



UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia Applicata

Corso di laurea magistrale in Psicologia sociale, del lavoro e della comunicazione

Elaborato finale

Un'applicazione del Partial Credit Model per lo studio del funzionamento delle rating scale.

An application of the Partial Credit Model to the study of the functioning of rating scales.

Relatrice

Prof.ssa Colledani Daiana

Laureanda: Ruscica Ketty

Matricola: 3050383

Anno Accademico: 2022-2023

INDICE

INTRODUZIONE	1
1. I TRATTI DI PERSONALITA'	3
1.1 DEFINIZIONI	3
1.2 APPROCCIO LESSICALE E TEORIE CONTEMPORANEE	4
1.3 OUTCOMES DELLA PERSONALITA'	6
2. L'ANSIA	9
2.1 DEFINIZIONI DEL COSTRUTTO	9
2.2 CENNI STORICI	11
2.3 METODI DI MISURAZIONE DEL COSTRUTTO	16
2.3.1 STATE-TRAIT ANXIETY INVENTORY (STAI)	16
2.3.2 BECK ANXIETY INVENTORY (BAI)	18
2.3.3 ENDLER MULTIDIMENSIONAL ANXIETY SCALES (EMAS)	19
3. LA MISURAZIONE IN PSICOLOGIA	21
3.1 I TEST PSICOLOGICI	22
3.2 CARATTERISTICHE METRICHE	23
3.3 VANTAGGI E LIMITI DELLA MISURAZIONE	26
3.4 ITEM RESPONSE THEORY	27
3.4.1 MODELLO DI RASCH	29
3.4.2 ESTENSIONI DEL MODELLO DI RASCH	30
4. LA RICERCA	35
4.1 IPOTESI DI RICERCA	35
4.2 METODO	37
4.2.1 PARTECIPANTI	37
4.2.2 PROCEDURA E SOMMINISTRAZIONE	38
4.2.3 STRATEGIA D'ANALISI	39
4.3 RISULTATI	42
4.4 INTEPRETAZIONE E DISCUSSIONE DELLO STUDIO	52
4.4.1 LIMITI	55
4.4.2 SVILUPPI FUTURI	55
4.5 CONCLUSIONI	56
BIBLIOGRAFIA	59

INTRODUZIONE

Il presente lavoro di ricerca mostra come il funzionamento delle *rating scales* possa essere efficacemente verificato attraverso l'applicazione del *Partial Credit Model* (PCM; Masters, 1982).

Nello sviluppo di test e questionari la letteratura solitamente raccomanda di prevedere l'uso di più categorie di risposta agli item (Bollen, 1981; Maydeu-Olivares, 2017; Taherdoost, 2019) in quanto si ritiene che questo possa massimizzare l'informazione raccolta limitando il numero di domande da porre ai rispondenti. Tuttavia, tale strategia in alcune situazioni può rivelarsi non ottimale. Infatti, la presenza di troppe opzioni di risposta può creare confusione nei rispondenti e condurre ad una svantaggiosa condizione definita “*threshold disordering*”.

Il presente lavoro di ricerca intende mostrare l'importanza di non dare per scontato il funzionamento delle *rating scales* e l'utilità di verificarlo empiricamente. L'elaborato mostra come il modello di Rasch consenta di operare tale verifica. In particolare, il funzionamento della scala di risposta della scala di tratto del questionario Trait-State Anxiety Inventory (STAI; Pedrabissi & Santinello, 1989) è stato verificato attraverso l'applicazione del modello PCM. Il modello è stato applicato ai dati raccolti attraverso la somministrazione del questionario in due diverse versioni, una con scala di risposta standard a 4 categorie ed una allungata a 8 categorie, sviluppata appositamente ai fini di questo lavoro di ricerca.

Per avere uno sguardo chiaro sui costrutti considerati e sulle metodologie utilizzate, si è ritenuto opportuno partire da una rassegna della letteratura volta a dare un fondamento metodologico a tutto il lavoro.

Nello specifico, nel primo capitolo vengono definiti i tratti di personalità e prese in considerazione alcune tra le teorie più conosciute che si sono susseguite nel tempo, considerando il contributo di Galton, Allport, Cattell, ed Eysenck.

A seguire, viene preso in esame uno tra i costrutti più interessanti del campo psicologico odierno, ovvero l'ansia. Il costrutto è presentato nel suo aspetto generale, attraverso alcune definizioni, e le caratteristiche di alcuni particolari stati ansiosi sono delineate facendo riferimento alla specifica sintomatologia, ai particolari fattori di rischio e con accenni di tipo teorico. Infine, vengono descritti alcuni strumenti per la valutazione dell'ansia, tra cui lo STAI che costituisce lo strumento oggetto del lavoro di ricerca.

Ampio spazio è dedicato alla descrizione del concetto di misurazione in psicologia, ponendo attenzione ad alcune proprietà cruciali dei test psicologici quali validità e affidabilità. Un capitolo è dedicato all'Item Response Theory e ad alcune estensioni del modello di Rasch, importanti per comprendere le analisi che sono state effettuate nel lavoro di ricerca.

Infine, l'ultimo capitolo è dedicato alla presentazione del lavoro di ricerca, di cui vengono dettagliatamente descritti obiettivi, metodi e risultati. Ampio spazio è dedicato anche alla discussione degli stessi e alla definizione di possibili futuri sviluppi di ricerca.

1. I TRATTI DI PERSONALITA'

1.1 DEFINIZIONI

Le persone differiscono le une dalle altre esternamente ed internamente. È possibile vedere ad occhio nudo alcune differenze fisiche tra gli individui, ma ciò che importa particolarmente l'ambito psicologico, sono quelle caratteristiche che coinvolgono il modo di comportarsi, di pensare e di sentire (Ashton, 2023).

Gli individui possono essere tutti accumulati da alcuni processi biologici e psicologici come lo sviluppo, la percezione, l'apprendimento e la memoria solo per citarne alcuni, ma ciascuno di noi ha un insieme distintivo di convinzioni, motivazioni, emozioni e tratti di personalità che ci rendono unici (Nolen-Hoeksema et al., 2014).

Queste ultime sono caratteristiche psicologiche - differenze di personalità - che riteniamo importanti nel definire chi è una persona.

Allport, importante psicologo statunitense definì la personalità come *“un'organizzazione dinamica, interna alla persona di sistemi psicofisiologici che determinano il suo adattamento all'ambiente”* (Allport, p.48, 1937).

Una definizione più articolata presente in letteratura invece descrive la personalità come quell'insieme di *“differenze tra gli individui in una tipica tendenza a comportarsi, pensare, o sentire in alcuni modi concettualmente correlati, attraverso una varietà di situazioni rilevanti e in un periodo di tempo abbastanza lungo* (Ashton, pp.33-36, 2023). Questa definizione sottolinea come le differenze individuali acquistino un significato solo nella misura in cui vi sia un confronto con altri, e come questa sia un'inclinazione o una predisposizione a comportarsi e pensare in un dato modo. Inoltre, consente di notare che la manifestazione dei tratti di personalità avviene attraverso

comportamenti, pensieri e sentimenti che hanno tra di loro qualche elemento comune, e come questo si manifesti in una varietà di contesti e nel lungo periodo (Ashton, 2023).

1.2 APPROCCIO LESSICALE E TEORIE CONTEMPORANEE

Analizzando in modo più sistematico il tema della personalità, è possibile seguire una sequenza temporale, trovando nell'approccio lessicale il contributo probabilmente più importante. L'approccio lessicale nasce dall'idea di utilizzare il dizionario come fonte per individuare, descrivere e classificare le caratteristiche della personalità degli individui. Più precisamente, tale approccio, sostiene che le persone nel descrivere gli altri tendono a selezionare ed usare quelle parole che ritengono rappresentare un determinato tipo di tratto, in relazione anche a quanto (livelli alti o bassi) i tratti siano presenti nelle persone. A lungo andare, in ogni lingua, dovrebbero quindi stabilirsi dei set di parole frequentemente usate per descrivere i tratti di personalità più importanti. Questi termini nel tempo dovrebbero costituire un elenco completo di caratteristiche salienti dei tratti di personalità (Ashton, 2023).

Francis Galton fu il primo studioso ad utilizzare un dizionario come mezzo per individuare un set di termini descrittivi della personalità, sottolineando come alcuni di questi condividessero tra loro aspetti significativi. Ovvero, le varie parole si sovrapponevano in parte l'una con l'altra, e in parte mantenevano un proprio significato unico (Galton, 1884).

Successivamente Allport e Odbert (1936) cercarono di creare un inventario di questo lessico. Questo fu reso possibile attraverso l'analisi dei termini presenti nel *Webster's Unabridged Dictionary*. Gli autori giunsero ad un elenco di quasi 18.000 parole che potevano essere utilizzate per descrivere le persone. Di queste parole, 4500 potevano

essere attribuite ad alcuni tratti di personalità. Gli autori cercarono di raggruppare i termini individuati in tre categorie: *tratti cardinali* cioè tratti principali che sono dominanti/salienti nella vita di un individuo (es. stabilità emotiva, narcisismo), *tratti centrali* disposizioni significative in una gamma limitata di situazioni (es. timidezza o onestà), e *disposizioni secondarie* che sono meno evidenti e si manifestano solo in circostanze particolari (es. una persona generalmente accomodante, potrebbe diventare irascibile se sottoposta a molta pressione) (Novikova, 2013). Il loro lavoro è stato poi integrato da diversi autori, come Cattell negli anni '40 e Norman negli anni '60 (Goldberg, 1993).

Cattell (1947) decise, dato l'ampio numero di variabili considerate, di raggruppare tra di loro sinonimi e termini semanticamente simili, arrivando ad un gruppo di 35 variabili bipolari. Cercò poi di indentificare le principali dimensioni della personalità rappresentate dai termini considerati. Per fare ciò utilizzò la tecnica dell'analisi fattoriale¹, dalla quale emersero 12 fattori di personalità. Un importante contributo di questo autore fu lo sviluppo del *Sixteen Personality Factors Questionnaire* (16PF). Studi successivi, tuttavia, analizzarono i fattori emersi dal lavoro di Cattell e mostrarono che solo cinque di essi erano replicabili: Estroversione, Piacevolezza, Coscienziosità, Stabilità emotiva vs Nevroticismo, Cultura o come definito in seguito Apertura Mentale (Tupes, 1961). Tali dimensioni furono riconosciute come centrali anche da altri studiosi e successivamente diedero origine a quelli che oggi definiamo i Big Five, ovvero i grandi cinque fattori di personalità (Goldberg, 1981).

¹ Analisi fattoriale: metodo statistico che si propone di ricondurre le correlazioni di un elevato numero di variabili all'azione di un numero più contenuto di fattori ipotizzati alla base della variazione delle variabili (Galimberti, 2005)

Il modello dei Big Five (anche detto Five Factor Model) diviene un punto di convergenza tra più punti di vista. Una teoria contemporanea, sviluppata da McCrae e Costa, che riconosce cinque dimensioni centrali per descrivere le differenze individuali: l'apertura all'esperienza (O), Coscienziosità (C), Estroversione (E), Piacevolezza (A) e Nevroticismo (N). Spesso questi fattori vengono indicati con l'acronimo OCEAN, e corrispondono in parte ai fattori globali di Cattell (McCrae, 2003). L'adattamento italiano è riconducibile a Caprara (2005), il quale con i colleghi realizzò il Big Five Questionnaire (BFQ). È interessante notare che una recente revisione del modello Big Five, chiamata modello HEXACO, aggiunge una sesta dimensione ai cinque grandi fattori: Onestà – Umiltà (Ashton & Lee, 2007).

Altre fondamentali teorie furono quelle di Eysenck (1960) che individuò un modello gerarchico di personalità che includeva tre fattori: Estroversione, Nevroticismo, Psicoticismo misurabili con *l'Eysenck Personality Inventory (EPI)* e *l'Eysenck Personality Questionnaire (EPQ)*.

A seguito di questa panoramica teorica sui vari studi della personalità possiamo dedurre l'importanza dei tratti di personalità nel comprendere il comportamento umano. Al tempo stesso però emerge che vi sia ancora un dibattito aperto sul numero esatto dei fattori che la compongono. Questo suggerisce, a mio avviso, l'importanza di incrementare, per la qualità dell'esame diagnostico, l'ausilio di strumenti che utilizzano sofisticate tecniche matematiche ed un'attenta analisi del lessico quotidiano.

1.3 OUTCOMES DELLA PERSONALITA'

Un aspetto principale da tenere in considerazione relativo alla personalità è quello dell'influenza che essa esercita su alcuni outcomes nella vita delle persone.

