalare	A parità di abilità vera si ottiene lo stesso punteggio?	Spesso violata: lo stesso QI non indica lo stesso livello di g
Stretta	L'errore di misura è lo stesso?	Spesso violata, ma con conseguenze minori

L'invarianza scalare è la più importante per l'interpretazione. Se un gruppo ottiene punteggi più bassi in alcuni subtest a parità di abilità generale,

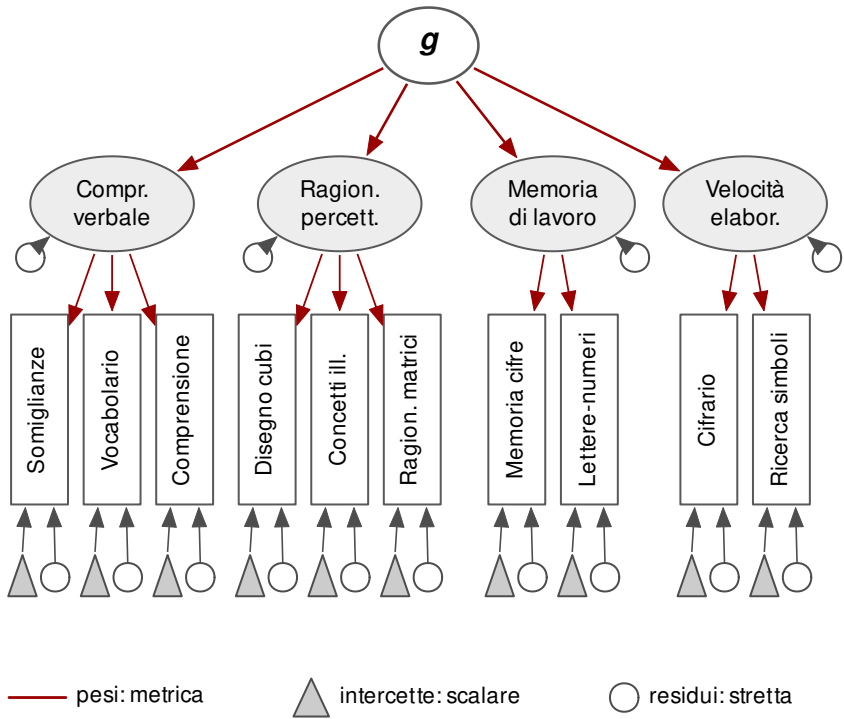


Figura 19.4: Dove si collocano i parametri che l'invarianza di misura confronta fra gruppi, nel modello gerarchico della WISC-IV. Pesi (frecce rosse, da g agli indici e dagli indici ai subtest): invarianza metrica. Intercette, cioè il punteggio atteso in un subtest quando l'abilità vera è esattamente nella media (triangoli): invarianza scalare. Varianze residue, cioè la parte di un indice o di un subtest non spiegata dal fattore superiore (cerchi): invarianza stretta.

il QI totale di quel gruppo sottostima l'abilità generale. E la mancanza di invarianza non riguarda solo i gruppi clinici: può interessare fasce d'età, generi, gruppi linguistici e culturali, persone bilingui.

Una lettura rigida concluderebbe che senza invarianza completa non si possa usare lo strumento. In pratica lo si usa comunque, perché spesso è l'unico disponibile o perché le alternative presentano lo stesso problema; perché tarature separate per ogni gruppo non sarebbero praticabili; perché lo scopo può essere solo escludere una disabilità intellettiva; perché la violazione può essere parziale e incidere poco sul calcolo dei punteggi. Serve allora sapere dove sta il problema.

Analisi su dati italiani della WISC-IV, che confrontano il campione normativo con ampi campioni clinici di bambini con disturbi dell'apprendimento o ADHD, indicano che in entrambi i gruppi i subtest di memoria di lavoro e velocità di elaborazione hanno medie più basse di quanto l'abilità generale farebbe prevedere, mentre gli altri no. Nei disturbi dell'apprendimento, inoltre, il fattore generale è rappresentato meno bene in quasi tutti gli indici, tranne il ragionamento percettivo (Toffalini et al., 2022). Per l'autismo, uno studio sulla WISC-V ha trovato una buona invarianza dei legami fattoriali ma non dei punteggi attesi: a parità di abilità generale, i bambini autistici ottenevano punteggi più bassi nella memoria di cifre e nel cifrario. Un modello basato sul solo indice di abilità generale risultava invece pienamente invariante (Stephenson et al., 2021).

La conseguenza pratica è la stessa nei tre casi. Quando memoria di lavoro e velocità sono compromesse per ragioni specifiche, il QI totale sottostima il funzionamento intellettivo generale, e l'**indice di abilità generale**, che le esclude, ne è una stima migliore. È ciò che accadeva nel caso illustrativo del capitolo precedente: un QI totale di 84 e un indice di abilità generale di 100 descrivono la stessa bambina in modo molto diverso, e il secondo è più vicino alle sue capacità di ragionamento.

Lo stesso meccanismo può nascondere un'abilità molto alta. Un bambino con un ragionamento eccellente e un disturbo dell'apprendimento può ottenere un QI totale soltanto nella media, perché le cadute in memoria di lavoro e velocità abbassano il totale; e un rendimento scolastico accettabile, perché il ragionamento compensa in parte la difficoltà specifica. Le due

rarietà si mascherano a vicenda. Leggere gli indici, in questi casi, non è un vezzo da profilo: è l'unico modo di vedere entrambe.

19.6 Valutare bene

La valutazione dell'intelligenza nel neurosviluppo richiede più attenzione che nella popolazione generale, non meno. Il QI resta indispensabile per escludere o confermare una disabilità intellettiva. Il profilo può orientare gli interventi. Ma i profili medi dei gruppi non sono marcatori diagnostici; i punteggi possono non avere lo stesso significato che hanno nella popolazione di riferimento; e l'indice più informativo dipende dalla persona che si ha davanti. Scegliere quale numero riportare è già un atto interpretativo, e va motivato.

In breve

- La disabilità intellettiva richiede limitazioni sia intellettive sia adattive; la gravità si stabilisce oggi soprattutto in base al bisogno di supporto.
- Nell'autismo convivono un'alta frequenza di disabilità intellettiva nelle casistiche cliniche e una correlazione genetica debolmente positiva con l'intelligenza; la ricerca sottorappresenta i casi a basso funzionamento.
- In ADHD e disturbi dell'apprendimento le cadute medie più evidenti riguardano memoria di lavoro e velocità di elaborazione; la specificità dei disturbi dell'apprendimento è imposta per definizione.
- Nei gruppi clinici manca spesso l'invarianza scalare: il QI totale può sottostimare l'abilità generale, e l'indice di abilità generale è spesso preferibile.
- I profili medi descrivono gruppi e non permettono inferenze diagnostiche sul singolo caso.

Riferimenti bibliografici

Parte VIII

Approfondimenti: oltre g

20 Altre intelligenze o vecchi costrutti con un nome nuovo?

La visione dominante dell'intelligenza è potente. Un singolo fattore generale riassume gran parte della varianza interpretabile nei test cognitivi, si misura in modo rapido e affidabile, prevede esiti importanti. Proprio questa potenza alimenta due domande. Il fattore generale è davvero tutto? Esistono altre intelligenze che i test non colgono? Questo approfondimento affronta la seconda domanda attraverso il caso più noto, l'intelligenza emotiva. Il prossimo affronta una domanda ancora più radicale: se g esista davvero.

20.1 Una buona idea

Le emozioni esistono e contano. Regolare le proprie emozioni aiuta a funzionare meglio. Capire quelle degli altri e comportarsi di conseguenza è altrettanto importante. Nessuno lo contesta. L'idea di **intelligenza emotiva** nasce qui: esisterebbe una capacità di percepire, comprendere, usare e regolare le emozioni, diversa dall'intelligenza cognitiva e altrettanto importante (Salovey & Mayer, 1990). Negli anni Novanta un libro di grande successo la presentò al pubblico come fattore decisivo del successo nella vita, persino più del QI (Goleman, 1995).

Le domande da porre sono quelle che si pongono per qualsiasi costrutto. Esiste davvero una capacità mentale identificabile dietro questa etichetta? È misurabile in modo abbastanza attendibile per la ricerca e la clinica? Aggiunge qualcosa a ciò che sappiamo già misurare? E, se sì, è necessario chiamarla «intelligenza»?

20.2 Due modi di misurarla

L'intelligenza emotiva si misura in due modi molto diversi, che corrispondono in pratica a due costrutti diversi.

Il primo la tratta come un'**abilità**. Si presentano scenari e si chiede di riconoscere l'emozione più probabile di un personaggio, o di valutare quanto una certa azione aiuterebbe qualcuno a mantenere il buon umore. Le risposte vengono confrontate con un criterio. Qui sorge il primo problema: qual è la risposta corretta? Spesso quella scelta dalla maggioranza, o da un gruppo di esperti. È un criterio molto diverso dalla soluzione di una matrice.

Il secondo la tratta come un **tratto** e la misura con questionari autodescrittivi. Le affermazioni riguardano l'autostima, la soddisfazione per la propria vita, la facilità a calmarsi, il controllo degli impulsi, la capacità di esprimere affetto, l'assertività (Petrides & Furnham, 2001). Chi conosce i questionari di personalità vi riconoscerà molti contenuti familiari.

20.3 Che cosa misurano davvero

Gli stessi proponenti della versione come abilità hanno ridimensionato gli entusiasmi giornalistici (Mayer et al., 2008). I punteggi correlano attorno a 0,2–0,3 con il rendimento scolastico e la prestazione lavorativa: valori apprezzabili, ma lontani dalle promesse. Le prove di abilità emotiva correlano in misura moderata con l'intelligenza verbale, e le loro diverse componenti correlano fra loro in modo molto variabile, da deboli a moderate.