All'interno del libro "*Individual Differences and Personality*" (Ashton, 2023) vengono descritte alcune delle aree che possono essere influenzate dalla personalità di un individuo. Tra queste si trovano: la soddisfazione nelle relazioni e nel matrimonio, fortemente correlata ad alti livelli di Piacevolezza, Stabilità Emotiva e Coscienziosità (Malouff et al., 2010). Inoltre, si evidenzia l'effetto sulle scelte amicali e delle relazioni tra pari, che sembrano avvenire sulla base della somiglianza percepita rispetto a sé in termini di livelli di onestà (umiltà) e apertura all'esperienza (Lee et al., 2009). Un altro dominio interessato è quello dei comportamenti relativi alla salute: la longevità risulta essere associata ad alti livelli di coscienziosità e stabilità emotiva (Jokela et al., 2020), mentre bassi livelli sembrano essere correlati all'abuso di alcol o droghe illecite e fumo (Roberts & Bogg, 2004). Inoltre, un elevato rischio di malattie cardiache è associato a tratti di personalità di ostilità (o rabbia) (Williams, 2010). Inoltre, risultano essere associati alla personalità anche il rendimento scolastico, predetto positivamente dalla Coscienziosità e in particolare dalla motivazione al successo (Poropat, 2011) e il rendimento lavorativo modestamente correlato sia alla Coscienziosità (Donovan & Hurtz, 2000) e al fattore HEXACO Onestà-Umiltà (Pletzer et al., 2019). La personalità risulta essere predittore di attività criminali in coloro che hanno scarso controllo degli impulsi, ma anche tratti di volontà di sfruttare e manipolare gli altri (Gottfredson & Hirschi, 1990). Infine, le valutazioni di soddisfazione per la propria vita complessiva sono associate ai tratti del Big Five di Estroversione e Stabilità emotiva, come l'autostima, l'allegria e la mancanza di tendenze ansiose o depressive. Esiste dunque nelle persone una differenza nella tendenza a provare emozioni positive o negative, che scaturisce poi in una diversa percezione di soddisfazione (Furr & Funder, 1998). Tutte

queste aree coinvolte mettono dunque in luce quanto sia importante porre attenzione alla personalità e riconoscerla.

2. L'ANSIA

2.1 DEFINIZIONI DEL COSTRUTTO

Fra i tratti di personalità nessuno è popolare, specie oggi giorno, quanto l'ansia, che viene quasi quotidianamente nominata per una moltitudine di situazioni diverse.

L'ansia, infatti, indica una pluralità di costrutti che nel tempo sono andati via via evolvendosi (Sanavio & Sica, 1997).

Il termine deriva dal tardo latino *anxia*, derivato da *anxius* che significa ansioso, e nel vocabolario italiano viene definito come “uno stato di agitazione, di forte apprensione dovuto a timore, incertezza, attesa di qualcosa” (Treccani, 2023).

Nel lessico psicodiagnostico, per ansia si intende: “anticipazione apprensiva di un pericolo o di un evento negativo futuro, accompagnata da sentimenti di disforia e/o da sintomi fisici di tensione. Gli elementi esposti al rischio possono appartenere sia al mondo interno che a quello esterno” (APA, 1995).

Nel *Diagnostic and Statistical Manual of Mental Disorders Fifth edition* (DSM-5) i disturbi d'ansia si differenziano per la tipologia di oggetti o situazioni che li provocano, mettendo in luce come sia fondamentale nella diagnosi considerare attentamente la situazione (APA, 2014).

I disturbi d'ansia vengono suddivisi in ordine evolutivo, secondo l'età tipica di esordio:

- Disturbo d'ansia di separazione
- Mutismo selettivo
- Fobia specifica (animale, ambiente naturale, sangue-iniezione-ferite)
- Disturbo d'ansia sociale (Fobia) legata o meno alla performance

- Disturbo di panico
- Agorafobia
- Disturbo d'ansia generalizzata
- Disturbo d'ansia indotto da sostanze/farmaci (con esordio durante intossicazione, astinenza, dopo l'uso)
- Disturbo d'ansia dovuto a un'altra condizione medica
- Disturbo d'ansia con altra specificazione
- Disturbo d'ansia senza specificazione

È inoltre importante sottolineare che, nonostante spesso vengano utilizzati come sinonimi ansia e paura sono costrutti diversi. La prima è una risposta orientata al futuro, di lunga durata, focalizzata su una minaccia diffusa ed associata a tensione muscolare o vigilanza in vista del pericolo futuro, mentre la seconda è una risposta orientata al presente, di breve durata, appropriata ad una minaccia identificabile e specifica ed associata a picchi di attivazione automatica necessaria alla lotta o alla fuga (APA, 2023).

Occorre comunque tenere a mente che l'ansia non è sempre connotata di accezione negativa, ma che viene sperimentata dalle persone come una normale risposta allo stress, che aiuta a far fronte a situazioni/ambienti pericolosi o a rispondere ad eventuali minacce, aumentando ad esempio la percezione della probabilità o gravità di potenziali eventi negativi. Tale range di normalità però viene meno nel momento in cui gli individui sperimentano l'ansia ad elevati livelli, o in risposta ad eventi presumibilmente non stressanti, o protratti nel tempo anche quando l'evento scatenante sia concluso. Ciò porta ad avere un impatto negativo sulla qualità della vita e ad incidere col normale funzionamento della persona (APA, 2016). Questi aspetti caratterizzano quella che

viene definita come ansia morbosa, la quale ha un effetto paralizzante sul pensiero e sull'azione degli individui (Bennet, 1946).

I sintomi dell'ansia possono essere divisi in soggettivi e oggettivi. Per soggettivi intendiamo quei sintomi che sono vissuti come esperienze psicologiche, ad esempio, la paura, il terrore, le preoccupazioni emotive, la paura di morire, i pensieri ossessivi riguardanti la sicurezza, ecc. Quelli oggettivi invece sono anche detti somatici ed includono nausea, palpitazioni, dolore addominale, vertigini, vampate di calore, affanno, mal di testa, o altre esperienze confondibili con altre malattie fisiche. Il quadro clinico, dunque, nello stato d'ansia è una combinazione di caratteristiche psicologiche e somatiche (Emilien et al., 2002).

Tra i fattori di rischio e di vulnerabilità che predispongono l'esordio dei disturbi d'ansia troviamo fattori biologici, genetici, relazionali e ambientali. Questi possono essere di natura endogena (geni, temperamento) o ambientale (traumi, negligenza nelle cure da parte del caregiver) (Tambelli, 2017).

2.2 CENNI STORICI

Il costrutto dell'ansia, nel corso degli anni, è stato studiato attraverso diversi approcci e prospettive. Le teorie dell'ansia possono, quindi, essere suddivise in psicoanalitiche, di apprendimento/comportamentali, fisiologiche, fenomenologiche/esistenziali e cognitive (Strongman, 1995).

La teoria psicoanalitica più importante è quella di Freud che considera l'ansia come un fenomeno quotidiano, di significativa importanza che spiegherebbe le nevrosi. Freud sviluppò due diverse teorie. Nella prima considerava l'ansia come una libido trasformata attraverso la rimozione. Più precisamente, nel momento in cui viene

represso un atto istintivo (motivato sessualmente), si genererebbe uno stato di ansia, la quale sfocerebbe poi in qualche sintomo con l'intento di impedire lo sviluppo di ulteriore ansia (Freud, 1917). Nella seconda teoria, invece, l'autore inverte il legame ansia-rimozione considerando l'esperienza ansiosa come causa della rimozione. L'ansia diventa dunque un segnale dell'ego su un pericolo reale o potenziale. La minaccia provoca ansia che conduce alla repressione per far uscire la persona dal pericolo (Freud, 1926). Queste teorie sembrano evidenziare tre aspetti dell'ansia: una sensazione spiacevole, un processo di scarica e la percezione dei fenomeni coinvolti in questa scarica. Partendo da tali concetti distingue due tipi di ansia, al quale se ne può aggiungere un terzo (Ghadirian, 1981):

- ansia oggettiva: origina da una fonte esterna, il pericolo è uno stimolo oggettivamente pericoloso (Freud, 1924) e viene percepito come minaccia che suscita uno stato di apprensione proporzionale alla grandezza del pericolo. Da questa condizione si può arrivare allo stato fobico verso un oggetto specifico.
- ansia nevrotica: deriva da conflitti interni, tra Es ed Io più precisamente (Endler & Kocovski, 2001), cioè desideri repressi e conflitti inconsci.
- ansia morale: espressione di vergogna o senso di colpa che nasce dal conflitto tra Es ed Super-io (Endler & Kocovski, 2001)

Sullivan, successivamente a Freud, considera l'ansia un fenomeno sociale, interpersonale piuttosto che intrapsichico (Sullivan, 1953).

La prospettiva comportamentale, invece, vede come importanti contributi i lavori di Pavlov e Watson e si basa sull'assunto che gli individui imparano a evitare gli stimoli nocivi attraverso alcuni meccanismi di mediazione. Questi normalmente sono la paura o l'ansia. L'elemento centrale di queste teorie è che uno stimolo condizionato accoppiato

ad uno stimolo incondizionato (che sembra essere nocivo e causare dolore) porterà, dopo diversi accoppiamenti, a una risposta condizionata che è paura/ansia. Queste sono viste come pulsioni secondarie o acquisite che sono sorte attraverso un processo di condizionamento classico (Strongman, 1995). Chi si discosta leggermente da queste teorie è Eysenck, che basa la teoria dell'ansia sulla sua teoria della personalità che vede, come abbiamo anticipato nel capitolo precedente, due dimensioni principali:

l'estroversione/introversione e il nevroticismo. Eysenck ritiene che l'individuo nevrotico sia particolarmente sensibile agli stimoli che provocano ansia, rendendolo più incline a questo stato e ad altre emozioni come il senso di colpa. Tuttavia, l'ansia può anche essere appresa, in quanto degli eventi traumatici possono portare a delle risposte ansiose disadattive. Qui, quindi, l'ansia è vista come paura condizionata (Eysenck, 1957).

La prospettiva fisiologia e neurofisiologica si concentra prevalentemente nell'individuare quali parti del sistema nervoso centrale potrebbero essere coinvolte nelle emozioni, e nell'ansia in particolare. L'importante teoria al quale dovremmo fare riferimento per l'ansia è quella di Gray, che deriva dalle concettualizzazioni di Eysenck. L'autore si è concentrato sul sistema di inibizione comportamentale (*behavioral inhibition system BIS*), il quale sopprime qualsiasi comportamento che minacci un risultato indesiderato, regolando dunque gli stati affettivi e il comportamento di evitamento. Questa mediazione è data dal sistema di lotta/fuga. Parla di un complesso sistema setto-ippocampale come alla base dell'ansia, anche se sono coinvolte anche altre parti del cervello come la proiezione neocorticale del sistema setto-ippocampale nel lobo frontale e le afferenze monoaminergiche che originano dal tronco encefalico (Gray, 1982).

La prospettiva fenomenologica origina con Kierkegaard (1844), il quale considera l'ansia come una parte integrante dell'esperienza umana che viene sperimentata ogni qualvolta che un individuo si trova, dalla nascita in poi, di fronte ad una scelta/possibilità. Successivamente Fisher integra diverse teorie in una che descrive cinque componenti dell'esperienza ansiosa:

- 1- Identità: pietre miliari che indicano modi di vivere,
- 2- Mondo: rete di relazioni e coinvolgimenti per ogni sfera dell'identità
- 3- Motivazione: modo in cui permangono le due componenti precedenti
- 4- Azione: raggiungimento di un traguardo
- 5- Abilità: valutazione di una competenza incerta

Nel momento in cui queste componenti vengono minacciate può scaturire l'ansia (Fischer, 1970).

Infine, nella prospettiva cognitiva, troviamo le importanti teorie di Eysenck e Öhman. Partendo dalla prima, Eysenck individua nelle persone degli schemi del sé presupponendo che appartengano al sistema cognitivo, e che dipendono dalla rilevanza personale di ogni particolare tratto per l'individuo. Questa teoria nasce dall'analisi alla base delle cause sottostanti le differenze nelle persone di suscettibilità allo stress. L'autore finisce col dimostrare che coloro che hanno un'ansia alta o bassa in effetti differiscono nella struttura e nei processi cognitivi, come anche nella memoria. Questo perché hanno delle preoccupazioni più frequenti o perché potrebbero essere più suscettibili all'ansia in alcune situazioni rispetto ad altre.

Öhman propone, successivamente, quella che definisce una teoria dell'elaborazione delle informazioni dell'ansia (**Figura 1**), suddivisa in cinque aspetti principali:

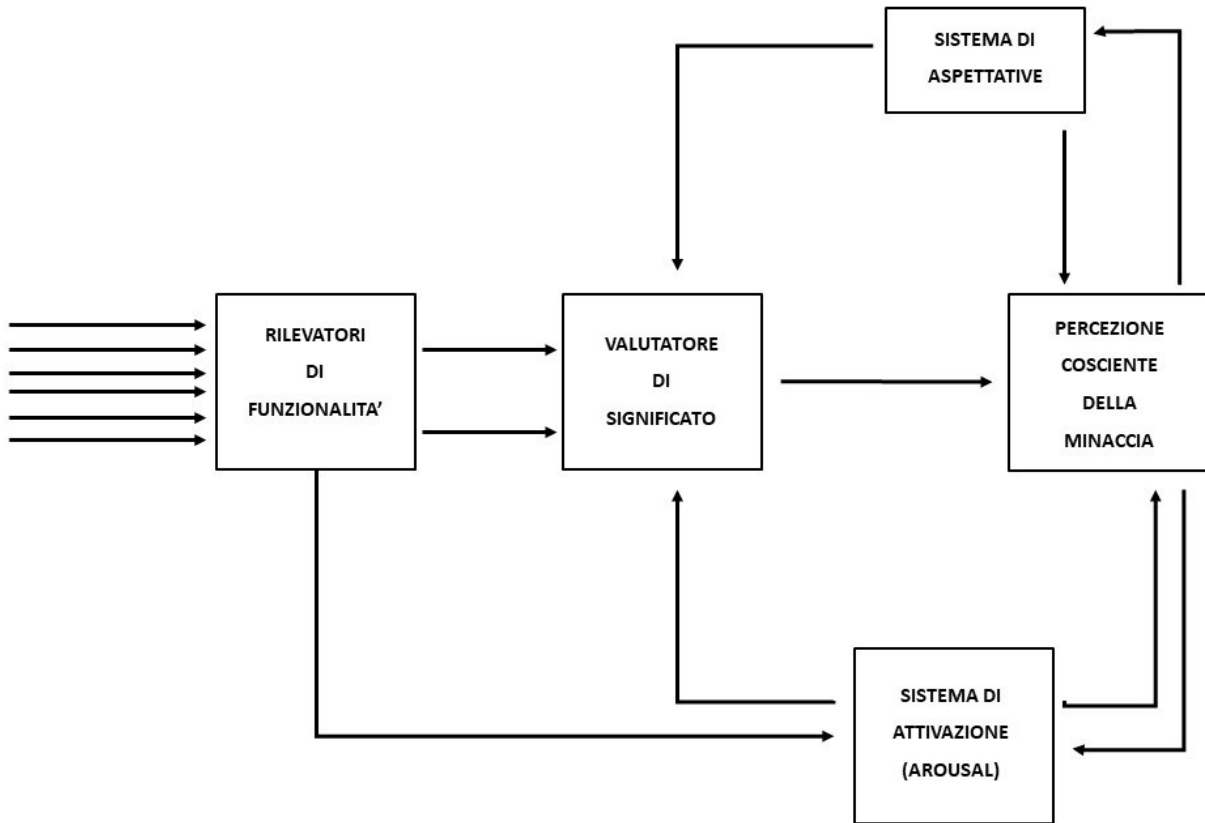


Figura 1 *Modello di elaborazione delle informazioni dell'ansia* (Ohman, 1993)

- 1- le informazioni sullo stimolo entrano nei rilevatori di funzionalità che trasmettono le informazioni ai valutatori di significatività
- 2- I valutatori di significatività valutano automaticamente la pertinenza e il significato degli stimoli, tenendo in considerazione le informazioni immagazzinate in memoria.
- 3- Il valutatore di significato fornisce un input al sistema di percezione cosciente.

4- Il sistema di aspettativa si basa sull'organizzazione dell'emozione nella memoria, il che influenza il modo di reagire alle informazioni.

5- Il sistema di percezione cosciente, dunque integrando l'input con tutti questi processi, sceglie un'azione adatta: se l'evitamento o la fuga è possibile, il risultato è la paura, in caso contrario il risultato è l'ansia. (Ohman, 1993)

Per fare un esempio, se uno studente con dislessia ha immagazzinato in memoria delle emozioni negative di vergogna/imbarazzo per un'esperienza di lettura ad alta voce a scuola, in età adulta davanti ad un intervento in un seminario, il suo sistema di eccitazione amplificherà la richiesta, facendo sì che il valutatore di significato percepirà la situazione come una minaccia, innescando emozioni negative e distorcendo le informazioni (Abbott-Jones, 2021).

2.3 METODI DI MISURAZIONE DEL COSTRUTTO

2.3.1 STATE-TRAIT ANXIETY INVENTORY (STAI)

Cattell fu il primo, nel 1958, ad aver distinto due accezioni differenti nel costrutto di ansia. Una riferita allo stato emotivo di un individuo in un determinato momento, ed una riferita ad una variabile di personalità che può differenziare individui diversi (Cattell & Scheier, 1958). Ma è grazie a Spielberger se si è diffusa internazionalmente l'idea che l'ansia fosse un costrutto multidimensionale, distinguibile in ansia di tratto (Trait-Anxiety) e ansia di stato (State-Anxiety). L'ansia di stato è una condizione transitoria (variabile nel tempo) ed emozionale dell'organismo caratterizzata da sentimenti soggettivi percepiti a livello cosciente di tensione ed attivazione del sistema nervoso autonomo. Viene accostata al concetto di Energia cinetica, intesa come una reazione che avviene in un determinato momento e ad una determinata intensità.

L'ansia di tratto è, invece, la predisposizione di un individuo a rispondere con un incremento dell'intensità dell'ansia di stato a situazioni percepite come minacciose. Sono delle differenze relativamente stabili. Viene accostata al concetto di Energia potenziale, intesa come differenza nella disposizione a manifestare un certo tipo di reazione (Spielberger et al., 1968).

Nel 1964 inizia dunque la costruzione dello STAI (State-Trait Anxiety Inventory), uno strumento di autovalutazione (self-report) costituito da due scale (S Anxiety e T Anxiety) di 20 item ciascuna con scala di risposta a 4 punti, dal quale si ricavano due punteggi attraverso il ricorso ad apposite griglie di scoring (Sanavio & Sica, 1997).

La prima valuta l'ansia di stato, chiedendo come si sentono gli intervistati "in questo momento", utilizzando elementi che misurano sentimenti soggettivi di apprensione, tensione, nervosismo, preoccupazione, ottenendo una misura dell'intensità dei sentimenti "in questo momento": 1) per niente, 2) un po', 3) moderatamente e 4) moltissimo.

La seconda valuta gli aspetti relativamente stabili della "predisposizione all'ansia", compresi gli stati generali di calma, fiducia e sicurezza, ottenendo una misura della frequenza dei sentimenti "in generale": 1) quasi mai, 2) qualche volta, 3) spesso e 4) quasi sempre.

Esistono anche una STAI per bambini (STAIC) con lo stesso numero di item, e delle versioni brevi (Julian, 2011).

Una prima versione viene pubblicata nel 1970: lo STAI-X. Lo strumento è stato successivamente rivisitato nel 1983 (STAI-Y). La revisione ha comportato la modifica di 6 item per ciascuna delle scale, con lo scopo di differenziare meglio il costrutto di

ansia da quello di depressione e offrendo un miglior bilanciamento fra gli item che indicavano presenza di ansia e quelli che ne indicavano assenza. In generale, si tratta di un questionario carta-matita (oggi anche online) che può essere somministrato sia in modo individuale sia in gruppo. Richiede una compilazione di circa 10-15 minuti, evitando l'uso del termine ansia e somministrando sempre prima la scala di Stato e poi quella di Tratto affinché non vi sia un'influenza da parte del proprio clima emotivo (Sanavio & Sica, 1997).

Tra le caratteristiche psicometriche, l'attendibilità è stata valutata tramite test-retest. I risultati per la scala di Stato sono compresi tra .34 e .62 ($r = .34 - .62$ dopo un intervallo di 30 giorni, $r = .36$ e $.51$ dopo un intervallo di 60 giorni), mentre sono compresi tra .65 e .75 per la scala di Tratto ($r = .71 - .75$ dopo un intervallo di 30 giorni, $r = .65 - .68$ dopo un intervallo di 60 giorni). Tali valori suggeriscono una buona stabilità temporale per la scala di tratto e, come atteso, una bassa stabilità temporale per la scala di stato. I coefficienti di coerenza interna sono risultati abbastanza elevati: tra .86 e .95 per la scala di Stato, tra .89 e .91 per la scala di Tratto (Sanavio & Sica, 1997).

2.3.2 BECK ANXIETY INVENTORY (BAI)

Il BAI è una breve misura dell'ansia con focus sui sintomi somatici e con una forte abilità nel discriminare l'ansia dalla depressione (Beck & Steer, 1988). È anch'essa una misura self-report, che valuta però alcuni sintomi specifici come ad esempio nervosismo, incapacità di rilassarsi, ecc. La scala è costituita da 21 item con scala di risposta Likert a 4 punti (da 0 per niente a 3 severamente). Può essere somministrato in formato individuale, inizialmente solo carta-matita, ora anche in modalità computerizzata. Il punteggio si ottiene sommando i vari punteggi di ogni singolo item,

con un totale che può variare da 0 a 63. L'interpretazione è la seguente: 0–9 ansia normale o assente; 10-18 ansia da lieve a moderata; 19–29 ansia da moderata a grave; e 30-63 grave ansia (Julian, 2011). La somministrazione richiede all'incirca 5-10 minuti, ed esistono traduzioni della scala in molte lingue: spagnolo, francese, tedesco, norvegese ed altre (Steer et al., 1993).

Nel valutare l'affidabilità emerge una coerenza interna elevata con valori alfa di Cronbach compresi tra 0.90 e 0.94, mentre i coefficienti test-retest sono compresi tra .62 e .93 (Beck & Steer, 1991).

2.3.3 ENDLER MULTIDIMENSIONAL ANXIETY SCALES (EMAS)

L'EMAS è un inventario che segue la prospettiva interazionista della personalità, che cerca quindi di mediare tra le teorie classiche (che cercano spiegazione nel comportamento del singolo) e le teorie situazionali (che cercano spiegazione nella situazione), individuando come migliore previsore del livello dell'ansia l'interazione tratti \times situazione. Considera dunque il versante situazionale, il versante personale e l'interazione tra i due (Endler & Magnusson, 1976).

L'ansia viene raffigurata e scomposta in quattro potenziali situazioni specifiche: situazione di valutazione sociale o minaccia interpersonale, situazioni di minaccia da danno o pericolo fisico, situazioni di minaccia data da situazioni ambigue, situazioni di minaccia in situazioni innocue o di normale routine quotidiana. Da un lato, dunque, troviamo un elenco di stimoli ansiogeni con item che descrivono la situazione stimolo, e dall'altro lato troviamo un elenco di possibili risposte d'ansia. Uno strumento che offre un formato di quattro situazioni specifiche messe in relazione a quindici possibili risposte d'ansia. Le scale di Endler diventano dunque un connubio tra un inventario

delle paure, un inventario dell'ansia di tratto e uno di ansia di stato (Sanavio & Sica, 1997).

È costituito da tre parti con scala di risposta a 5 punti: per niente – molto. La scala EMAS-S è una scala di stato con 20 item (10 per la dimensione emozionale, 10 per la dimensione cognitiva) che chiede al soggetto come si sente nel particolare momento in cui compila il questionario. La scala EMAS-T è una scala di tratto con 15 item che chiede al soggetto di indicare come abitualmente reagisce nelle quattro situazioni specifiche. La scala EMAS-P è una scala con 8 item che chiede al soggetto di valutare la situazione attuale in rapporto alle quattro situazioni specifiche della scala EMAS-T. L'inventario è autosomministrato, e non è necessario somministrare tutte e tre le scale. Lo scoring dei punteggi avviene attraverso la trasformazione dei punteggi in ranghi percentili e punti t in rapporto a sesso ed età. Più è alto il punteggio, più la variabile è accentuata nella direzione dell'ansia (Endler et al., 1962).

Lo strumento è stato tradotto e adattato anche al contesto italiano. Il manuale riporta norme italiane recenti, ma al tempo stesso porta con sé un problema di utilizzo dovuto all'eccessiva sintesi sull'interpretazione dei punteggi (Endler et al., 1991).

3. LA MISURAZIONE IN PSICOLOGIA

La misurazione in psicologia può essere descritta come *“una procedura per identificare gli elementi del mondo reale con degli elementi del sistema logico astratto, attraverso precise definizioni ed alcune regole precise”*. Il fine di tale procedura è quello di identificare elementi del mondo psichico con elementi del sistema numerico (Rubini, 1984).

Esistono molteplici livelli di misurazione. Abbiamo la possibilità di classificare attraverso la scala nominale, ovvero raggruppare diversi elementi in classi o categorie, in base al possesso di una definita caratteristica. Possiamo poi costituire una graduatoria tramite scala ordinale, ottenendo una relazione di intensità differenziata nella direzione maggiore o minore. Possiamo misurare su scala a intervalli uguali fissando convenzionalmente un punto zero e considerando un'unità di misura definita. Oppure possiamo misurare su scala a rapporti uguali avendo uno zero assoluto ed un'unità di misura convenzionale (Antonelli & Rubini, 2017).

In psicologia, nel processo diagnostico di raccolta ed analisi delle informazioni diviene un'importantissima risorsa quella del test psicologico, in quanto esso consente di quantificare e misurare i costrutti psicologici, consentendo quindi un progressivo approfondimento dei singoli casi, attraverso esclusione e falsificazione di ipotesi, come un imbuto che va via via restringendosi. Nelle fasi iniziali è opportuno dunque utilizzare il colloquio per un'analisi ad ampio spettro, mentre in fasi più avanzate sarà opportuno l'utilizzo dei test che offrono l'analisi di costrutti più specifici e mirati. Questionari e inventari di personalità analizzano l'autoriferito del paziente e possono/devono essere integrati con il colloquio clinico e con informazioni provenienti da fonti osservative (Sanavio & Sica, 1997).

Un esempio di modello per l'assessment è quello multidimensionale, che vede uno spazio tridimensionale comprendente una dimensione cognitivo-verbale costituita dalle informazioni che il soggetto riferisce nel corso di un colloquio o intervista. Le informazioni in questo canale sono influenzate dal contesto, dalla relazione interpersonale e da altre innumerevoli variabile definite *biases* (proprie o dello psicologo). Il secondo canale di questo spazio multidimensionale è quello comportamentale-motorio costituito dall'osservazione diretta del comportamento della persona. Infine, troviamo lo spazio psicofisiologico, costituito da informazioni ricavabili dalla registrazione di risposte elettrodermiche, dall'attività elettromiografica, dalla temperatura cutanea, dalla frequenza cardiaca e quella respiratoria. Queste tre dimensioni sono da considerare tra loro connesse ma relativamente indipendenti (Sanavio & Sica, 1997).

3.1 I TEST PSICOLOGICI

Entrando un po' più nello specifico, un test psicologico viene definito come “una procedura sistematica per osservare il comportamento di una persona e descriverlo con l'aiuto di una scala numerica o insieme di categorie” (Cronbach, 1970) dove vengono presentati degli stimoli (domande, problemi, compiti) in modo da ottenere particolari risposte valutabili ed interpretabili quantitativamente attraverso alcuni criteri standard (Pedrabissi & Santinelli, 1997).