Per la versione come tratto il problema è la sovrapposizione con la personalità. I punteggi correlano fortemente, in senso negativo, con il nevroticismo e, positivamente, con estroversione e coscienziosità. Una meta-analisi ha trovato che l'intelligenza emotiva di tratto coincide quasi completamente con un fattore generale della personalità, che riassume i poli socialmente desiderabili dei cinque grandi tratti (van der Linden et al., 2017).

La domanda decisiva è quindi quella della **validità incrementale**: tolti il fattore cognitivo generale e i tratti di personalità già noti, resta qualcosa che prevede gli esiti? Le analisi più accurate rispondono che resta poco. Nel lavoro, l'intelligenza emotiva aggiunge una piccola quota di previsione rispetto ad abilità cognitive e personalità, più per i questionari che per le prove di abilità, e in parte proprio perché i questionari includono contenuti di autovalutazione e motivazione che altri strumenti non misurano (Joseph & Newman, 2010).

C'è poi un problema pratico, rilevante soprattutto nella selezione del personale, dove questi strumenti sono molto usati. Un questionario autodescrittivo si può falsare: chi vuole fare buona impressione sa quali risposte sono desiderabili. In un esperimento, i partecipanti a cui era stato chiesto di presentarsi nel modo migliore possibile hanno alzato in modo sensibile i propri punteggi in un questionario di intelligenza emotiva, mentre non sono riusciti a farlo in una prova di abilità emotiva (Day & Carroll, 2008). È lo stesso motivo per cui nessuno misurerebbe l'intelligenza cognitiva chiedendo alle persone quanto si ritengono brillanti. Una misura di prestazione ha i suoi problemi di criterio; una misura autodescrittiva ha quelli di sincerità e di conoscenza di sé.

20.4 Una lista di qualità desiderabili

Il problema di fondo è concettuale. Nelle versioni più diffuse, l'intelligenza emotiva è un insieme troppo ampio e malleabile per essere un costrutto unitario: empatia, autoregolazione, ottimismo, leadership, abilità sociali, motivazione, resilienza. Assomiglia più a un elenco di caratteristiche desiderabili che a una capacità mentale con una struttura propria.

E perché chiamarla «intelligenza»? Già per il fattore cognitivo generale la parola è un'etichetta che accettiamo per la sua utilità pratica, consapevoli del rischio di reificazione. Estenderla a costrutti psicometricamente più deboli non aggiunge chiarezza. Aggiunge prestigio.

Questo non significa che competenze emotive e sociali non esistano o non contino. Significa che, per studiarle bene, conviene descriverle con costrut-

ti precisi e misure valide, e verificare ogni volta che non siano vecchie conoscenze travestite.

20.5 Jingle, jangle e la proliferazione delle etichette

L'intelligenza emotiva non è un caso isolato. Autoefficacia, grinta, resilienza, mentalizzazione, soft skills, flessibilità, competenza emotiva, agency: la psicologia della personalità, della motivazione, del benessere e del lavoro produce di continuo nuove etichette. Due errori classici aiutano a orientarsi.

- La **fallacia jingle** consiste nel dare lo stesso nome a costrutti diversi: due questionari di «resilienza» possono misurare cose differenti.
- La **fallacia jangle** consiste nel dare nomi diversi allo stesso costrutto: una nuova scala con un nome nuovo può misurare, nella sostanza, un tratto di personalità noto da decenni (Lawson & Robins, 2021).

La seconda è particolarmente frequente in psicologia, dove ogni gruppo di ricerca tende a costruirsi i propri strumenti. Un'analisi di migliaia di questionari ha mostrato che la maggior parte delle misure psicologiche viene usata pochissime volte, spesso solo dagli autori che le hanno create (Elson et al., 2023). Il risultato è una letteratura frammentata, difficile da cumulare, in cui la novità del nome sostituisce la novità della scoperta.

Il rimedio non è vietare i nuovi costrutti, ma chiedere loro tre cose: una definizione che li distingua da quelli esistenti; una misura che non correli così tanto con costrutti noti da esserne un duplicato; e la prova che prevedano qualcosa di rilevante al di là di ciò che quei costrutti già prevedono. Il fattore generale dell'intelligenza ha superato queste prove. Le altre «intelligenze», per ora, molto meno.

In breve

- L'intelligenza emotiva come abilità si sovrappone in parte all'intelligenza cognitiva; come tratto autodescritto, quasi completamente alla personalità.

- La sua validità incrementale rispetto a intelligenza e personalità è piccola.
- Il costrutto è spesso un insieme eterogeneo di qualità desiderabili; chiamarlo «intelligenza» aggiunge prestigio più che chiarezza.
- Le fallacie jingle e jangle e la proliferazione di strumenti poco usati sono problemi diffusi; ogni nuovo costrutto deve dimostrare di essere distinto e utile.

Riferimenti bibliografici

21 Si può spiegare la positive manifold senza un fattore generale?

La positive manifold è uno dei risultati più solidi della psicologia. Su questo non c'è disaccordo. Il disaccordo riguarda la spiegazione. Quella standard, presentata nei capitoli sui modelli dell'intelligenza, ipotizza un insieme di cause comuni riassunte dal fattore generale g (Figura 21.1). Ma quelle cause restano in buona parte misteriose, e a molti ricercatori l'idea di una variabile latente non osservata sembra un'ipotesi troppo comoda.

Negli ultimi vent'anni sono emerse almeno due teorie che cercano di spiegare le correlazioni fra abilità senza un fattore generale. Sono alternative serie, che meritano di essere conosciute. Sono anche, a oggi, minoritarie. Questo approfondimento spiega perché.

21.1 Un vecchio precedente

L'idea non è nuova. Già nel 1916 Godfrey Thomson mostrò che una positive manifold può emergere anche senza un fattore generale: basta che ogni prova attinga a un campione casuale di molti processi elementari, e che prove diverse condividano in parte gli stessi processi (Thomson, 1916). Più processi due prove hanno in comune, più correlano. La **teoria del campionamento** di Thomson è stata riproposta in forma moderna, ed è compatibile con molti dei dati che sostengono g (Bartholomew et al., 2009). Mostra un punto importante: il fatto che un modello a un fattore descriva bene i dati non dimostra, da solo, che esista una singola causa comune.

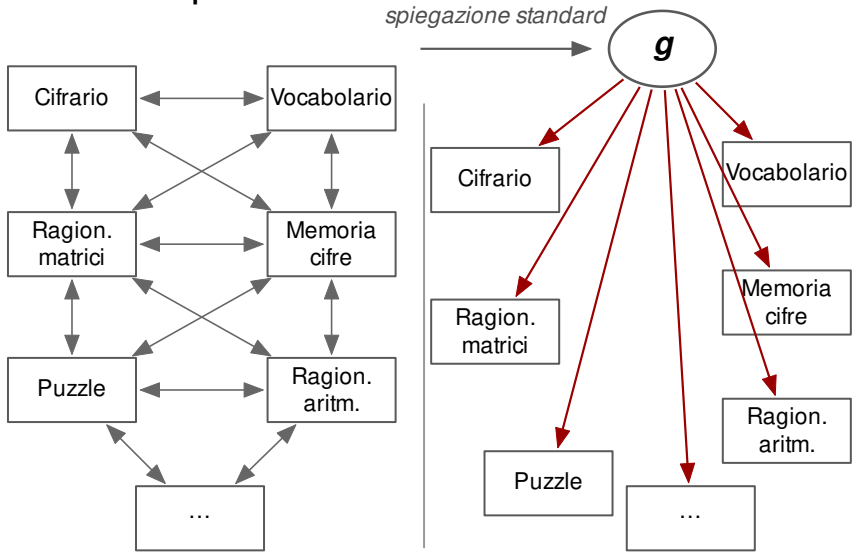
Il dato: tutte le prove correlano

Figura 21.1: Il dato e la spiegazione standard. Nel primo pannello il dato da spiegare, la positive manifold: le prove di una batteria correlano tutte positivamente fra loro (freccie doppie). Nel secondo la spiegazione standard: una causa comune, g , influenza tutte le prove, e le correlazioni ne sono la conseguenza.

21.2 La teoria della sovrapposizione dei processi

La **teoria della sovrapposizione dei processi** (POT) aggiorna l'idea di Thomson con il linguaggio della psicologia cognitiva (Kovacs & Conway, 2016). Ogni prova richiede sia processi specifici del dominio, sia processi esecutivi generali, come il controllo dell'attenzione e la memoria di lavoro. Questi ultimi sono coinvolti in molte prove e funzionano come un collo di bottiglia: limitano la prestazione ovunque. Le prove correlano perché condividono quei limiti.

La teoria spiega alcuni fatti noti. Il primo è che le prove più complesse, come i problemi di ragionamento con molte regole da tenere a mente, so-

no le più legate al fattore generale: sono anche quelle che richiedono più controllo esecutivo.

Il secondo fatto è noto da quasi un secolo come **legge dei rendimenti decrescenti** di Spearman: nei gruppi con abilità più alta le correlazioni fra prove diverse, e quindi il peso del fattore generale, tendono a essere più basse. Una meta-analisi ne ha confermato l'esistenza, con un effetto di entità modesta (Blum & Holling, 2017). La teoria della sovrapposizione dei processi lo spiega in modo naturale: se i processi esecutivi limitano la prestazione soprattutto quando sono deboli, le prove dipendono più spesso dagli stessi limiti nelle persone meno capaci. Anche un modello con un fattore generale può però accomodare il dato, ammettendo che il suo peso vari con il livello.

Ma a uno sguardo attento non è una vera alternativa all'ipotesi delle cause comuni. Sostituisce un fattore generale con un insieme di processi generali, cioè con altre cause comuni, descritte con maggiore dettaglio cognitivo. Ed è rimasta minoritaria soprattutto perché quei processi esecutivi non sono stati isolati e misurati in modo abbastanza chiaro da metterla alla prova.