Un test deve essere caratterizzato da stimoli uguali per tutti i soggetti e questi devono essere presentati in condizioni standard. Le risposte del soggetto devono poi essere codificate ed interpretate allo stesso modo, seguendo un legame risposta-variabile stabilito precedentemente in base a delle regole fisse (Boncori, 2006).

È generalmente composto da: materiale-stimolo (item), foglio di risposta, griglia di correzione o guida per la codifica, un manuale che contiene riferimenti teorici e metodologici, scopi, usi, informazioni sulla popolazione a cui è adatto, istruzioni dettagliate per la somministrazione e lo scoring e infine le norme statistiche per la trasformazione dei punteggi (Antonelli & Rubini, 2017).

Storicamente si attribuisce a Kraepelin uno dei primi tentativi di messa a punto di prove per la valutazione della personalità anormale, attraverso una prova di libere associazioni che consisteva nella presentazione di alcune parole-stimolo al quale il soggetto doveva rispondere con la prima parola che gli veniva in mente (Kraepelin, 1892). Il vero e proprio questionario di personalità fu però sviluppato per uso bellici, con il nome di *Personal Data Sheet*, noto anche come *Psychoneurotic Inventory*, per identificare velocemente i soggetti inabili al servizio militare per problemi emotivi, il quale costituì un modello per i successivi questionari (DuBois, 1970).

Esistono differenti tipi di test, e la principale classificazione è quella che li divide in test cognitivi e test non cognitivi (di personalità). I primi misurano il rendimento massimo o il rendimento minimo e comprende i test d'intelligenza, i test attitudinali o di abilità specifiche, i test di profitto e quelli neuropsicologici. I secondi, invece, misurano le reazioni tipiche e comprendono i questionari autodescrittivi, quelli che misurano interessi, atteggiamenti e valori e infine i test proiettivi (Antonelli & Rubini, 2017).

3.2 CARATTERISTICHE METRICHE

L'APA (1954) emanò negli anni '50 una serie di norme per la costruzione e per il corretto uso dei test psicologici, introducendo i concetti fondamentali di affidabilità e validità necessari per definire la bontà di uno strumento.

L'attendibilità (*reliability*), o affidabilità, o fedeltà, può essere definita in vari modi: "Grado di indipendenza dello strumento di misura dall'errore casuale" (Rubini, 1975), o ancora, "Grado di accuratezza e di precisione di una procedura di misurazione" (Pedrabissi & Santinello, 1997). Uno strumento di misura è dunque affidabile quando i punteggi ottenuti da uno stesso soggetto allo stesso test sono stabili in misurazioni ripetute, anche compiute da persone diverse, in situazioni diverse. L'attendibilità di un test viene indicata da un coefficiente di attendibilità (r_{tt}), definito come rapporto tra varianza del punteggio vero (s_V^2) e varianza del punteggio osservato (s_X^2). Tale coefficiente assume valori compresi tra 1 e 0 e può essere interpretato in termini percentuali. I metodi di calcolo del coefficiente di attendibilità sono molteplici. È possibile utilizzare metodi che prevedono due somministrazioni di un test (metodo *test-retest*; forme parallele dello stesso test) o metodi ad una sola somministrazione (divisione a metà del test, detta anche *split-half*; partendo dai singoli item, attraverso *alpha di Cronbach*) (Antonelli & Rubini, 2017).

La validità viene generalmente intesa come congruenza tra il costrutto oggetto di misurazione e ciò che il test restituisce come misura effettiva. Un test è valido se misura ciò che dichiara di misurare (Kline, 1993). Un'altra definizione di validità definisce tale proprietà come, il grado di precisione e di accuratezza con cui (il test) misura ciò che si propone di misurare (Pedrabissi & Santinello, 1997; Rubini, 1984). Può essere indicata anche come la distanza tra il punteggio osservato di un soggetto e il suo punteggio vero. L'approccio tradizionale definisce diversi tipi di validità:

- esteriore o di facciata: si riscontra quando il test sembra misurare ciò che dichiara di misurare o se la struttura e gli item che compongono il test appaiono motivanti e significativi per chi vi è sottoposto;

- di contenuto: rappresenta il grado in cui gli item che fanno parte del test costituiscono un campione rappresentativo dell'universo dei comportamenti possibili relativi al costrutto che si vuole misurare;
- di criterio: esprime il grado di associazione tra la misurazione del costrutto ottenuta tramite il test ed una variabile esterna utilizzata come criterio di riferimento. Prende il nome di concorrente quando le misure vengono acquisite contemporaneamente, mentre predittiva quando viene eseguita prima la misura del costrutto e poi del criterio. La validità concorrente può essere valutata mediante il calcolo del coefficiente di correlazione o mediante il metodo dei gruppi contrapposti. Il coefficiente di correlazione può essere: r di Pearson per le scale ad intervalli uguali, ρ di Spearman per misure su scala ordinale, R_{pbis} (correlazione punto biseriale) se la misura del criterio è dicotomica. Il metodo dei "gruppi contrapposti"² invece può utilizzare: il test t per campioni indipendenti e ANOVA se la misura del costrutto è su scala ad intervalli uguali. Oppure ancora il test U di Mann-Whitney e il test H di Kruskal-Wallis se la misura del costrutto è su scala ordinale (Pedrabissi & Santinello, 1997);
- di costrutto: esprime il grado di precisione con cui uno strumento misura ciò che intende misurare. I metodi per misurarla sono: l'analisi delle correlazioni tra le misure ottenute con lo strumento, le matrici multi-tratto multi-metodo (MTMM³) (Campbell & Fiske, 1959) e l'analisi fattoriale⁴. Quest'ultima può avere

² "gruppi contrapposti": metodo con il quale si confrontano i risultati ottenuti nel test da gruppi a composizione nota, cioè che sono già noti per essere diversi nel costrutto oggetto di misurazione

³ MTMM; l'assunto di base è che se il costrutto in esame rappresenta una caratteristica universale che si manifesta in situazioni diverse e può essere individuato in modi diversi, allora la sua misurazione non dovrebbe variare in funzione del metodo utilizzato per misurarla (Wothke, 1995)

⁴ Analisi fattoriale: insieme di tecniche per l'analisi delle relazioni tra variabili, legate da un obiettivo comune, quello di rappresentare un numero elevato di variabili per mezzo di un numero inferiore di variabili ipotetiche (o latenti) chiamate fattori.

un'ottica esplorativa, quando si utilizza nel momento in cui non si hanno idee circa il numero di dimensioni che sottendono le variabili rilevate, oppure un'ottica confermativa quando lo scopo è quello di verificare ipotesi specifiche rispetto alla struttura dei dati.

Le prime due tipologie di validità descritte non vengono generalmente espresse da indici metrici, mentre le ultime due vengono espresse da indici metrici. Approcci più moderni, invece, identificano la validità di un test in generale con la validità di costrutto (Antonelli & Rubini, 2017).

3.3 VANTAGGI E LIMITI DELLA MISURAZIONE

Come si evince da questi importanti passaggi, i test psicologici sono un importante strumento che consente di classificare (indipendentemente dal particolare ambito in cui viene svolta l'attività di testing), effettuare diagnosi e impostare programmi di azione, favorire l'autoconoscenza, valutare programmi realizzati in ambito organizzativo, educativo, sociale e psicologico e infine effettuare programmi di ricerca, per sostenere e sviluppare la ricerca comportamentale teorica e applicativa (Boncori, 2006).

Al tempo stesso però occorre tenere ben presente che per far sì che la misura sia valida ed attendibile è necessario rispettare la standardizzazione delle condizioni di applicazione di un test, cercando di escludere tutti quei fattori che potrebbero inquinare la raccolta dei dati. Questi elementi che potrebbero produrre delle distorsioni nelle risposte dei soggetti possono essere presenti in tutto il processo di testing, sia in fase di consegna delle istruzioni, sia durante l'esecuzione del test. Alcune tra le più riconosciute fonti di bias sono le caratteristiche che riguardano l'esaminatore come il genere, o l'atteggiamento nella fase di somministrazione del test che potrebbe, ad esempio,

esporre al rischio di suggerire delle risposte, o al contrario essere freddo e meno disponibile (Antonelli & Rubini, 2017).

Vi sono poi i *response bias*, che sono tutti dovuti ad una tendenza sistematica, e non casuale, a rispondere alle domande in base a qualcosa di diverso dal contenuto specifico dell'item (Paulhus, 1991). Alcuni esempi sono la desiderabilità sociale (tendenza di un individuo a porsi in luce positiva), l'acquiescenza (tendenza a dichiararsi d'accordo con un'affermazione indipendentemente dal contenuto della stessa), la tendenza a dare risposte estreme (tendenza a scegliere le categorie di risposta ai poli della scala di risposta 1 o 5 ad esempio). Ancora, vi possono essere degli effetti contestuali dovuti all'ordine delle domande come, ad esempio, l'effetto priming (quando la prima domanda ha un effetto su quella successiva), o ancora delle distorsioni collegate al grado di motivazione o di informazioni sull'argomento in questione (Zammuner, 1998).

3.4 ITEM RESPONSE THEORY

La costruzione di strumenti di assessment psicologico si basa principalmente su due grandi cornici teoriche: la teoria classica dei test (*Classical test theory* CTT) e la modern test theory. In quest'ultimo approccio assumono grande rilievo i modelli basati sull'*Item Response Theory* (IRT), nominata anche *Latent trait theory* o *Item characteristic curve theory*.

La CTT si focalizza sul punteggio totale del test e lo considera come una quantità che può essere scomposta in un punteggio vero e una componente di errore. L'assunto di base è che in ogni misurazione siano strutturalmente presenti due componenti: una che riflette la quantità di tratto presente nei rispondenti (punteggio vero) e che viene valutata attraverso le loro risposte agli item ed una che costituisce l'errore di misura, determinato

da distorsioni nella procedura di testing (punteggio di errore). La quota di errore casuale prende il nome di varianza d'errore s_e^2 . Quanto più è piccola la sua deviazione standard s_e , tanto più la misura è attendibile. La CTT, dunque, attribuisce grande importanza al concetto di affidabilità del test, basandosi però principalmente sui punteggi totali (Antonelli & Rubini, 2017).

L'IRT si sviluppa intorno agli inizi del 1900 e a partire dagli anni '50-'60 in poi viene perfezionata ed approfondita da molti altri autori come Lord e Novick (1968), Wright (1977) e Rasch (1960) solo per citarne alcuni. I modelli IRT hanno inizialmente trovato applicazione soprattutto nell'ambito dei test cognitivi, costituiti essenzialmente da item a scelta multipla nel formato dicotomico "giusto-sbagliato", "vero-falso". I modelli IRT unidimensionali condividono il presupposto che un singolo costrutto latente sia il determinante causale delle risposte a ciascuno degli item del test. Rispetto alla CTT, l'IRT attribuisce maggiore rilevanza ai singoli item, e assume che la probabilità di osservare una risposta corretta ad un item da parte di un soggetto sia funzione di una caratteristica del soggetto, cioè il suo livello di tratto latente (θ), e di una o più caratteristiche degli item (Weiss & Yoes, 1991). Le caratteristiche degli item che vengono considerate danno origine a diversi modelli, di complessità crescente. Il *Simple Logistic Model* (o Modello di Rasch o Modello ad 1 Parametro) considera il solo parametro di difficoltà (livello di abilità richiesto affinché un soggetto abbia le stesse probabilità di superare o fallire l'item), il Modello a 2 Parametri affianca al parametro di difficoltà quello di discriminatività (capacità dell'item di discriminare tra soggetti che hanno differenti livelli di tratto), il Modello a 3 Parametri considera oltre ai precedenti il parametro guessing (influenza del caso nel determinare il superamento dell'item) e il

Modello a 4 Parametri aggiunge anche il parametro careless (influenza della distrazione nel determinare una risposta errata all'item); (Harvey & Hammer, 1999).

L'IRT presenta un grande potere innovativo e interessanti potenzialità applicative, riuscendo a superare i limiti della CTT che non riesce a chiarire del tutto il rapporto esistente tra le risposte agli item e le caratteristiche degli item stessi.

L'unidimensionalità è una caratteristica necessaria per l'applicazione di questi modelli, ma è noto che nelle scienze sociali e comportamentali molti dei costrutti in cui ci si imbatte sono di natura multidimensionale (Cai et al, 2016). Per questo motivo sono state sviluppate estensioni dei modelli IRT per dati multidimensionali.

Inoltre, estensioni dei modelli IRT sono state sviluppate anche per gestire dati politomici con scale ordinali (es. scale Likert) e con scale nominali non ordinate, rendendo l'approccio applicabile a qualsiasi strumento di valutazione psicologica standardizzata (Thissen & Steinberg, 1985). È sempre opportuno, dunque, riflettere e compiere delle scelte di metodo in funzione del tipo di risposte da valutare o dal numero di parametri da considerare nel modello.