21.3 Il mutualismo

Il **mutualismo** propone qualcosa di concettualmente più radicale (van der Maas et al., 2006). Immaginiamo che, all'inizio dello sviluppo, abilità diverse siano indipendenti. Durante la crescita, però, si sostengono a vicenda, in un intreccio di influenze reciproche che agisce soprattutto nel lungo periodo (Figura 21.2).

Un esempio aiuta (Figura 21.3). Un vocabolario più ricco aiuta a capire concetti e spiegazioni, quindi a rappresentarsi meglio i problemi e a ragionare, prima su problemi esposti a parole e poi anche su quelli non verbali. Un ragionamento migliore aiuta a ricavare il significato di parole nuove dal contesto, quindi ad arricchire il vocabolario. Ripetuti per anni, questi scambi reciproci producono abilità correlate anche partendo da abilità indipendenti, come mostra una semplice simulazione (Figura 21.4). La positive manifold sarebbe un fenomeno emergente, non il segno di una causa comune.

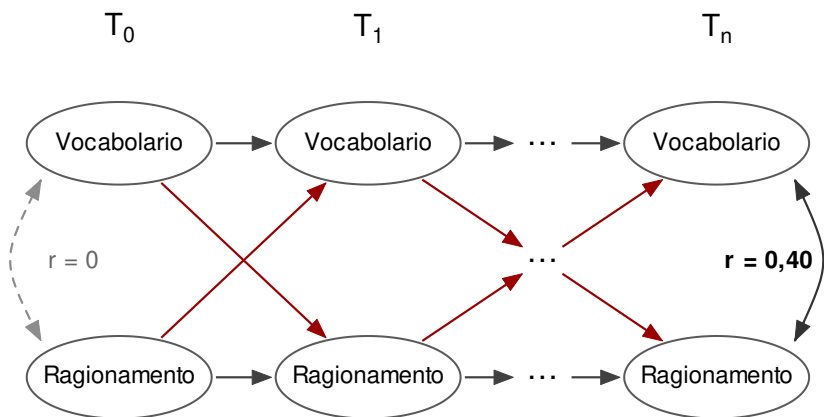


Figura 21.3: Il meccanismo del mutualismo in due abilità. All’inizio dello sviluppo (T_0) vocabolario e ragionamento non sono correlati; a ogni passo ciascuna abilità contribuisce alla crescita dell’altra (freccie rosse), e dopo molti scambi (T_n) le due abilità risultano correlate. La correlazione finale è illustrativa. Adattamento da uno schema dell’autore, sulla base di Kievit et al. (2017).

21.4 Le critiche

Le critiche al mutualismo come teoria generale dell’intelligenza sono però numerose.

- **La positive manifold c’è fin dall’inizio.** Emerge appena le abilità diventano misurabili, nei primi anni di vita, e non sembra rafforzarsi in modo marcato con l’età, come il mutualismo prevederebbe (Gignac, 2014).
- **Il fattore generale resta più parsimonioso.** Una volta considerato g , le relazioni residue fra abilità specifiche sono minime o nulle. Una rete complessa di influenze reciproche non aggiunge capacità esplicativa (Gignac, 2016). Anche nei dati sui gemelli, i modelli fattoriali descrivono le relazioni fra abilità almeno altrettanto bene dei modelli di rete (Knyspel & Plomin, 2024).

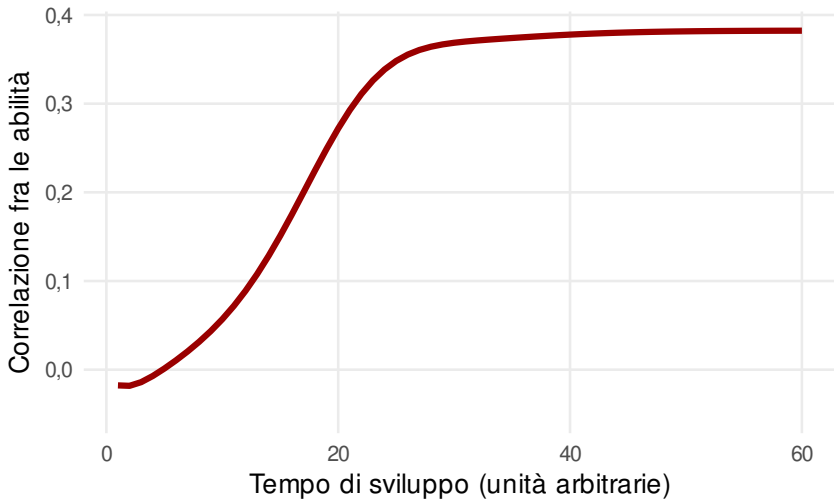


Figura 21.4: Simulazione di un processo mutualistico. Due abilità partono indipendenti e crescono nel tempo verso un livello massimo diverso per ogni persona e per ogni abilità, scelto a caso; a ogni passo ciascuna riceve una piccola spinta proporzionale al livello dell'altra. La correlazione fra le due, inizialmente nulla, cresce durante lo sviluppo. Parametri scelti a scopo illustrativo.

- **Le previsioni sono imprecise.** Quali abilità influenzano quali? In quale periodo? Con quale intensità? Senza risposte precise, la teoria è difficile da falsificare.
- **Le evidenze longitudinali sono limitate.** Riguardano di solito due o tre abilità alla volta e usano modelli, come quelli a ritardi incrociati, che non distinguono bene fra influenze reciproche e differenze stabili fra persone (Hamaker et al., 2015).

21.5 Che cosa sopravvive

Il mutualismo sopravvive soprattutto in una versione compatibilista: un fattore generale e una rete di influenze reciproche possono coesistere (Savi et al., 2019). È plausibile che abilità diverse si rinforzino a vicenda in

alcuni periodi critici dello sviluppo, in misura diversa da persona a persona, anche in presenza di cause comuni. Come cornice concettuale resta utile per pensare lo sviluppo in modo dinamico, e in particolare il neurosviluppo, dove difficoltà in un'area possono propagarsi ad altre. Il prezzo della versione compatibilista è che perde la sua radicalità: se servono anche variabili latenti, la rete non è più un'alternativa al fattore generale, ma un'aggiunta.

Il bilancio, oggi, è questo. Le alternative a g hanno il merito di ricordare che un fattore statistico non è una spiegazione causale e che la positive manifold potrebbe avere origini diverse da una singola causa comune. Ma nessuna ha ancora prodotto previsioni precise che il modello gerarchico non riesca a spiegare altrettanto bene. Finché non lo faranno, restano ipotesi da seguire con interesse, non sostituti del fattore generale.

In breve

- La teoria del campionamento di Thomson mostra che una positive manifold può emergere anche senza un singolo fattore generale.
- La teoria della sovrapposizione dei processi attribuisce le correlazioni a processi esecutivi generali: aggiunge dettaglio cognitivo, ma resta una teoria di cause comuni.
- Il mutualismo spiega la positive manifold con influenze reciproche fra abilità durante lo sviluppo: è l'unica vera alternativa logica alle cause comuni.
- Le critiche riguardano la presenza precoce della positive manifold, la maggiore parsimonia di g e l'imprecisione delle previsioni; la versione compatibilista sopravvive come cornice per lo sviluppo.

Riferimenti bibliografici

Riferimenti bibliografici

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf
- Arden, R., Luciano, M., Deary, I. J., Reynolds, C. A., Pedersen, N. L., Plassman, B. L., McGue, M., Christensen, K., & Visscher, P. M. (2016). The association between intelligence and lifespan is mostly genetic. *International Journal of Epidemiology*, 45(1), 178–185. <https://doi.org/10.1093/ije/dyv112>
- Association, A. P. (2022). *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association Publishing. <https://doi.org/10.1176/appi.books.9780890425787>
- Astle, D. E., & Fletcher-Watson, S. (2020). Beyond the core-deficit hypothesis in developmental disorders. *Current Directions in Psychological Science*, 29(5), 431–437.
- Astle, D. E., Holmes, J., Kievit, R., & Gathercole, S. E. (2022). Annual Research Review: The transdiagnostic revolution in neurodevelopmental disorders. *Journal of Child Psychology and Psychiatry*, 63(4), 397–417. <https://doi.org/10.1111/jcpp.13481>
- Baron, R. M., & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182. <https://doi.org/10.1037/0022-3514.51.6.1173>
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44(1), 1–26. <https://doi.org/10.1111/j.1744-6570.1991.tb00688.x>
- Bartholomew, D. J., Deary, I. J., & Lawn, M. (2009). A new lease of life for Thomson's bonds model of intelligence. *Psychological Review*, 116(3), 567–579. <https://doi.org/10.1037/a0016262>

- Basten, U., Hilger, K., & Fiebach, C. J. (2015). Where smart brains are different: A quantitative meta-analysis of functional and structural brain imaging studies on intelligence. *Intelligence*, 51, 10–27. <https://doi.org/10.1016/j.intell.2015.04.009>
- Basten, U., Stelzel, C., & Fiebach, C. J. (2013). Intelligence is differentially related to neural effort in the task-positive and the task-negative brain network. *Intelligence*, 41(5), 517–528. <https://doi.org/10.1016/j.intell.2013.07.006>
- Binet, A., & Simon, T. (1904). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'année psychologique*, 11(1), 191–244. <https://doi.org/10.3406/psy.1904.3675>
- Binet, A., & Simon, T. (1916). *The Development of Intelligence in Children (The Binet-Simon Scale)* (E. S. Kite, Trad.). Williams & Wilkins.
- Blau, N., van Spronsen, F. J., & Levy, H. L. (2010). Phenylketonuria. *The Lancet*, 376(9750), 1417–1427. [https://doi.org/10.1016/S0140-6736\(10\)60961-0](https://doi.org/10.1016/S0140-6736(10)60961-0)
- Blum, D., & Holling, H. (2017). Spearman's law of diminishing returns. A meta-analysis. *Intelligence*, 65, 60–66. <https://doi.org/10.1016/j.intell.2017.07.004>
- Boring, E. G. (1923). Intelligence as the Tests Test It. *New Republic*, 36, 35–37.
- Bratsberg, B., & Rogeberg, O. (2018). Flynn effect and its reversal are both environmentally caused. *Proceedings of the National Academy of Sciences*, 115(26), 6674–6678. <https://doi.org/10.1073/pnas.1718793115>
- Brigham, C. C. (1923). *A Study of American Intelligence*. Princeton University Press.
- Brigham, C. C. (1930). Intelligence tests of immigrant groups. *Psychological Review*, 37(2), 158–165. <https://doi.org/10.1037/h0072570>
- Briley, D. A., & Tucker-Drob, E. M. (2013). Explaining the Increasing Heritability of Cognitive Ability Across Development: A Meta-Analysis of Longitudinal Twin and Adoption Studies. *Psychological Science*, 24(9), 1704–1713. <https://doi.org/10.1177/0956797613478618>
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376. <https://doi.org/10.1038/nrn3475>