3.4.1 MODELLO DI RASCH

Il modello di Rasch (anche noto come 1PL) è uno dei più semplici modelli IRT e prende il nome dal matematico Georg Rasch che lo elaborò nel 1960. Richiede l'utilizzo di un solo parametro per rappresentare il processo di risposta agli item. Questo parametro è chiamato difficoltà ed è spesso indicato come b . Operativamente parlando è definito come il punteggio su θ che è associato con una probabilità del 50% di fornire una risposta corretta all'item. Il parametro b è definito direttamente in funzione di θ . La probabilità di una risposta corretta all'item varia in funzione del livello di abilità, e

graficamente viene rappresentata da una funzione non lineare monotona crescente a forma di S, che prende il nome di Curva Caratteristica di un item (*Item Characteristic Curve*, ICC) (Baker, 2001). Nel modello ad un parametro, tutti gli item del test presentano la stessa forma di ICC, ma si contraddistinguono nella posizione destra-sinistra sull'asse orizzontale θ , ovvero la loro difficoltà (Harvey & Hammer, 1999). In altri termini, presuppone che la probabilità di dare una risposta corretta ad un item sia una funzione logistica della relativa distanza tra la posizione dell'item e quella del rispondente su una scala lineare. Un esempio concreto potrebbe essere la probabilità di rispondere correttamente all'item dovuta alla sola differenza tra il livello di tratto di ansia (θ) della persona e il livello di ansia espresso dall'item (b), calcolabile con la seguente formula: (Pallant & Tennant, 2007)

$$P_{\theta} = \frac{e(\theta_s - \beta_i)}{1 + e(\theta_s - \beta_i)}$$

Quando si ha a che fare con costrutti non cognitivi, come ad esempio l'ansia sopracitata, è opportuno parlare di livello di affettività (livello di tratto necessario per fornire una risposta indicativa del tratto, un item affettivo richiede molto tratto per fornire una risposta che ne indichi la presenza; item molto affettivi ricevono poche risposte indicative della presenza di tratto) dell'item e livelli di tratto latente.

3.4.2 ESTENSIONI DEL MODELLO DI RASCH

Dal modello di Rasch presero vita ulteriori sviluppi sia per item dicotomici che politomici.

Per quanto riguarda gli item dicotomici, un'estensione è il Modello a 2 Parametri (2PL) che aggiunge al parametro dell'affettività/difficoltà, quello della discriminatività. Con questo parametro si intende la capacità dell'item di discriminare tra soggetti che hanno

differenti livelli di tratto (chi alti, chi bassi). La probabilità di una risposta corretta a qualsiasi item del test è funzione dell'abilità del soggetto e dei parametri difficoltà (collocazione della curva sul tratto) e discriminatività (inclinazione/pendenza della curva). Gli item con parametri di discriminazione bassi sono meno informativi e le informazioni sono distribuite lungo una parte maggiore dell'intervallo di abilità. Mentre nell'1PL le ICC sono tutte della stessa forma e cambiano solo la collocazione lungo il tratto, nel modello 2PL anche se le curve raggiungono il loro massimo dove abilità del soggetto e difficoltà dell'item coincidono, le loro forme dipendono fortemente dal parametro della discriminazione (Harvey & Hammer, 1999).

Esiste poi il Modello a 3 Parametri (3PL), sempre per item dicotomici, che come si evince dal nome, individua tre principali parametri: difficoltà, discriminatività e parametro guessing. Quest'ultimo indica l'influenza del caso nel determinare il superamento dell'item da parte di rispondenti che non conoscono la risposta (Faraci, 2009).

Ultima estensione per item dicotomici è il Modello a 4 Parametri (4PL), che viene utilizzato in maniera molto minore rispetto agli altri in quanto molto recente. Tale modello aggiunge ai precedenti parametri quello del carelessness, che indica l'influenza della distrazione nel determinare una risposta errata all'item da parte dei soggetti con elevati livelli di tratto. Quindi in altre parole, tale parametro, indica l'influenza della disattenzione o distrazione nel fornire all'item una risposta errata da parte di una persona che dovrebbe conoscere la risposta corretta (Ho et al, 2012).

Per quanto riguarda, invece, gli item politomici, troviamo come estensioni del modello di Rasch, il *Partial credit model* (PCM) (Masters, 1982) e il *Rating scale model* (RSM)

(Andrich, 1978). Questi modelli, come il modello di Rasch, considerano un solo parametro ma a differenza del modello di Rasch, che usa dati dicotomici, si prestano ad analizzare item politomici (Masters, 1982).

I modelli RSM e PCM descrivono la probabilità che una persona fornisca un certo tipo di risposta ad un item politomico, tenendo conto del livello di tratto della persona e dell'affettività (o difficoltà) dell'item.

I modelli PCM e RSM consentono di stimare un insieme di soglie, ciascuna delle quali rappresenta la posizione sul continuum dei tratti latenti dove le categorie di risposta adiacenti sono ugualmente probabili (es. in una scala Likert categorie di risposta adiacenti potrebbero essere “molto” e “moltissimo”, o ancora “mai” e “quasi mai”). Si presuppone che vi siano tante soglie quante sono le categorie di risposta meno 1 ($k - 1$). Ogni soglia rappresenta il livello di tratto che è necessario per scegliere una determinata categoria di risposta al posto della precedente. Queste soglie sono necessarie per collegare il livello di tratto della persona con la categoria di risposta che questa persona dovrebbe scegliere. In particolare, maggiore è il livello di tratto della persona, maggiore dovrebbe essere la probabilità che questa persona scelga una categoria di risposta elevata (Andrich, 1978).

Più precisamente, il RSM prevede che le soglie siano le stesse per tutti gli item, mentre il PCM consente loro di differire da un item all'altro. Pertanto, il RSM fornisce una valutazione complessiva del funzionamento della scala di valutazione a livello dell'intero test, mentre il PCM fornisce una valutazione più dettagliata a livello di ogni singolo item (Masters, 1982; Andrich, 1978).

Tutti questi modelli, e l'IRT più in generale, offrono un notevole contributo nella costruzione di diversi strumenti che trovano applicazione in svariati ambiti, tra cui l'orientamento scolastico o professionale, la ricerca psicosociale, la gestione e selezione del personale (Faraci, 2009), e più recentemente anche nella misura di atteggiamenti e disposizioni comportamentali (Hambleton & Swaminathan, 1985).

4. LA RICERCA

4.1 IPOTESI DI RICERCA

Il presente lavoro di ricerca intende mostrare come il funzionamento delle *rating scales* possa essere verificato attraverso l'utilizzo del PCM.

Le *rating scales* utilizzano termini descrittivi relativi al costrutto oggetto dell'item, noto come "tratto latente" o "variabile" (Stanley, 1972, p. 290). Le categorie di risposta che vengono presentate agli intervistati hanno lo scopo di ottenere indicazioni ordinali sulle loro posizioni riguardo tali variabili di interesse. Questi tipi di scale di valutazione vengono utilizzate frequentemente nella costruzione dei questionari, in quanto si ritiene siano un mezzo per ottenere più informazioni dagli item rispetto a quelle che si otterrebbero da un formato di risposta dicotomico "sì/no", "giusto/sbagliato", ecc. (Linacre, 2002a).

Occorre tenere a mente però che l'intervistato nel momento della compilazione potrebbe essere incerto sulla categoria di risposta da scegliere, oppure potrebbe non comprenderla fornendo così delle risposte confuse (Wright & Linacre, 1992). Proprio per questo motivo è fondamentale verificare il funzionamento della scala di valutazione.

Come esposto in precedenza i modelli di Rasch per item politomici prevedono il calcolo di alcuni parametri detti soglie (τ) di Rasch-Andrich o semplicemente soglie, che consentono di testare empiricamente il funzionamento della scala. Le soglie che separano due categorie di risposta adiacenti sono i punti sulla variabile latente in corrispondenza dei quali queste due categorie hanno la stessa probabilità di essere selezionate dall'intervistato. Questa uguale probabilità è necessariamente 0.5 (50%) nel caso dei modelli dicotomici, mentre negli altri casi tale percentuale può variare, purché la somma delle probabilità di tutte le risposte possibili sia 1 (100%; Penta et al., 2008). Dunque,

supponendo che ci siano k categorie, allora ci sono $k - 1$ soglie. La prima soglia (τ_{1-2}) è il livello di tratto in cui un individuo ha la stessa probabilità di scegliere la categoria 1 o 2, la seconda soglia (τ_{2-3}) è il livello di tratto in cui un individuo ha la stessa probabilità di scegliere la categoria 2 o 3, e l'ultima soglia ($\tau_{(k-1)-k}$) è il livello di tratto in cui un individuo ha la stessa probabilità di scegliere la categoria $k - 1$ o k . Se le categorie riflettono livelli crescenti della variabile latente, come previsto, allora le soglie devono avanzare in modo monotono con le categorie. Tuttavia, è empiricamente possibile che le soglie non riflettano l'ordine previsto. Questa condizione, denominata "*threshold disordering*", ovvero "disordine delle soglie", fornisce la prova che la scala di valutazione non funziona come previsto. Tale disordine può voler dire che vi è una discordanza tra quantità di variabile latente e categorie scelte, oppure che una categoria rappresenta un intervallo troppo ristretto sulla variabile latente o che agli intervistati sono state presentate troppe categorie (Linacre, 1999; 2002a).

Proprio quest'ultimo aspetto è lo spunto di riflessione motore di questo lavoro. La ricerca prende in considerazione la scala di tratto del questionario *State-Trait Anxiety Inventory* STAI (Spielberger et al, 1983), che è uno strumento molto noto che misura i livelli di ansia. Lo studio vuole verificare e confrontare il funzionamento della scala originale a 4 punti con quello di una versione estesa dello stesso STAI composta da un numero doppio di categorie (cioè una scala a 8 punti), sviluppata appositamente per i fini di questo lavoro.

La domanda di ricerca nasce dall'interesse di verificare empiricamente se l'utilizzo di scale di risposta con un ampio numero di categorie consenta effettivamente la raccolta di maggiori informazioni. Per fare ciò si è dunque pensato di prendere in esame una scala ben validata come quella dello STAI e confrontarla con una scala più ampia creata appositamente.

Per entrambe le scale vengono effettuate in primo luogo delle analisi fattoriali esplorative e confermative, per verificarne la dimensionalità.

Dopodiché viene utilizzato il PCM per valutare il funzionamento delle soglie per ogni singolo item. Attraverso questo modello si riesce a vedere se le categorie di risposta vengono scelte dai rispondenti senza confusione. L'obiettivo di ricerca è dunque di verificare se lo strumento misura meglio il costrutto latente utilizzando la scala di risposta standard a 4 punti, o attraverso una nuova scala di risposta che utilizza 8 categorie.

4.2 METODO

4.2.1 PARTECIPANTI

I partecipanti sono stati reclutati attraverso un campionamento di convenienza e hanno compilato il questionario online individualmente, volontariamente, in maniera anonima e senza ricompensa monetaria. È stato richiesto ai partecipanti il consenso informato, e sono stati rispettati tutti gli standard di comportamento necessari per garantire la corretta realizzazione delle attività di ricerca con la finalità di proteggere i diritti dei partecipanti e delle persone coinvolte.

Sono stati reclutati due campioni, il primo campione è costituito da 508 soggetti (69,90% donne e 27,20% uomini) con un'età media di 29 anni ($SD = 11,71$) che hanno risposto agli item della scala in formato standard (i.e. 4 categorie di risposta), mentre il secondo campione è costituito da 437 soggetti (78% donne e 20,6% uomini) con un'età media di 28 anni ($SD = 9,71$) che hanno risposto agli item della scala con formato allungato di risposta (i.e. 8 categorie di risposta). Una rappresentazione grafica della distribuzione di genere nei due campioni è riportata in **Figura 2**.

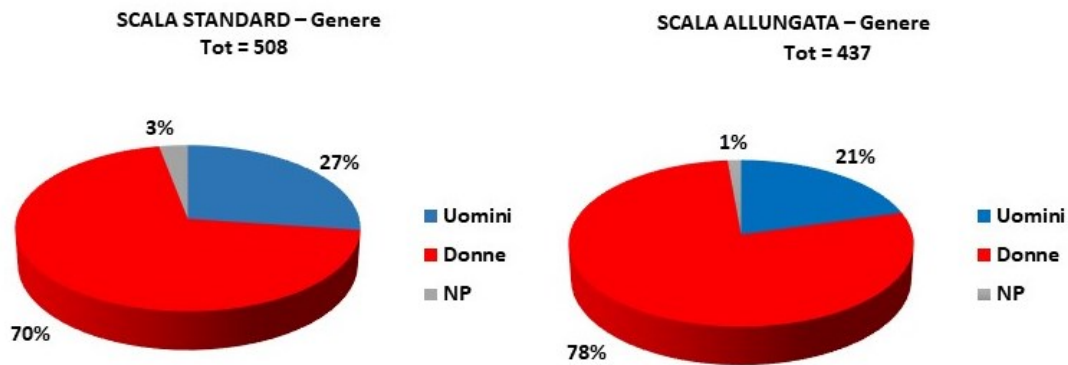


Figura 2 *Genere dei rispondenti alle scale STAI-T standard e allungata.*
 Nota: NP = preferisce non rispondere

Analizzando il livello di educazione dei rispondenti (**Figura 3**) emerge che in entrambe le condizioni i partecipanti sono stati per la maggior parte persone con educazione elevata (laurea e diploma superiore).