- Calvin, C. M., Batty, G. D., Der, G., Brett, C. E., Taylor, A., Pattie, A., Čukić, I., & Deary, I. J. (2017). Childhood intelligence in relation to major causes of death in 68 year follow-up: prospective population study. *BMJ*, j2708. <https://doi.org/10.1136/bmj.j2708>
- Calvin, C. M., Deary, I. J., Fenton, C., Roberts, B. A., Der, G., Leckenby, N., & Batty, G. D. (2011). Intelligence in youth and all-cause-mortality: systematic review with meta-analysis. *International Journal of Epidemiology*, 40(3), 626–644. <https://doi.org/10.1093/ije/dyq190>
- Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early Childhood Education: Young Adult Outcomes From the Abecedarian Project. *Applied Developmental Science*, 6(1), 42–57. https://doi.org/10.1207/S1532480XADS0601_05
- Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children–Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment*, 29(4), 458–472. <https://doi.org/10.1037/pas0000358>
- Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511571312>
- Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clinical Psychological Science*, 2(2), 119–137. <https://doi.org/10.1177/2167702613497473>
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1–22. <https://doi.org/10.1037/h0046743>
- Cattell, R. B. (1987). *Intelligence: Its Structure, Growth and Action* (Vol. 35). North-Holland.
- Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The Fourth Law of Behavior Genetics. *Current Directions in Psychological Science*, 24(4), 304–312. <https://doi.org/10.1177/0963721415580430>
- Childhood cognitive ability and adult academic attainment: evidence from three British cohort studies. (2010). *Longitudinal and Life Course Studies*, 1(3). <https://doi.org/10.14301/llcs.v1i3.93>

- Cornoldi, C. (2007). *L'intelligenza*. Il Mulino.
- Cornoldi, C., Belacchi, C., Giofrè, D., Martini, A., & Tressoldi, P. (2010). The mean Southern Italian children IQ is not particularly low: A reply to R. Lynn (2010). *Intelligence*, 38(5), 462–470. <https://doi.org/10.1016/j.intell.2010.06.003>
- Cornoldi, C., & Vecchi, T. (2003). *Visuo-spatial Working Memory and Individual Differences*. Psychology Press.
- Cox, S. R., Ritchie, S. J., Fawns-Ritchie, C., Tucker-Drob, E. M., & Deary, I. J. (2019). Structural brain imaging correlates of general intelligence in UK Biobank. *Intelligence*, 76, 101376. <https://doi.org/10.1016/j.intell.2019.101376>
- Crespi, B. J. (2016). Autism as a disorder of high intelligence. *Frontiers in Neuroscience*, 10, 300. <https://doi.org/10.3389/fnins.2016.00300>
- Crompton, C. J., Ropar, D., Evans-Williams, C. V. M., Flynn, E. G., & Fletcher-Watson, S. (2020). Autistic peer-to-peer information transfer is highly effective. *Autism*, 24(7), 1704–1712. <https://doi.org/10.1177/1362361320919286>
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12(11), 671–684. <https://doi.org/10.1037/h0043943>
- Cucina, J. M., Burtnick, S. K., De la Flor Musso, M. E., Walmsley, P. T., & Wilson, K. J. (2024). Meta-analytic validity of cognitive ability for hands-on military job proficiency. *Intelligence*, 104, 101818. <https://doi.org/10.1016/j.intell.2024.101818>
- Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D., Hagenaars, S. P., Ritchie, S. J., Marioni, R. E., Fawns-Ritchie, C., Liewald, D. C. M., Okely, J. A., Ahola-Olli, A. V., Barnes, C. L. K., Bertram, L., Bis, J. C., Burdick, K. E., Christoforou, A., DeRosse, P., ... Deary, I. J. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications*, 9(1), 2098. <https://doi.org/10.1038/s41467-018-04362-x>
- Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., Ke, X., Le Hellard, S., Christoforou, A., Luciano, M., McGhee, K., Lopez, L., Gow, A. J., Corley, J., Redmond, P., Fox, H. C., Haggarty, P., Whalley, L. J., McNeill, G., ... Deary, I. J. (2011). Genome-wide association stu-

- dies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*, 16(10), 996–1005. <https://doi.org/10.1038/mp.2011.85>
- Davies, N. M., Hill, W. D., Anderson, E. L., Sanderson, E., Deary, I. J., & Davey Smith, G. (2019). Multivariable two-sample Mendelian randomization estimates of the effects of intelligence and education on health. *eLife*, 8, e43990. <https://doi.org/10.7554/eLife.43990>
- Day, A. L., & Carroll, S. A. (2008). Faking emotional intelligence (EI): comparing response distortion on ability and trait-based EI measures. *Journal of Organizational Behavior*, 29(6), 761–784. <https://doi.org/10.1002/job.485>
- Dearborn, W. F. (1921). Intelligence and its measurement: A symposium—XII. *Journal of Educational Psychology*, 12(4), 210–212. <https://doi.org/10.1037/h0065003>
- Deary, I. J. (2012a). Intelligence. *Annual Review of Psychology*, 63(1), 453–482. <https://doi.org/10.1146/annurev-psych-120710-100353>
- Deary, I. J. (2012b). Looking for «System Integrity» in Cognitive Epidemiology. *Gerontology*, 58(6), 545–553. <https://doi.org/10.1159/000341157>
- Deary, I. J., Cox, S. R., & Hill, W. D. (2022). Genetic variation, brain, and intelligence differences. *Molecular Psychiatry*, 27(1), 335–353. <https://doi.org/10.1038/s41380-021-01027-y>
- Deary, I. J., Hill, W. D., & Gale, C. R. (2021). Intelligence, health and death. *Nature Human Behaviour*, 5(4), 416–430. <https://doi.org/10.1038/s41562-021-01078-9>
- Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35(1), 13–21. <https://doi.org/10.1016/j.intell.2006.02.001>
- Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R., & Starr, J. M. (2000). The Stability of Individual Differences in Mental Ability from Childhood to Old Age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence*, 28(1), 49–55. [https://doi.org/10.1016/S0160-2896\(99\)00031-8](https://doi.org/10.1016/S0160-2896(99)00031-8)
- Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The Impact of Childhood Intelligence on Later Life: Following Up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality*

- and *Social Psychology*, 86(1), 130–147. <https://doi.org/10.1037/0022-3514.86.1.130>
- Demontis, D., Walters, G. B., Athanasiadis, G., Walters, R., Therrien, K., Nielsen, T. T., Farajzadeh, L., Voloudakis, G., Bendl, J., Zeng, B., Zhang, W., Grove, J., Als, T. D., Duan, J., Satterstrom, F. K., Bybjerg-Grauholm, J., Bækved-Hansen, M., Gudmundsson, O. O., Magnusson, S. H., ... Børglum, A. D. (2023). Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nature Genetics*, 55(2), 198–208. <https://doi.org/10.1038/s41588-022-01285-8>
- Duncan, G. J., & Magnuson, K. (2013). Investing in Preschool Programs. *Journal of Economic Perspectives*, 27(2), 109–132. <https://doi.org/10.1257/jep.27.2.109>
- Dutton, E., van der Linden, D., & Lynn, R. (2016). The negative Flynn Effect: A systematic literature review. *Intelligence*, 59, 163–169. <https://doi.org/10.1016/j.intell.2016.10.002>
- Dworak, E. M., Revelle, W., & Condon, D. M. (2023). Looking for Flynn effects in a recent online U. S. adult sample: Examining shifts within the SAPA Project. *Intelligence*, 98, 101734. <https://doi.org/10.1016/j.intell.2023.101734>
- Ekstein, S., Glick, B., Weill, M., Kay, B., & Berger, I. (2011). Down syndrome and attention-deficit/hyperactivity disorder (ADHD). *Journal of Child Neurology*, 26(10), 1290–1295. <https://doi.org/10.1177/0883073811405201>
- Elson, M., Hussey, I., Alsalti, T., & Arslan, R. C. (2023). Psychological measures aren't toothbrushes. *Communications Psychology*, 1(1), 25. <https://doi.org/10.1038/s44271-023-00026-9>
- Federal Trade Commission. (2016). *Lumosity to Pay \$2 Million to Settle FTC Deceptive Advertising Charges for Its «Brain Training» Program*. Comunicato stampa. <https://www.ftc.gov/news-events/news/press-releases/2016/01/lumosity-pay-2-million-settle-ftc-deceptive-advertising-charges-its-brain-training-program>
- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, 53(1), 25–44. <https://doi.org/10.1016/j.jsp.2014.10.002>

- Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, 95(1), 29–51. <https://doi.org/10.1037/0033-2909.95.1.29>
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101(2), 171–191. <https://doi.org/10.1037/0033-2909.101.2.171>
- Flynn, J. R. (2020). *Secular Changes in Intelligence: The “Flynn Effect”* (R. J. Sternberg, A. c. Di; 2^a ed., pp. 940–963). Cambridge University Press. <https://doi.org/10.1017/9781108770422.040>
- Frazier, T. W., Demaree, H. A., & Youngstrom, E. A. (2004). Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. *Neuropsychology*, 18(3), 543–555. <https://doi.org/10.1037/0894-4105.18.3.543>
- Frinco, M., Sironi, E. M., & Frinco, R. (2024). The Psychometric Properties of Test. Why They Are Essential for Clinical and Academic Purposes. *Psicologia clinica dello sviluppo*, 1/2024. <https://doi.org/10.1449/113146>
- Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156–168. <https://doi.org/10.1177/2515245919847202>
- Galton, F. (1869). *Hereditary Genius: An Inquiry into Its Laws and Consequences*. Macmillan.
- Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. Basic Books.
- Genç, E., Fraenz, C., Schlüter, C., Friedrich, P., Hossiep, R., Voelkle, M. C., Ling, J. M., Güntürkün, O., & Jung, R. E. (2018). Diffusion markers of dendritic density and arborization in gray matter predict differences in intelligence. *Nature Communications*, 9(1), 1905. <https://doi.org/10.1038/s41467-018-04268-8>
- Gignac, G. E. (2014). Dynamic mutualism versus g factor theory: An empirical test. *Intelligence*, 42, 89–97. <https://doi.org/10.1016/j.intell.2013.11.004>
- Gignac, G. E. (2016). Residual group-level factor associations: Possibly negative implications for the mutualism theory of general intelligence. *Intelligence*, 55, 69–78. <https://doi.org/10.1016/j.intell.2016.01.007>
- Gignac, G. E., & Kretschmar, A. (2017). Evaluating dimensional distinct-

- ness with correlated-factor models: Limitations and suggestions. *Intelligence*, 62, 138–147. <https://doi.org/10.1016/j.intell.2017.04.001>
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78. <https://doi.org/10.1016/j.paid.2016.06.069>
- Giofrè, D., Pastore, M., Cornoldi, C., & Toffalini, E. (2019). Lumpers vs. splitters: Intelligence in children with specific learning disorders. *Intelligence*, 76, 101380. <https://doi.org/10.1016/j.intell.2019.101380>
- Giofrè, D., Toffalini, E., Altoè, G., & Cornoldi, C. (2017). Intelligence measures as diagnostic tools for children with specific learning disabilities. *Intelligence*, 61, 140–145. <https://doi.org/10.1016/j.intell.2017.01.014>
- Gobet, F., & Sala, G. (2023). Cognitive Training: A Field in Search of a Phenomenon. *Perspectives on Psychological Science*, 18(1), 125–141. <https://doi.org/10.1177/17456916221091830>
- Goleman, D. (1995). *Emotional Intelligence*. Bantam Books.
- Gottfredson, L. S. (1997). Mainstream Science on Intelligence: An Editorial with 52 Signatories, History, and Bibliography. *Intelligence*, 24(1), 13–23. [https://doi.org/10.1016/S0160-2896\(97\)90011-8](https://doi.org/10.1016/S0160-2896(97)90011-8)
- Gottfredson, L. S. (2004). Intelligence: Is It the Epidemiologists' Elusive "Fundamental Cause" of Social Class Inequalities in Health? *Journal of Personality and Social Psychology*, 86(1), 174–199. <https://doi.org/10.1037/0022-3514.86.1.174>
- Grassi, M., & Talamini, F. (2025). *Do musicians have better short-term memory than nonmusicians? A multi-lab study*. Preprint, PsyArXiv. https://doi.org/10.31234/osf.io/kghua_v1
- Gregory, H. (2015). *McNamara's Folly: The Use of Low-IQ Troops in the Vietnam War*. Infinity Publishing.
- Grieco, J., Pulsifer, M., Seligsohn, K., Skotko, B., & Schwartz, A. (2015). Down syndrome: Cognitive and behavioral functioning across the lifespan. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, 169(2), 135–149. <https://doi.org/10.1002/ajmg.c.31439>
- Guilford, J. P. (1967). *The Nature of Human Intelligence*. McGraw-Hill.
- Haier, R. J. (2017). *The Neuroscience of Intelligence*. Cambridge University Press.

- Haier, R. J., Siegel, B. V., MacLachlan, A., Soderling, E., Lottenberg, S., & Buchsbaum, M. S. (1992). Regional glucose metabolic changes after learning a complex visuospatial/motor task: a positron emission tomographic study. *Brain Research*, 570(1-2), 134–143. [https://doi.org/10.1016/0006-8993\(92\)90573-R](https://doi.org/10.1016/0006-8993(92)90573-R)
- Haier, R. J., Siegel, B. V., Nuechterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., Browning, H. L., & Buchsbaum, M. S. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence*, 12(2), 199–217. [https://doi.org/10.1016/0160-2896\(88\)90016-5](https://doi.org/10.1016/0160-2896(88)90016-5)
- Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The Science of Sex Differences in Science and Mathematics. *Psychological Science in the Public Interest*, 8(1), 1–51. <https://doi.org/10.1111/j.1529-1006.2007.00032.x>
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20(1), 102–116. <https://doi.org/10.1037/a0038889>
- Hamilton, A. F. de C. (2013). Reflecting on the mirror neuron system in autism: A systematic review of current theories. *Developmental Cognitive Neuroscience*, 3, 91–105. <https://doi.org/10.1016/j.dcn.2012.09.008>
- Haworth, C. M. A., Wright, M. J., Luciano, M., Martin, N. G., de Geus, E. J. C., van Beijsterveldt, C. E. M., Bartels, M., Posthuma, D., Boomsma, D. I., Davis, O. S. P., Kovas, Y., Corley, R. P., DeFries, J. C., Hewitt, J. K., Olson, R. K., Rhea, S.-A., Wadsworth, S. J., Iacono, W. G., McGue, M., ... Plomin, R. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry*, 15(11), 1112–1120. <https://doi.org/10.1038/mp.2009.55>
- Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, 94(1-2), 114–128. <https://doi.org/10.1016/j.jpubeco.2009.11.001>
- Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., Chittoor, G., Ahlskog, R., Lind, P. A., Palviainen, T., van der Zee, M. D., Cheesman, R., Mangino, M., Wang, Y., Li, S., Klaric, L., Ratliff, S. M., Bielak, L. F., Nygaard, M., ... Davies, N. M. (2022). Within-sibship genome-wide association analyses decrease bias in estimates of direct

- genetic effects. *Nature Genetics*, 54(5), 581–592. <https://doi.org/10.1038/s41588-022-01062-7>
- Ilieva, I. P., Hook, C. J., & Farah, M. J. (2015). Prescription Stimulants' Effects on Healthy Inhibitory Control, Working Memory, and Episodic Memory: A Meta-analysis. *Journal of Cognitive Neuroscience*, 27(6), 1069–1089. https://doi.org/10.1162/jocn_a_00776
- Intelligence and Its Measurement: A Symposium. (1921). *Journal of Educational Psychology*, 12(3), 123–147. <https://doi.org/10.1037/h0076078>
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, 105(19), 6829–6833. <https://doi.org/10.1073/pnas.0801268105>
- Jensen, A. R. (1989). The Milwaukee Project: Preventing Mental Retardation in Children at Risk. *Developmental Review*, 9(3), 234–258. [https://doi.org/10.1016/0273-2297\(89\)90030-0](https://doi.org/10.1016/0273-2297(89)90030-0)
- Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: consistent results from three test batteries. *Intelligence*, 32(1), 95–107. [https://doi.org/10.1016/S0160-2896\(03\)00062-X](https://doi.org/10.1016/S0160-2896(03)00062-X)
- Johnson, W., Carothers, A., & Deary, I. J. (2008). Sex differences in variability in general intelligence: A new look at the old question. *Perspectives on Psychological Science*, 3(6), 518–531. <https://doi.org/10.1111/j.1745-6924.2008.00096.x>
- Joseph, D. L., & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology*, 95(1), 54–78. <https://doi.org/10.1037/a0017286>
- Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences*, 30(2), 135–154. <https://doi.org/10.1017/S0140525X07001185>
- Kaufman, A. S. (2009). *IQ Testing 101*. Springer Publishing Company.
- Kell, H. J., Lubinski, D., & Benbow, C. P. (2013). Who rises to the top? Early indicators. *Psychological Science*, 24(5), 648–659. <https://doi.org/10.1177/0956797612457784>
- Kendler, K. S., Turkheimer, E., Ohlsson, H., Sundquist, J., & Sundquist, K.