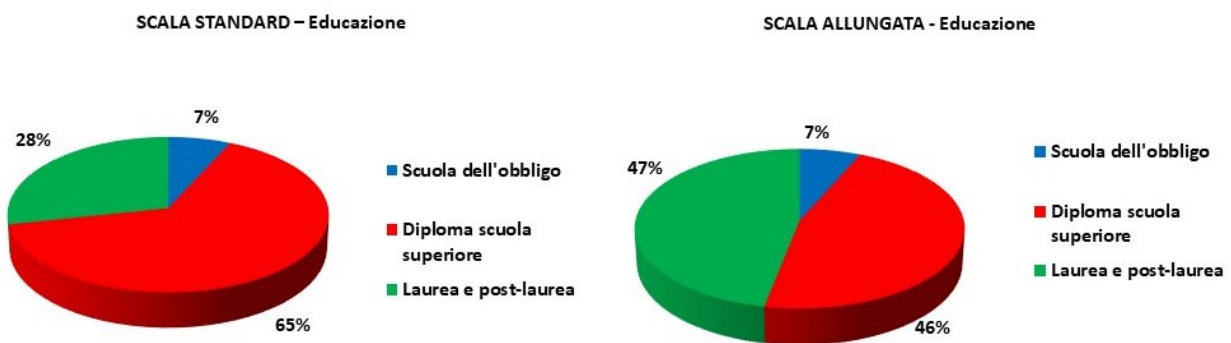


Figura 3 *Livello di educazione dei rispondenti alle scale STAI-T standard e allungata*

4.2.2 PROCEDURA E SOMMINISTRAZIONE

Ai partecipanti è stato presentato un questionario online realizzato con *Qualtrics* diviso in due parti principali:

1. Caratteristiche demografiche – 5 domande;
2. Vissuti individuali – 20 domande.

La parte di questionario sui vissuti individuali, è costituita principalmente dalla versione italiana (Pedrabissi & Santinello, 1989) dello STAI (Spielberger et al., 1983). Lo STAI è composto da 20 item che valutano i sentimenti di ansia generale e la predisposizione relativamente stabile a vedere le situazioni stressanti come minacciose. Un esempio di item è "*Ho pensieri disturbanti*".

Sono state prese in considerazione due versioni dello STAI, che differivano l'una dall'altra solo nella scala utilizzata per rispondere agli item. Una versione utilizzava la scala originale a 4 punti con le seguenti categorie di risposta: 1 "quasi mai", 2 "a volte", 3 "spesso", 4 "quasi sempre". L'altra versione invece comprendeva delle ulteriori categorie di risposta, create appositamente dai ricercatori per scopi di ricerca. Le categorie di risposta in questa nuova scala estesa erano 8: 1 "mai", 2 "quasi mai", 3 "raramente", 4 "a volte", 5 "spesso", 6 "molto spesso"; 7 "quasi sempre"; 8 "sempre" (come si può notare le quattro categorie della scala originale sono incluse nella scala allungata). In quanto segue, queste due versioni dello STAI saranno indicate rispettivamente come STAI-STD e STAI-LONG.

4.2.3 STRATEGIA D'ANALISI

In primo luogo, è stata accertata la dimensionalità delle due versioni dello STAI attraverso una *parallel analysis* (Horn, 1965) con il pacchetto "*Psych*" per il software *R* (R Core Team, 2021). La tecnica PA consente di individuare il numero di fattori adatto a rappresentare la struttura fattoriale di una scala. Il metodo prevede di trattenere tanti fattori quanti sono quelli i cui autovalori stimati sui dati reali superano quelli stimati su un certo numero di matrici di dati simulati. In questo lavoro, in particolare, il numero di fattori da considerare è stato determinato valutando il numero di autovalori, stimati sui

dati reali, con autovalori superiori al 95° percentile della distribuzione degli autovalori stimati su 1.000 matrici di correlazione simulate.

Dopodiché è stata effettuata un'analisi fattoriale esplorativa (EFA) al fine di esaminare le proprietà degli item della scala. L'analisi è stata svolta con il software Mplus 7.4 (Muthén & Muthén, 2012) e utilizzando il metodo di stima *Maximum Likelihood Robust* (MLR) (Yuan & Bentler, 2000) con rotazione obliqua Geomin. Tali analisi sono state condotte sia per la versione standard del questionario, che per la versione allungata.

Successivamente, sempre per entrambe le scale, è stata effettuata l'analisi fattoriale confermativa (CFA) utilizzando il metodo di stima *Maximum Likelihood Mean and Variance Adjusted* (MLMV; Muthén & Muthén, 2012).

La bontà di adattamento dei modelli è stata valutata per mezzo di diversi indici di fit. χ^2 è sensibile alla dimensione del campione, quindi per valutare i modelli sono stati presi in considerazione anche altri indici di adattamento. In particolare i valori CFI - *Comparative Fit Index* (Bentler, 1990) vicini a 0,95 (da 0,90 a 0,95 per un adattamento accettabile), e i valori di SRMS - *Standardized Root Mean Square Residual* (Bentler, 1995) e RMSEA - *Root Mean Square Error of Approximation* inferiori a 0,06 (da 0,06 a 0,08 per un adattamento accettabile; Browne & Cudeck, 1993).

La verifica della struttura fattoriale è stata un passaggio necessario in quanto il modello di Rasch richiede la monodimensionalità dei dati.

Infine, verificata la monodimensionalità è stata eseguita l'analisi con il PCM (Masters, 1982), appartenente alla famiglia dei modelli di Rasch (1960). Questo modello consente di valutare il funzionamento della scala di valutazione per ogni singolo item, analizzando l'ordine delle soglie. Se le categorie riflettono livelli crescenti di variabile latente, allora

le soglie essere ordinate in modo crescente (ordine monotono). Qualora avanzino in modo disordinato sarebbe indice di un malfunzionamento della scala di valutazione.

Sono state, inoltre, esaminate le statistiche basate sul χ^2 chiamate infit e outfit (Wright & Masters, 1982) che consentono di valutare la bontà degli item. L'infit è un indice sensibile a risposte inattese di piccola entità rispetto alle attese del modello, mentre l'outfit è più sensibile a risposte inattese di grande entità (Linacre, 2002b). Se ad esempio consideriamo una scala a 7 punti, una risposta osservata di 1 influisce principalmente nell'infit se la risposta attesa è 2, mentre influisce nell'outfit se la risposta attesa è 7. In questa ricerca l'infit e l'outfit sono stati calcolati per ogni singolo item. Valori superiori a 2 per un item sono stati considerati come un indice del mal funzionamento dello stesso. In genere, valori superiori a 2 indicano item mal formulati, o che misurano una dimensione diversa da quella misurata dagli altri item, in tal caso si parlerebbe di multidimensionalità (Linacre, 2002b; Smith, 2002).

Infine, è stato esaminato il *Person Separation Reliability R*, una misura di consistenza interna dello strumento che può variare da 0 a 1. Più il valore è vicino a 1, maggiore è la probabilità che le differenze tra punteggi latenti esprimano vere differenze tra i livelli di ansia di tratto. *R* è l'equivalente dell' α di Cronbach nel framework Rasch, ed è considerato un migliore indice di consistenza interna perché le misure stimate per ciascun rispondente sono su scala lineare (Smith, 2001; Wright & Stone, 1999), e perché considera l'effettiva varianza d'errore del campione invece della varianza d'errore di un rispondente medio. (Smith, 2002).

Le analisi PCM sono state effettuate con il pacchetto “*eRm*” (Mair et al., 2016) del software *R* (R Core Team, 2021).

4.3 RISULTATI

I risultati delle PA sono riportati nel **Grafico 1** e nel **Grafico 2**. I risultati suggeriscono che sia la scala STAI-STD, che la scala STAI-LONG possano essere rappresentate da due componenti.

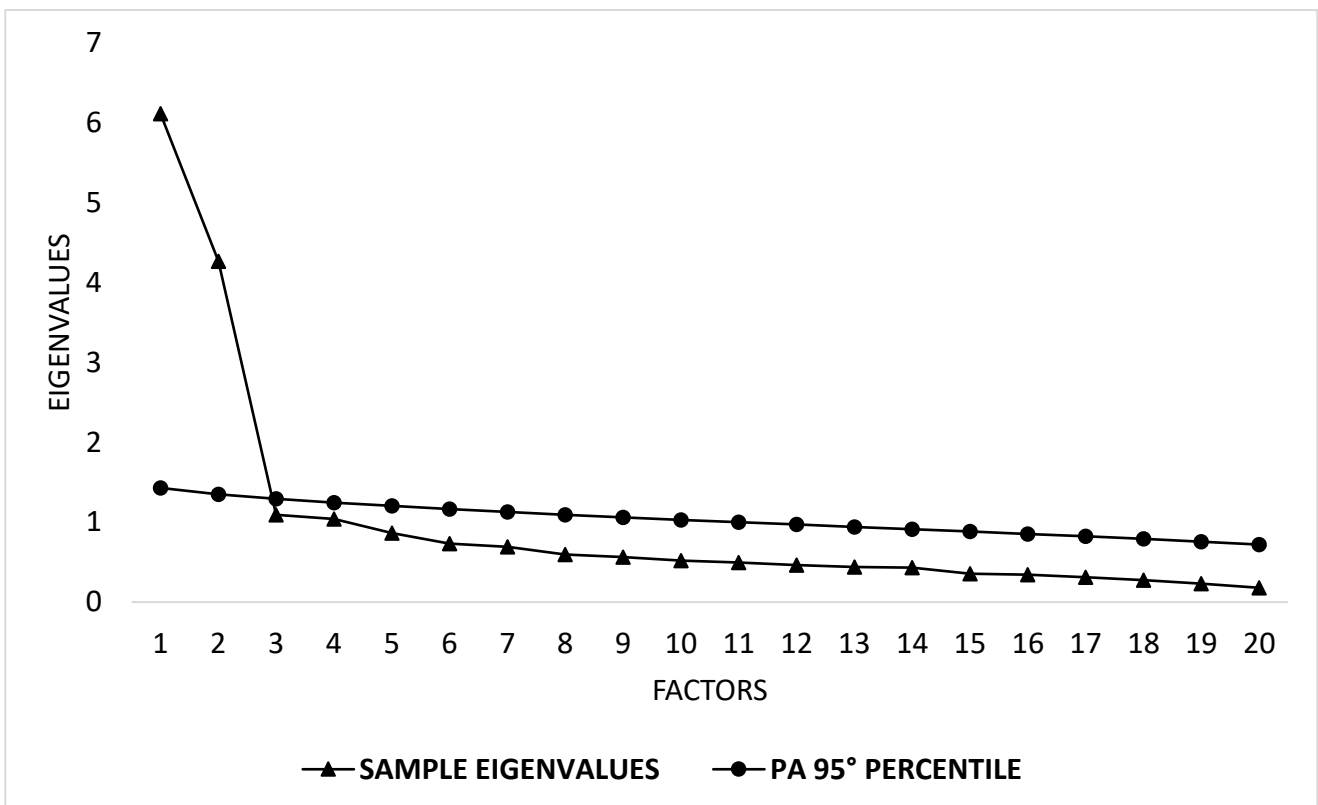


Grafico 1 *Parallel analysis della scala STAI-STD*

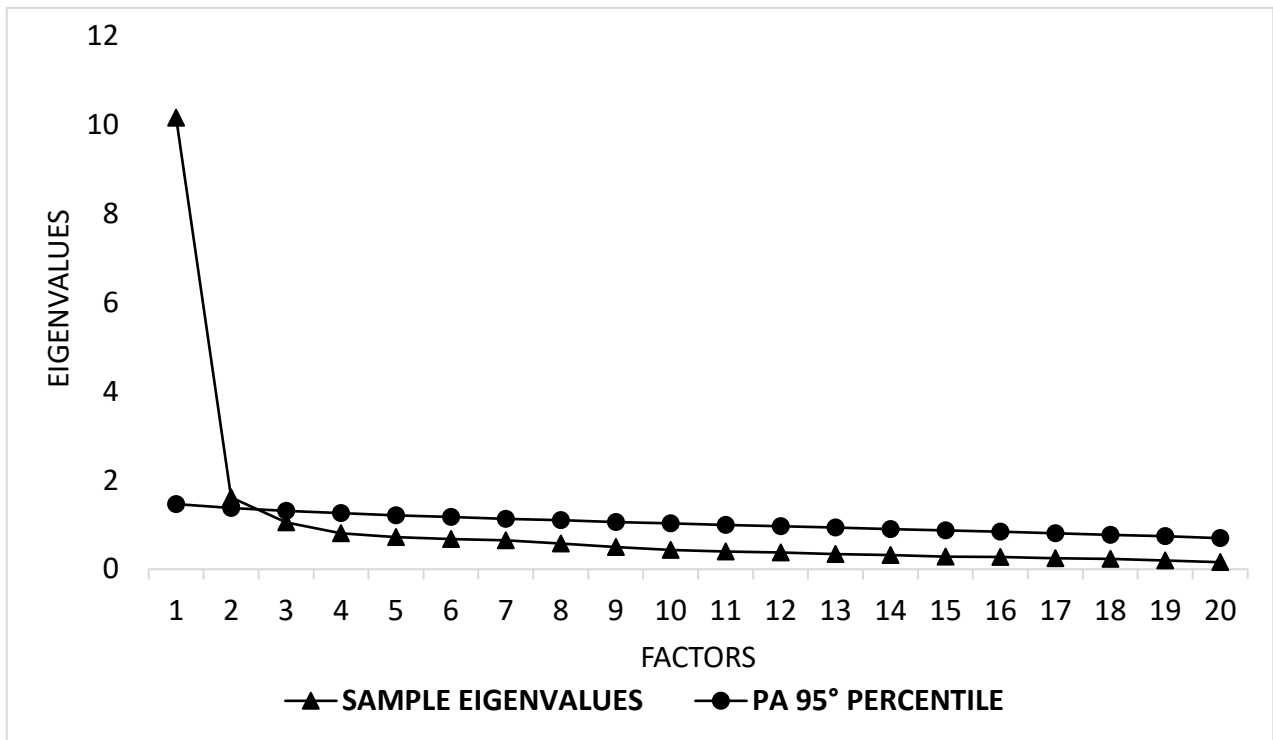


Grafico 2 *Parallel analysis della scala STAI-LONG*

Sulla base di questi risultati, sono state effettuate due EFA, una per la scala STAI-STD e una per la scala STAI-LONG, chiedendo due fattori come emerso dalla PA.