- (2015). Family environment and the malleability of cognitive ability: A Swedish national home-reared and adopted-away cosibling control study. *Proceedings of the National Academy of Sciences*, 112(15), 4612–4617. <https://doi.org/10.1073/pnas.1417106112>
- Kievit, R. A., Lindenberger, U., Goodyer, I. M., Jones, P. B., Fonagy, P., Bullmore, E. T., the Neuroscience in Psychiatry Network, & Dolan, R. J. (2017). Mutualistic Coupling Between Vocabulary and Reasoning Supports Cognitive Development During Late Adolescence and Early Adulthood. *Psychological Science*, 28(10), 1419–1431. <https://doi.org/10.1177/0956797617710785>
- Kilpeläinen, T. O., Qi, L., Brage, S., Sharp, S. J., Sonestedt, E., Demerath, E., Ahmad, T., Mora, S., Kaakinen, M., Sandholt, C. H., Holzapfel, C., Autenrieth, C. S., Hyppönen, E., Cauchi, S., He, M., Kutalik, Z., Kumari, M., Stančáková, A., Meidtner, K., ... Loos, R. J. F. (2011). Physical Activity Attenuates the Influence of FTO Variants on Obesity Risk: A Meta-Analysis of 218,166 Adults and 19,268 Children. *PLoS Medicine*, 8(11), e1001116. <https://doi.org/10.1371/journal.pmed.1001116>
- Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., Gillberg, C. G., Forssberg, H., & Westerberg, H. (2005). Computerized Training of Working Memory in Children With ADHD-A Randomized, Controlled Trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44(2), 177–186. <https://doi.org/10.1097/00004583-200502000-00010>
- Knyspel, J., & Plomin, R. (2024). Comparing factor and network models of cognitive abilities using twin data. *Intelligence*, 104, 101833. <https://doi.org/10.1016/j.intell.2024.101833>
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., ... Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454–477. <https://doi.org/10.1037/abn0000258>
- Kovacs, K., & Conway, A. R. A. (2016). Process Overlap Theory: A Unified Account of the General Factor of Intelligence. *Psychological Inquiry*, 27(3), 151–177. <https://doi.org/10.1080/1047840X.2016.1153946>

- Kovas, Y., & Plomin, R. (2006). Generalist genes: implications for the cognitive sciences. *Trends in Cognitive Sciences*, 10(5), 198–203. <https://doi.org/10.1016/j.tics.2006.03.001>
- Kush, J. C., & Canivez, G. L. (2019). Construct validity of the WISC–IV Italian edition: A bifactor examination of the standardization sample. *International Journal of School & Educational Psychology*, 9(1), 73–87. <http://doi.org/10.1080/21683603.2019.1638854>
- Lawson, K. M., & Robins, R. W. (2021). Sibling Constructs: What Are They, Why Do They Matter, and How Should You Handle Them? *Personality and Social Psychology Review*, 25(4), 344–366. <https://doi.org/10.1177/10888683211047101>
- Lee, J. J., Wedow, R., Okbay, A., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, 50(8), 1112–1121. <https://doi.org/10.1038/s41588-018-0147-3>
- Lee, Y., Berry, C. M., & Gonzalez-Mulé, E. (2019). The importance of being humble: A meta-analysis and incremental validity analysis of the relationship between honesty-humility and job performance. *Journal of Applied Psychology*, 104(12), 1535–1546. <https://doi.org/10.1037/apl0000421>
- Levy, F., Hay, D. A., McStephen, M., Wood, C., & Waldman, I. (1997). Attention-Deficit Hyperactivity Disorder: A Category or a Continuum? Genetic Analysis of a Large-Scale Twin Study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36(6), 737–744. <https://doi.org/10.1097/00004583-199706000-00009>
- Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*, 51(3), 485–504. <https://doi.org/10.1002/ejsp.2752>
- Lubinski, D., & Benbow, C. P. (2006). Study of Mathematically Precocious Youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science*, 1(4), 316–345. <https://doi.org/10.1111/j.1745-6916.2006.00019.x>
- Lynn, R. (2010). In Italy, north–south differences in IQ predict differences in income, education, infant mortality, stature, and literacy. *Intelligence*, 38(1), 93–100. <https://doi.org/10.1016/j.intell.2009.07.004>
- Machts, N., Kaiser, J., Schmidt, F. T. C., & Möller, J. (2016). Accuracy of

- teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review*, 19, 85–103. <https://doi.org/10.1016/j.edur.2016.06.003>
- Maenner, M. J., Warren, Z., Williams, A. R., Amoakohene, E., Bakian, A. V., Bilder, D. A., Durkin, M. S., Fitzgerald, R. T., Furnier, S. M., Hughes, M. M., Ladd-Acosta, C. M., McArthur, D., Pas, E. T., Salinas, A., Vehorn, A., Williams, S., Esler, A., Grzybowski, A., Hall-Lande, J., ... Shaw, K. A. (2023). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2020. *MMWR. Surveillance Summaries*, 72(2), 1–14. <https://doi.org/10.15585/mmwr.ss7202a1>
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., ... Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747–753. <https://doi.org/10.1038/nature08494>
- Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Hendrickson, T. J., Malone, S. M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., ... Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, 603(7902), 654–660. <https://doi.org/10.1038/s41586-022-04492-9>
- Martens, M. A., Wilson, S. J., & Reutens, D. C. (2008). Research Review: Williams syndrome: a critical review of the cognitive, behavioral, and neuroanatomical phenotype. *Journal of Child Psychology and Psychiatry*, 49(6), 576–608. <https://doi.org/10.1111/j.1469-7610.2008.01887.x>
- Maxwell, S. E., & Cole, D. A. (2007). Bias in Cross-Sectional Analyses of Longitudinal Mediation. *Psychological Methods*, 12(1), 23–44. <https://doi.org/10.1037/1082-989X.12.1.23>
- Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human Abilities: Emotional Intelligence. *Annual Review of Psychology*, 59(1), 507–536. <https://doi.org/10.1146/annurev.psych.59.103006.093646>
- McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002).

- Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, 38(1), 115–142. <https://doi.org/10.1037/0012-1649.38.1.115>
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37(1), 1–10. <https://doi.org/10.1016/j.intell.2008.08.004>
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. <https://doi.org/10.1037/met0000144>
- Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working Memory Training Does Not Improve Performance on Measures of Intelligence or Other Measures of “Far Transfer”. *Perspectives on Psychological Science*, 11(4), 512–534. <https://doi.org/10.1177/1745691616635612>
- Meredith, W. (1993). Measurement Invariance, Factor Analysis and Factorial Invariance. *Psychometrika*, 58(4), 525–543. <https://doi.org/10.1007/BF02294825>
- Milton, D. E. M. (2012). On the ontological status of autism: the «double empathy problem». *Disability & Society*, 27(6), 883–887. <https://doi.org/10.1080/09687599.2012.710008>
- Molloy, M. F., Taxali, A., Angstadt, M., Greathouse, T., Toda-Thorne, K., McCurry, K. L., Weigard, A., Kardan, O., Burchell, L., Dziubinski, M., Choi, J., Vandersluis, M., Michael, C., Heitzeg, M. M., & Sripada, C. (2025). Regional, but not brain-wide, graph theoretic measures are robustly and reproducibly linked to general cognitive ability. *Cerebral Cortex*, 35(4), bhaf074. <https://doi.org/10.1093/cercor/bhaf074>
- Mutluer, T., Aslan Genç, H., Özcan Morey, A., Yapici Eser, H., Ertinmaz, B., Can, M., & Munir, K. (2022). Population-Based Psychiatric Comorbidity in Children and Adolescents With Autism Spectrum Disorder: A Meta-Analysis. *Frontiers in Psychiatry*, 13, 856208. <https://doi.org/10.3389/fpsy.2022.856208>
- Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and Unknowns. *American Psychologist*, 51(2), 77–101. <https://doi.org/10.1037/0003-066X.51.2.77>
- Neubauer, A. C., & Fink, A. (2009). Intelligence and neural efficiency. *Neu-*

- rosience & Biobehavioral Reviews*, 33(7), 1004–1023. <https://doi.org/10.1016/j.neubiorev.2009.04.001>
- Nye, C. D., Ma, J., & Wee, S. (2022). Cognitive Ability and Job Performance: Meta-analytic Evidence for the Validity of Narrow Cognitive Abilities. *Journal of Business and Psychology*, 37(6), 1119–1139. <https://doi.org/10.1007/s10869-022-09796-1>
- Oh, I.-S., Le, H., & Roth, P. L. (2023). Revisiting Sackett et al.'s (2022) rationale behind their recommendation against correcting for range restriction in concurrent validation studies. *Journal of Applied Psychology*, 108(8), 1300–1310. <https://doi.org/10.1037/apl0001078>
- Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, 3(2), 173–182. <https://doi.org/10.1038/s41562-018-0506-1>
- Owen, A. M., Hampshire, A., Grahm, J. A., Stenton, R., Dajani, S., Burns, A. S., Howard, R. J., & Ballard, C. G. (2010). Putting brain training to the test. *Nature*, 465(7299), 775–778. <https://doi.org/10.1038/nature09042>
- Penke, L., Maniega, S. M., Bastin, M. E., Valdés Hernández, M. C., Murray, C., Royle, N. A., Starr, J. M., Wardlaw, J. M., & Deary, I. J. (2012). Brain white matter tract integrity as a neural foundation for general intelligence. *Molecular Psychiatry*, 17(10), 1026–1030. <https://doi.org/10.1038/mp.2012.66>
- Pennington, B. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, 101(2), 385–413. <https://doi.org/10.1016/j.cognition.2006.04.008>
- Petrides, K. V., & Furnham, A. (2001). Trait emotional intelligence: psychometric investigation with reference to established trait taxonomies. *European Journal of Personality*, 15(6), 425–448. <https://doi.org/10.1002/per.416>
- Pietschnig, J., Penke, L., Wicherts, J. M., Zeiler, M., & Voracek, M. (2015). Meta-analysis of associations between human brain volume and intelligence differences: How strong are they and what do they mean? *Neuroscience & Biobehavioral Reviews*, 57, 411–432. <https://doi.org/10.1016/j.neubiorev.2015.09.017>
- Pietschnig, J., & Voracek, M. (2015). One Century of Global IQ Gains: A Formal Meta-Analysis of the Flynn Effect (1909–2013). *Perspectives on Psychological Science*, 10(3), 282–306. <https://doi.org/10.1177/174569>