I risultati delle EFA hanno mostrato degli indici di fit mediocri ma la soluzione è risultata essere ben interpretabile. Gli indici per la scala standard erano: $\chi^2(151) = 606.242, p \cong .00$; RMSEA = .077 [.071, .084]; CFI = .892; SRMR = .04. Per la scala allungata, invece, gli indici di fit erano: $\chi^2(151) = 738.527, p \cong .00$; RMSEA = .094 [.088, .101]; CFI = .90; SRMR = .041.

Guardando la configurazione dei *factor loading* si è notato che per entrambe le scale emergevano due componenti: una contenente gli item formulati in modo positivo (i.e., reverse code) e una contenente gli item formulati in modo negativo (vedi **Tabella 1** e **Tabella 2**). Nello specifico, una componente è risultata essere costituita da 9 item strutturati con un orientamento positivo, che denotano assenza di ansia e assenza di

sensazioni spiacevoli (ad esempio “*Sono felice*”), ed una parte costituita da 11 item con un orientamento negativo, che descrivono presenza di ansia e sensazioni spiacevoli (ad esempio “*Mi sento teso e irrequieto*”).

Nella maggior parte degli item ci sono indici di saturazione elevati (compresi tra .44 e .92). Nella scala allungata emergono però degli item con *cross loading*: Item13 “*Mi sento sicuro*”, Item5 “*Mi sento un fallito*” e Item7 “*Sono calmo e padrone di me*”.

Tabella 1 EFA STAI-STD

	F1	F2
ITEM2	0.104*	0.697*
ITEM4	-0.040	0.665*
ITEM5	-0.004	0.821*
ITEM8	0.034	0.824*
ITEM9	-0.021	0.553*
ITEM11	0.095*	0.774*
ITEM12	-0.005	0.727*
ITEM15	-0.027	0.837*
ITEM17	-0.058	0.723*
ITEM18	0.051	0.622*
ITEM20	0.061	0.516*
ITEM1	0.735*	-0.007
ITEM3	0.689*	0.038
ITEM6	0.546*	-0.060
ITEM7	0.618*	0.003
ITEM10	0.801*	-0.013
ITEM13	0.648*	0.030
ITEM14	0.440*	-0.024
ITEM16	0.796*	0.003
ITEM19	0.356*	0.053
F1		
F2	0.013	

nota. * $p < 0.05$

F1 = fattore positivo

F2 = fattore negativo

L'azzurro indica gli item che saturano il primo fattore.

Il rosa indica gli item che saturano il secondo fattore.

Tabella 2 EFA STAI-LONG

	F1	F2
ITEM2	0.124	0.605*
ITEM4	0.229*	0.461*
ITEM5	0.357*	0.470*
ITEM7	0.374*	0.401*
ITEM8	0.103	0.689*
ITEM9	-0.170*	0.824*
ITEM11	0.066	0.759*
ITEM12	0.176*	0.664*
ITEM15	0.234*	0.645*
ITEM17	-0.159*	0.812*
ITEM18	0.004	0.747*
ITEM20	-0.099	0.864*
ITEM1	0.782*	0.048
ITEM3	0.637*	0.206*
ITEM6	0.441*	0.129
ITEM10	0.920*	-0.091
ITEM13	0.492*	0.350*
ITEM14	0.290*	0.261*
ITEM16	0.865*	0
ITEM19	0.434*	0.032
F1		
F2	0.686*	

nota. * $p < 0.05$

F1 = fattore positivo

F2 = fattore negativo

L'azzurro indica gli item che saturano il primo fattore.

Il rosa indica gli item che saturano il secondo fattore.

I modelli CFA hanno confermato la struttura a 2 fattori per entrambe le scale.

Nella versione standard il fit è: $\chi^2(168) = 589.342, p \cong .00$; RMSEA = .070 [.064, .077]; CFI = .91; SRMR = .049.

Nella **Tabella 3** si possono osservare i *factor loading* del modello CFA applicato alla scala STAI-STD, essi mostrano valori elevati (compresi tra .30 e .84). Nel modello, si è osservata una correlazione moderata e positiva tra i fattori latenti ($r = .040, p < .05$) ed una correlazione moderata e positiva ($r = .424, p < .05$) fra gli item 10 e 16 (evidenziata dagli indici di modifica).

Il modello è stato quindi stimato correlando i residui dell'Item16 "Sono contento" e dell'Item10 "Sono felice". Questi due item di fatto sono molto simili e correlano tra loro.

Tabella 3 CFA STAI-STD

	F1	F2
ITEM2	0.022	0.700
ITEM4	0.027	0.664
ITEM5	0.013	0.820
ITEM8	0.016	0.825
ITEM9	0.031	0.551
ITEM11	0.018	0.777
ITEM12	0.022	0.726
ITEM15	0.015	0.836
ITEM17	0.021	0.720
ITEM18	0.027	0.623
ITEM20	0.033	0.519
ITEM1	0.729	0.023
ITEM3	0.723	0.024
ITEM6	0.547	0.032
ITEM7	0.653	0.029
ITEM10	0.726	0.024
ITEM13	0.683	0.024
ITEM14	0.458	0.036
ITEM16	0.723	0.026
ITEM19	0.382	0.043
F1		
F2	0.040	
ITEM16		
ITEM10	con	0.424

nota. F1 = fattore positivo
F2 = fattore negativo
L'azzurro indica gli item che saturano il primo fattore.
Il rosa indica gli item che saturano il secondo fattore.

Tabella 4 CFA STAI-LONG

	F1	F2
ITEM2	0.024	0.685
ITEM4	0.033	0.642
ITEM5	0.020	0.765
ITEM8	0.028	0.772
ITEM9	0.028	0.666
ITEM11	0.020	0.801
ITEM12	0.016	0.807
ITEM15	0.015	0.836
ITEM17	0.029	0.666
ITEM18	0.025	0.732
ITEM20	0.021	0.767
ITEM2	0.024	0.685
ITEM1	0.781	0.021
ITEM3	0.825	0.018
ITEM6	0.542	0.039
ITEM10	0.748	0.024
ITEM13	0.806	0.022
ITEM14	0.524	0.044
ITEM16	0.779	0.024
ITEM19	0.473	0.041
ITEM7	0.740	0.024
F1		
F2	0.845	
ITEM10		
ITEM16	con	0.558

nota. F1 = fattore positivo
F2 = fattore negativo
L'azzurro indica gli item che saturano il primo fattore.
Il rosa indica gli item che saturano il secondo fattore.

Per la scala allungata si sono ottenuti degli indici di fit simili a quelli della scala standard: $\chi^2(168) = 683.787$. $p \cong .00$; RMSEA = .084 [.077. .090]; CFI = .90; SRMR = .052.

In **Tabella 4** si possono osservare i *factor loading* ottenuti applicando il modello a due fattori. Anche in questo caso i valori sono molto elevati. Si osserva inoltre una correlazione elevata e positiva ($r = .588$, $p < .05$) fra gli item 10 e 16.

Riassumendo, dalle analisi fattoriali emerge che vi è un fattore negativo composto da 11 item: 2 “*Mi sento teso e irrequieto*”, 4 “*Vorrei poter essere felice come sembrano essere gli altri*”, 5 “*Mi sento un fallito*”, 8 “*Sento che le difficoltà si accumulano tanto da non poterle superare*”, 9 “*Mi preoccupo troppo di cose che in realtà non hanno importanza*”, 11 “*Mi vengono pensieri negativi*”, 12 “*Manco di fiducia in me stesso*”, 15 “*Mi sento inadeguato*”, 17 “*Pensieri di scarsa importanza mi passano per la mente e mi infastidiscono*”, 18 “*Vivo le delusioni con tanta partecipazione da non poter togliermele dalla testa*”, 20 “*Divento teso e turbato quando penso alle mie attuali preoccupazioni*”, e un fattore positivo composto da 9 item: 1 “*Mi sento bene*”, 3 “*Sono soddisfatto di me stesso*”, 6 “*Mi sento riposato*”, 7 “*Io sono calmo. tranquillo e padrone di me*”, 10 “*Sono felice*”, 13 “*Mi sento sicuro*”, 14 “*Prendo decisioni facilmente*”, 16 “*Sono contento*”, 19 “*Sono una persona costante*”.

Visto che il PCM richiede monodimensionalità, si è testata l'unifattorialità delle due dimensioni emerse dalle scale. Per fare ciò si è usata la PA e la CFA. I risultati hanno mostrato che le due dimensioni, positiva e negativa, sono entrambe monodimensionali in entrambe le scale (STAI-STD e STAI-LONG). (**Grafico 3 e Grafico 4**)

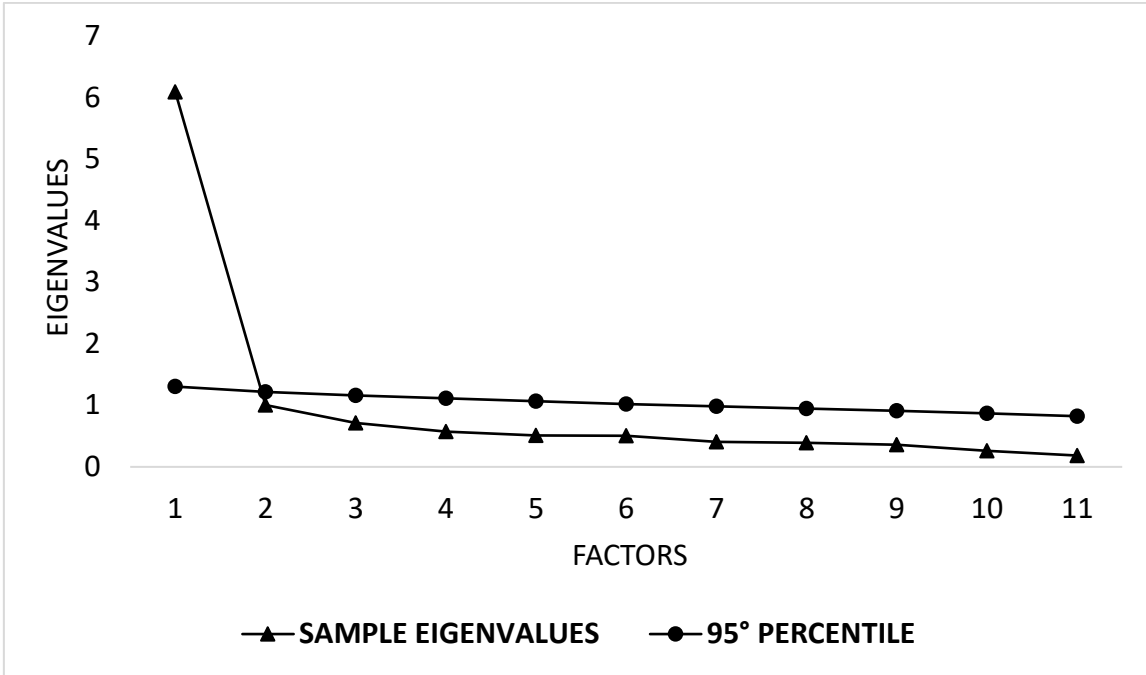


Grafico 3 *Monofattoriale STAI-STD: fattore negativo*

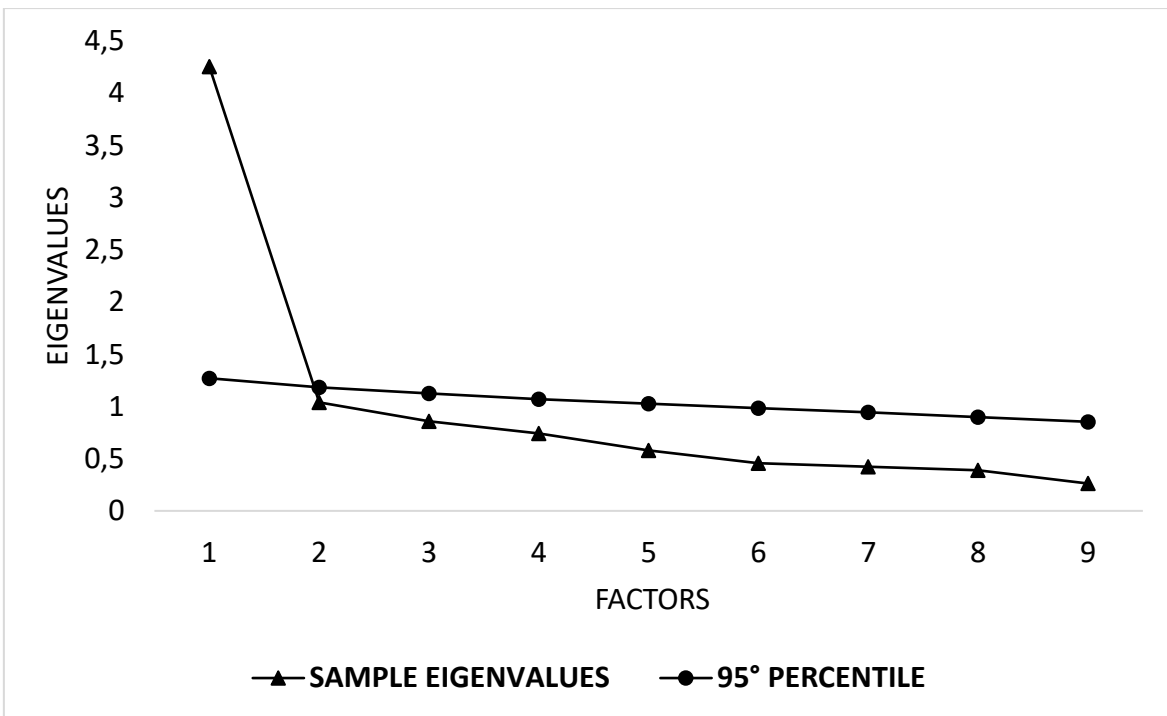


Grafico 4 *Monofattoriale STAI-STD: fattore positivo*

Risultato analogo si ottiene anche analizzando la monofattorialità dei fattori della scala allungata (**Grafico 5** e **Grafico 6**).