1615577701

- Pietschnig, J., Voracek, M., & Formann, A. K. (2010). Mozart effect–Shmozart effect: A meta-analysis. *Intelligence*, 38(3), 314–323. <https://doi.org/10.1016/j.intell.2010.03.001>
- Pizzigallo, E., Cornoldi, C., Buono, S., Città, S., Viola, F., & Toffalini, E. (2023). The Intellectual Profile of Adults with Specific Learning Disabilities. *Journal of Intelligence*, 11(12), 223. <https://doi.org/10.3390/jintelligence11120223>
- Pletzer, J. L., & Abrahams, L. (2025). Personality and job performance: A review of trait models and recent trends. *Current Opinion in Psychology*, 65, 102054. <https://doi.org/10.1016/j.copsyc.2025.102054>
- Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: five special findings. *Molecular Psychiatry*, 20(1), 98–108. <https://doi.org/10.1038/mp.2014.105>
- Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2001). *Behavioral Genetics* (4^a ed.). Worth Publishers.
- Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics*, 19(3), 148–159. <https://doi.org/10.1038/nrg.2017.104>
- Pohl, K. M., Thompson, W. K., Adeli, E., & Linguraru, M. G. (A c. Di). (2019). *Adolescent Brain Cognitive Development Neurocognitive Prediction: First Challenge, ABCD-NP 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-31901-4>
- Pokropek, A., Marks, G. N., Borgonovi, F., Koc, P., & Greiff, S. (2022). General or specific abilities? Evidence from 33 countries participating in the PISA assessments. *Intelligence*, 92, 101653. <https://doi.org/10.1016/j.intell.2022.101653>
- Protzko, J. (2015). The environment in raising early intelligence: A meta-analysis of the fadeout effect. *Intelligence*, 53, 202–210. <https://doi.org/10.1016/j.intell.2015.10.006>
- Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: a decennial review of the global literature. *Education Economics*, 26(5), 445–458. <https://doi.org/10.1080/09645292.2018.1484426>
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for

- psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Quintin, E.-M., Jo, B., Hall, S. S., Bruno, J. L., Chromik, L. C., Raman, M. M., Lightbody, A. A., Martin, A., & Reiss, A. L. (2016). The cognitive developmental profile associated with fragile X syndrome: A longitudinal investigation of cognitive strengths and weaknesses through childhood and adolescence. *Development and Psychopathology*, 28(4pt2), 1457–1469. <https://doi.org/10.1017/S0954579415001200>
- Rauscher, F. H., Shaw, G. L., & Ky, C. N. (1993). Music and spatial task performance. *Nature*, 365(6447), 611–611. <https://doi.org/10.1038/365611a0>
- Raven, J. (2000). The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology*, 41(1), 1–48. <https://doi.org/10.1006/cogp.1999.0735>
- Ritchie, S. J., Bates, T. C., & Deary, I. J. (2015). Is education associated with improvements in general cognitive ability, or in specific skills? *Developmental Psychology*, 51(5), 573–582. <https://doi.org/10.1037/a0038981>
- Ritchie, S. J., Cox, S. R., Shen, X., Lombardo, M. V., Reus, L. M., Alloza, C., Harris, M. A., Alderson, H. L., Hunter, S., Neilson, E., Liewald, D. C. M., Auyeung, B., Whalley, H. C., Lawrie, S. M., Gale, C. R., Bastin, M. E., McIntosh, A. M., & Deary, I. J. (2018). Sex Differences in the Adult Human Brain: Evidence from 5216 UK Biobank Participants. *Cerebral Cortex*, 28(8), 2959–2975. <https://doi.org/10.1093/cercor/bhy109>
- Ritchie, S. J., & Tucker-Drob, E. M. (2018). How Much Does Education Improve Intelligence? A Meta-Analysis. *Psychological Science*, 29(8), 1358–1369. <https://doi.org/10.1177/0956797618774253>
- Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: a review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine*, 50(3), 353–366. <https://doi.org/10.1017/S0033291719003404>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137–150. <https://doi.org/10.1037/met0000045>
- Rohrer, J. M. (2018). Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods*

- and Practices in Psychological Science, 1(1), 27–42. <https://doi.org/10.1177/2515245917745629>
- Rong, Y., Yang, C.-J., Jin, Y., & Wang, Y. (2021). Prevalence of attention-deficit/hyperactivity disorder in individuals with autism spectrum disorder: A meta-analysis. *Research in Autism Spectrum Disorders*, 83, 101759. <https://doi.org/10.1016/j.rasd.2021.101759>
- Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, 53, 118–137. <https://doi.org/10.1016/j.intell.2015.09.002>
- Russell, G., Mandy, W., Elliott, D., White, R., Pittwood, T., & Ford, T. (2019). Selection bias on intellectual ability in autism research: a cross-sectional review and meta-analysis. *Molecular Autism*, 10(1), 9. <https://doi.org/10.1186/s13229-019-0260-x>
- Sackett, P. R., Demeke, S., Bazian, I. M., Griebie, A. M., Priest, R., & Kuncel, N. R. (2024). A contemporary look at the relationship between general cognitive ability and job performance. *Journal of Applied Psychology*, 109(5), 687–713. <https://doi.org/10.1037/apl0001159>
- Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*, 107(11), 2040–2068. <https://doi.org/10.1037/apl0000994>
- Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2023). Revisiting the design of selection systems in light of new findings regarding the validity of widely used predictors. *Industrial and Organizational Psychology*, 16(3), 283–300. <https://doi.org/10.1017/iop.2023.24>
- Sala, G., & Gobet, F. (2017). Does Far Transfer Exist? Negative Evidence From Chess, Music, and Working Memory Training. *Current Directions in Psychological Science*, 26(6), 515–520. <https://doi.org/10.1177/0963721417712760>
- Sala, G., & Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory & Cognition*, 48(8), 1429–1441. <https://doi.org/10.3758/s13421-020-01060-2>
- Sala, G., Tatlidil, K. S., & Gobet, F. (2018). Video game training does not enhance cognitive ability: A comprehensive meta-analytic investigation. *Psychological Bulletin*, 144(2), 111–139. <https://doi.org/10.1037/bul0000139>

- Salovey, P., & Mayer, J. D. (1990). Emotional Intelligence. *Imagination, Cognition and Personality*, 9(3), 185–211. <https://doi.org/10.2190/DUGG-P24E-52WK-6CDG>
- Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., Stevens, C., Reichert, J., Mulhern, M. S., Artomov, M., Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., ... Buxbaum, J. D. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*, 180(3), 568–584.e23. <https://doi.org/10.1016/j.cell.2019.12.036>
- Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., Nagel, M., Awasthi, S., Barr, P. B., Coleman, J. R. I., Grasby, K. L., Hammerschlag, A. R., Kaminski, J. A., Karlsson, R., Krapohl, E., Lam, M., Nygaard, M., Reynolds, C. A., Trampush, J. W., ... Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, 50(7), 912–919. <https://doi.org/10.1038/s41588-018-0152-6>
- Savi, A. O., Marsman, M., van der Maas, H. L. J., & Maris, G. K. J. (2019). The Wiring of Intelligence. *Perspectives on Psychological Science*, 14(6), 1034–1061. <https://doi.org/10.1177/1745691619866447>
- Schmidt, F. L., & Hunter, J. (2004). General Mental Ability in the World of Work: Occupational Attainment and Job Performance. *Journal of Personality and Social Psychology*, 86(1), 162–173. <https://doi.org/10.1037/0022-3514.86.1.162>
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262–274. <https://doi.org/10.1037/0033-2909.124.2.262>
- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, 126(5), 1763–1768. <https://doi.org/10.1213/ANE.0000000000002864>
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40*. High/Scope Press.
- Serpico, D. (2022). *L'intelligenza tra natura e cultura*. Rosenberg & Sellier.
- Shayer, M., & Ginsburg, D. (2009). Thirty years on – a large anti-Flynn

- effect/ (II): 13- and 14-year-olds. Piagetian tests of formal operations norms 1976–2006/7. *British Journal of Educational Psychology*, 79(3), 409–418. <https://doi.org/10.1348/978185408X383123>
- Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence That Dyslexia May Represent the Lower Tail of a Normal Distribution of Reading Ability. *New England Journal of Medicine*, 326(3), 145–150. <https://doi.org/10.1056/NEJM199201163260301>
- Shewach, O. R., Sackett, P. R., & Quint, S. (2019). Stereotype threat effects in settings with features likely versus unlikely in operational test settings: A meta-analysis. *Journal of Applied Psychology*, 104(12), 1514–1534. <https://doi.org/10.1037/apl0000420>
- Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do «Brain-Training» Programs Work? *Psychological Science in the Public Interest*, 17(3), 103–186. <https://doi.org/10.1177/1529100616661983>
- Sjöwall, D., Roth, L., Lindqvist, S., & Thorell, L. B. (2013). Multiple deficits in ADHD: executive dysfunction, delay aversion, reaction time variability, and emotional deficits. *Journal of Child Psychology and Psychiatry*, 54(6), 619–627. <https://doi.org/10.1111/jcpp.12006>
- Skeels, H. M. (1966). Adult Status of Children with Contrasting Early Life Experiences: A Follow-Up Study. *Monographs of the Society for Research in Child Development*, 31(3), 1. <https://doi.org/10.2307/1165791>
- Snowling, M. J., Nash, H. M., Gooch, D. C., Hayiou-Thomas, M. E., & Hulme, C. (2019). Developmental Outcomes for Children at High Risk of Dyslexia and Children With Developmental Language Disorder. *Child Development*, 90(5), e548–e564. <https://doi.org/10.1111/cdev.13216>
- Snyderman, M., & Herrnstein, R. J. (1983). Intelligence tests and the Immigration Act of 1924. *American Psychologist*, 38(9), 986–995. <https://doi.org/10.1037/0003-066X.38.9.986>
- Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, 15(2), 201. <https://doi.org/10.2307/1412107>
- Spearman, C. (1927). *The Abilities of Man: Their Nature and Measurement*. Macmillan.
- Stanovich, K. E. (1986). Matthew Effects in Reading: Some Consequen-

- ces of Individual Differences in the Acquisition of Literacy. *Reading Research Quarterly*, 21(4), 360–407. <https://doi.org/10.1598/RRQ.21.4.1>
- Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on K–12 students' academic achievement: Findings of two second-order meta-analyses. *Review of Educational Research*, 86(4), 849–899. <https://doi.org/10.3102/0034654316675417>
- Stephenson, K. G., Beck, J. S., South, M., Norris, M., & Butter, E. (2021). Validity of the WISC-V in Youth with Autism Spectrum Disorder: Factor Structure and Measurement Invariance. *Journal of Clinical Child & Adolescent Psychology*, 50(5), 669–681. <https://doi.org/10.1080/15374416.2020.1846543>
- Stern, W. (1912). *Die psychologischen Methoden der Intelligenzprüfung*. Barth.
- Stern, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. *The Lancet Neurology*, 11(11), 1006–1012. [https://doi.org/10.1016/S1474-4422\(12\)70191-6](https://doi.org/10.1016/S1474-4422(12)70191-6)
- Sternberg, R. J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence*. Cambridge University Press.
- Sternberg, R. J. (A. c. Di). (2002). *Why Smart People Can Be So Stupid*. Yale University Press.
- Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, 35(5), 401–426. <https://doi.org/10.1016/j.intell.2006.09.004>
- Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect?: A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, 32(4), 349–362. <https://doi.org/10.1016/j.intell.2004.06.004>
- Szucs, D., & Ioannidis, J. Pa. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage*, 221, 117164. <https://doi.org/10.1016/j.neuroimage.2020.117164>
- Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PLOS ONE*, 12(10), e0186773. <https://doi.org/10.1371/journal.pone.0186773>
- te Nijenhuis, J., & van der Flier, H. (2013). Is the Flynn effect on g?: A meta-analysis. *Intelligence*, 41(6), 802–807. <https://doi.org/10.1016/j.intell.2013.05.004>

intell.2013.03.001

- Terman, L. M. (1916). *The Measurement of Intelligence*. Houghton Mifflin.
- Thiele, J. A., Faskowitz, J., Sporns, O., & Hilger, K. (2022). Multitask brain network reconfiguration is inversely associated with human intelligence. *Cerebral Cortex*, 32(19), 4172–4182. <https://doi.org/10.1093/cercor/bhab473>
- Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Arousal, Mood, and The Mozart Effect. *Psychological Science*, 12(3), 248–251. <https://doi.org/10.1111/1467-9280.00345>
- Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology*, 8(3), 271–281. <https://doi.org/10.1111/j.2044-8295.1916.tb00133.x>
- Thurstone, L. L. (1938). *Primary Mental Abilities*. University of Chicago Press.
- Toffalini, E., Buono, S., & Cornoldi, C. (2022). The structure, profile, and diagnostic significance of intelligence in children with ADHD are impressively similar to those of children with a specific learning disorder. *Research in Developmental Disabilities*, 129, 104306. <https://doi.org/10.1016/j.ridd.2022.104306>
- Toffalini, E., & Cornoldi, C. (2026). *Oltre il QI globale nell'IDS-2: Proposta di due nuovi indici per l'abilità intellettuale*. Presentazione al convegno ICAP 2026, Firenze. <https://enricotoffalini.github.io/slides/Toffalini-ICAP2026-IT.pdf>
- Toffalini, E., & Feraco, T. (2026). Musicians, Memory, and General Cognitive Ability: A Commentary on Grassi et al. (2025). *Advances in Methods and Practices in Psychological Science*, 9(3), 25152459261452948. <https://doi.org/10.1177/25152459261452948>
- Toffalini, E., Giofrè, D., & Cornoldi, C. (2017). Strengths and Weaknesses in the Intellectual Profile of Different Subtypes of Specific Learning Disorder: A Study on 1,049 Diagnosed Children. *Clinical Psychological Science*, 5(2), 402–409. <https://doi.org/10.1177/2167702616672038>
- Tucker-Drob, E. M., & Bates, T. C. (2016). Large Cross-National Differences in Gene $\times$ Socioeconomic Status Interaction on Intelligence. *Psychological Science*, 27(2), 138–149. <https://doi.org/10.1177/0956797615612727>
- Turkheimer, E. (2000). Three Laws of Behavior Genetics and What They

- Mean. *Current Directions in Psychological Science*, 9(5), 160–164. <https://doi.org/10.1111/1467-8721.00084>
- Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic Status Modifies Heritability of IQ in Young Children. *Psychological Science*, 14(6), 623–628. https://doi.org/10.1046/j.0956-7976.2003.psci_1475.x
- van Bergen, E., de Zeeuw, E. L., Hart, S. A., Boomsma, D. I., de Geus, E. J. C., & Kan, K.-J. (2025). Co-Occurrence and Causality Among ADHD, Dyslexia, and Dyscalculia. *Psychological Science*, 36(3), 204–217. <https://doi.org/10.1177/09567976241293999>
- van den Heuvel, M. P., Stam, C. J., Kahn, R. S., & Hulshoff Pol, H. E. (2009). Efficiency of Functional Brain Networks and Intellectual Performance. *Journal of Neuroscience*, 29(23), 7619–7624. <https://doi.org/10.1523/JNEUROSCI.1443-09.2009>
- Van der Hallen, R., Evers, K., Brewaeys, K., Van den Noortgate, W., & Wagemans, J. (2015). Global processing takes time: A meta-analysis on local–global visual processing in ASD. *Psychological Bulletin*, 141(3), 549–573. <https://doi.org/10.1037/bul0000004>
- van der Linden, D., Pekaar, K. A., Bakker, A. B., Schermer, J. A., Vernon, P. A., Dunkel, C. S., & Petrides, K. V. (2017). Overlap between the general factor of personality and emotional intelligence: A meta-analysis. *Psychological Bulletin*, 143(1), 36–52. <https://doi.org/10.1037/bul0000078>
- van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, 113(4), 842–861. <https://doi.org/10.1037/0033-295X.113.4.842>
- Van Hootegem, A., Røgeberg, O., Bratsberg, B., & Lyngstad, T. H. (2023). Correlation between cognitive ability and educational attainment weakens over birth cohorts. *Scientific Reports*, 13(1), 17747. <https://doi.org/10.1038/s41598-023-44605-6>
- van IJzendoorn, M. H., Juffer, F., & Poelhuis, C. W. K. (2005). Adoption and Cognitive Development: A Meta-Analytic Comparison of Adopted and Nonadopted Children's IQ and School Performance. *Psychological Bulletin*, 131(2), 301–316. <https://doi.org/10.1037/0033-2909.131.2.301>

- van IJzendoorn, M. H., Luijk, M. P. C. M., & Juffer, F. (2008). IQ of Children Growing Up in Children's Homes: A Meta-Analysis on IQ Delays in Orphanages. *Merrill-Palmer Quarterly*, 54(3), 341–366. <https://doi.org/10.1353/mpq.0.0002>
- Vernon, P. E. (1950). *The Structure of Human Abilities*. Methuen.
- Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006). Beyond g: Putting multiple intelligences theory to the test. *Intelligence*, 34(5), 487–502. <https://doi.org/10.1016/j.intell.2006.02.004>
- Warne, R. T. (2020). *In the Know: Debunking 35 Myths about Human Intelligence*. Cambridge University Press. <https://doi.org/10.1017/9781108593298>
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, 70(2), 129–133. <https://doi.org/10.1080/00031305.2016.1154108>
- Waterhouse, L. (2006). Multiple Intelligences, the Mozart Effect, and Emotional Intelligence: A Critical Review. *Educational Psychologist*, 41(4), 207–225. https://doi.org/10.1207/s15326985ep4104_1
- Watkins, M. W. (2006). Orthogonal higher order structure of the Wechsler Intelligence Scale for Children—Fourth edition. *Psychological Assessment*, 18(1), 123–125. <https://doi.org/10.1037/1040-3590.18.1.123>
- Watkins, M. W., & Smith, L. G. (2013). Long-term stability of the Wechsler Intelligence Scale for Children—Fourth Edition. *Psychological Assessment*, 25(2), 477–483. <https://doi.org/10.1037/a0031653>
- Wechsler, D. (1939). *The Measurement of Adult Intelligence*. Williams & Wilkins.
- Wechsler, D. (1958). *The Measurement and Appraisal of Adult Intelligence* (4^a ed.). Williams & Wilkins.
- Wechsler, D. (2008). *WAIS-IV Technical and Interpretive Manual*. Pearson.
- Whalley, L. J., & Deary, I. J. (2001). Longitudinal cohort study of childhood IQ and survival up to age 76. *BMJ*, 322(7290), 819. <https://doi.org/10.1136/bmj.322.7290.819>
- Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., van Baal, G. C. M., Boomsma, D. I., & Span, M. M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence*, 32(5), 509–537. <https://doi.org/10.1016/j.intell.2004.07.002>
- Wilmot, M. P., & Ones, D. S. (2019). A century of research on conscientiousness

- ness at work. *Proceedings of the National Academy of Sciences*, 116(46), 23004–23010. <https://doi.org/10.1073/pnas.1908430116>
- Wissler, C. (1901). The correlation of mental and physical tests. *The Psychological Review: Monograph Supplements*, 3(6), i–62. <https://doi.org/10.1037/h0092995>
- Wolfram, T. (2023). (Not just) Intelligence stratifies the occupational hierarchy: Ranking 360 professions by IQ and non-cognitive traits. *Intelligence*, 98, 101755. <https://doi.org/10.1016/j.intell.2023.101755>
- Woodrow, H. (1921). Intelligence and its measurement: A symposium–XI. *Journal of Educational Psychology*, 12(4), 207–210. <https://doi.org/10.1037/h0067806>
- World Health Organization. (2019). *International Classification of Diseases, 11th Revision (ICD-11)*. <https://icd.who.int/>
- Yerkes, R. M. (A c. Di). (1921). *Psychological Examining in the United States Army* (Vol. 15). Government Printing Office.
- Zajonc, R. B. (1965). Social Facilitation: A solution is suggested for an old unresolved social psychological problem. *Science*, 149(3681), 269–274. <https://doi.org/10.1126/science.149.3681.269>
- Zanon, F., & Priftis, K. (2025). Psychologists Must Be against the Use of Tests without Accessible Norms. *Giornale italiano di psicologia*, 2/2025. <https://doi.org/10.1421/116921>