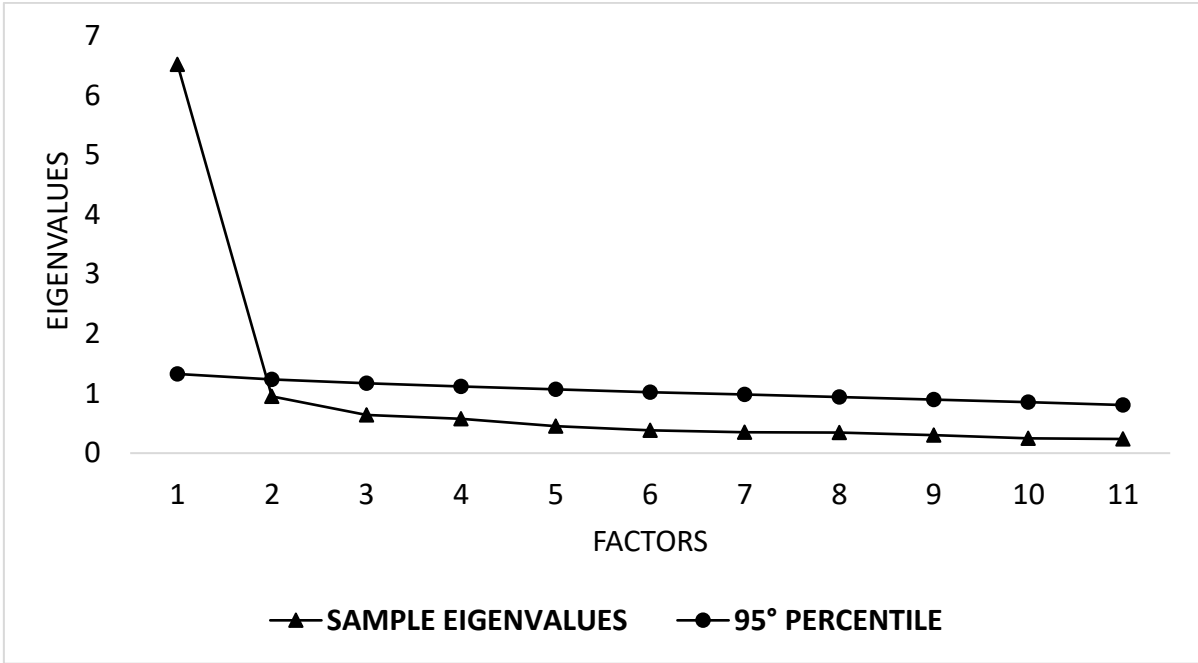


Grafico 5 *Monofattoriale STAI-LONG: fattore negativo*

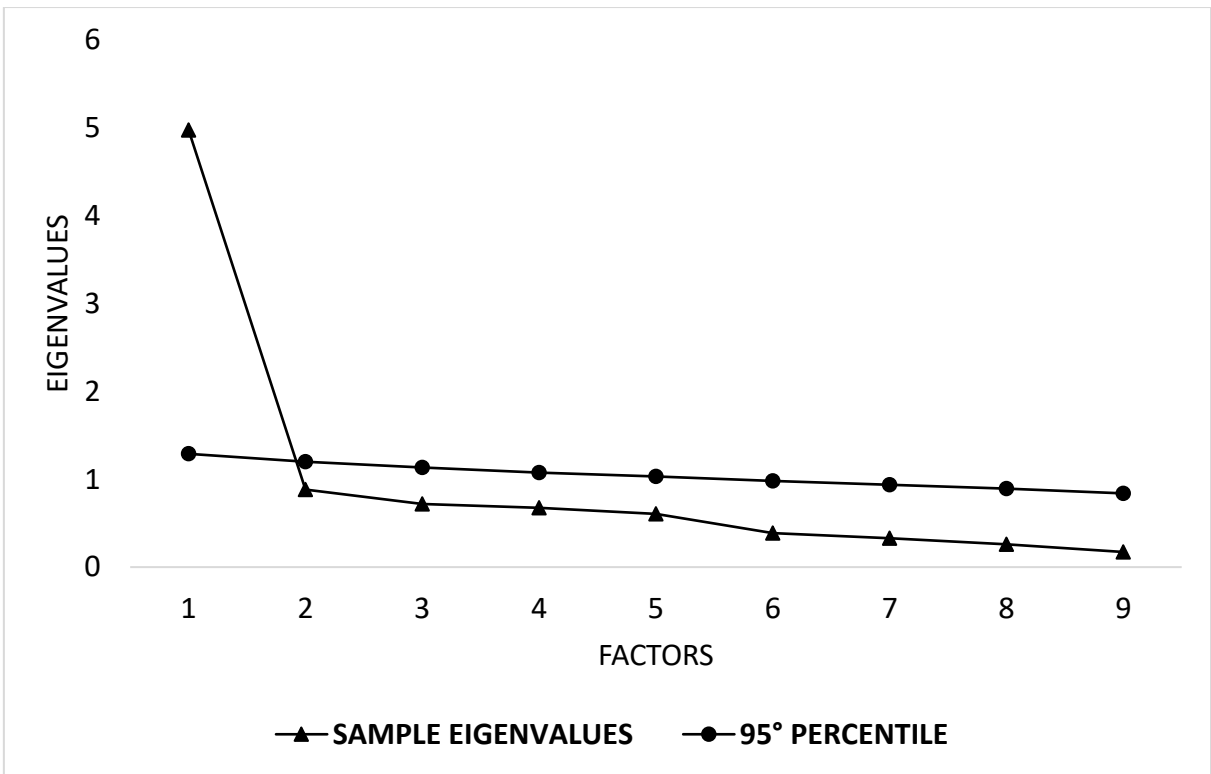


Grafico 6 *Monofattoriale STAI-LONG: fattore positivo*

Una volta accertata l'unidimensionalità, è stato stimato il modello PCM per le due dimensioni di ogni versione dello strumento: STAI-STD e STAI-LONG (vedi **Tabella 5 e Tabella 6**).

Relativamente alla scala STAI-STD si è osservato che le soglie sono ordinate, il che suggerisce che la scala a 4 punti è adeguata a rispondere agli item del questionario STAI. Questo si evince dal fatto che nello STAI-STD solo un item ha delle soglie disordinate, ovvero l'Item6 della scala positiva "*Mi sento riposato*". Nello specifico le soglie τ 1-2 e τ 2-3 sono tra loro invertite, ciò indica che i soggetti nel rispondere all'item non distinguono bene le categorie "a volte" e "spesso". Inoltre, l'*infit* e l'*outfit* sono tutti al di sotto del *cut-off* di 2. Questo suggerisce che gli item sono chiari e che contribuiscono a definire un costrutto essenzialmente monodimensionale. La consistenza interna delle dimensioni della scala standard è risultata essere molto buona sia con R che con α ($R = .85$. α di Cronbach = $.85$ per la scala positiva; $R = .90$. α di Cronbach = $.92$ per la scala negativa).

Per quanto riguarda invece la scala STAI-LONG si è osservato che le soglie sono tutte disordinate, sia nella componente positiva che in quella negativa. Questo disordine delle soglie significa sostanzialmente che le risposte con punteggio più alto (la scelta di una categoria alta di risposta) non riflettono necessariamente maggiori livelli di tratto. Ne consegue dunque che la scala a 8 punti non è adeguata a rispondere agli item del questionario in questione. Anche per questa scala sono emersi *infit* e *outfit* tutti al di sotto del *cut-off* di 2 e la consistenza interna è risultata essere buona ($R = .89$. α di Cronbach = $.89$ per la scala positiva; $R = .93$. α di Cronbach = $.93$ per la scala negativa).

Tabella 5 *Infit, Outfit e Soglie per ogni item della scala STAI-STD (positiva e negativa)*

Item	Infit	Outfit	τ_{1-2}	τ_{2-3}	τ_{3-4}
SCALA POSITIVA					
1	0.724	0.733	1.094	-0.704	-4.589
3	0.748	0.749	1.631	0.794	-2.137
6	1.074	1.062	2.745	2.957	0.762
7	0.867	0.876	1.721	1.083	-1.510
10	0.683	0.676	1.123	-0.475	-3.732
13	0.767	0.766	1.792	1.200	-1.720
14	1.226	1.249	1.986	1.726	-0.348
16	0.670	0.662	1.302	-0.535	-4.408
19	1.374	1.428	1.126	-0.080	-2.805
SCALA NEGATIVA					
2	0.889	0.894	1.546	1.227	-1.727
4	1.067	1.071	0.778	0.473	-1.067
5	0.850	0.788	-0.562	-0.856	-1.538
8	0.717	0.711	0.579	0.092	-1.644
9	1.221	1.261	1.349	0.770	-1.447
11	0.717	0.706	1.743	1.480	-0.776
12	0.891	0.899	0.882	0.539	-1.572
15	0.687	0.657	0.402	0.020	-1.995
17	0.868	0.871	0.865	0.416	-1.819
18	1.075	1.054	0.985	0.604	-1.313
20	1.240	1.214	1.581	1.183	-1.199

Nota. τ_{1-2} , τ_{2-3} e τ_{3-4} sono rispettivamente le soglie tra le categorie “quasi mai” e “a volte”, “a volte” e “spesso”, “spesso” e “quasi sempre”. I colori nella tabella rappresentano i valori delle soglie. Il verde indica valori elevati delle soglie, mentre i colori progressivamente degradanti verso il rosso indicano il progressivo ridursi dei valori delle soglie.

Tabella 6 *Infit, Outfit e soglie per ogni item della scala STAI-LONG (positiva e negativa)*

Item	Infit	Outfit	τ_{1-2}	τ_{2-3}	τ_{3-4}	τ_{4-5}	τ_{5-6}	τ_{6-7}	τ_{7-8}
SCALA POSITIVA									
1	0.645	0.634	3.558	3.295	3.742	2.711	-0.158	-2.548	-7.237
3	0.650	0.643	1.760	1.398	1.838	1.197	-0.831	-2.713	-6.028
6	1.164	1.165	1.760	1.398	1.838	1.197	-0.831	-2.713	-6.028
7	0.877	0.878	3.399	4.092	5.208	5.172	3.776	2.266	-1.095
10	0.693	0.689	2.027	1.867	2.449	1.794	-0.026	-1.969	-4.221
13	0.688	0.704	1.814	1.761	2.225	1.828	0.025	-1.916	-5.217
14	1.336	1.366	1.453	2.123	2.387	1.905	0.236	-1.476	-3.948
16	0.607	0.605	1.896	1.751	1.964	1.036	-1.643	-4.623	-8.974
19	1.387	1.390	0.896	0.635	0.868	-0.258	-2.703	-4.779	-7.808
SCALA NEGATIVA									
2	0.965	0.965	2.328	2.426	3.615	2.414	1.006	-0.814	-3.999
4	1.298	1.412	0.622	0.729	1.417	0.908	-0.219	-1.657	-2.920
5	0.963	0.922	-0.172	-0.727	-0.974	-2.797	-4.245	-6.230	-8.413
8	0.834	0.896	0.971	0.973	1.330	0.160	-1.333	-3.174	-5.201
9	1.061	1.097	1.197	1.404	2.149	1.360	-0.023	-1.473	-3.553
11	0.744	0.730	1.670	1.776	2.385	1.454	0.315	-1.248	-2.747
12	0.831	0.826	1.169	1.394	2.158	1.367	0.367	-1.509	-2.702
15	0.706	0.720	1.246	0.989	1.881	0.955	-0.690	-2.235	-4.118
17	1.030	1.021	1.560	1.879	2.383	1.222	-0.395	-1.990	-3.554
18	0.916	0.913	1.094	1.381	2.064	1.551	0.075	-1.503	-2.975
20	0.762	0.755	2.242	2.722	3.859	3.733	2.499	1.412	-0.219

Nota. τ_{1-2} , τ_{2-3} , τ_{3-4} , τ_{4-5} , τ_{5-6} , τ_{6-7} e τ_{7-8} sono rispettivamente le soglie comprese tra le categorie “mai” e “quasi mai”, “quasi mai” e “raramente”, “raramente” e “a volte”, “a volte” e “spesso”, “spesso” e “molto spesso”, “molto spesso” e “quasi sempre”, “quasi sempre” e “sempre”. I colori nella tabella rappresentano i valori delle soglie. Il verde indica valori elevati delle soglie, mentre i colori progressivamente degradanti verso il rosso indicano il progressivo ridursi dei valori delle soglie.

4.4 INTEPRETAZIONE E DISCUSSIONE DELLO STUDIO

Nel presente lavoro si è voluta verificare l'ipotesi implicita nelle *rating scales* secondo la quale le categorie di risposta degli item riflettono i livelli della variabile latente nei soggetti. Dunque, a risposte più alte nella scala dovrebbero corrispondere livelli crescenti di tratto. Inoltre, attraverso l'uso del PCM si è voluto verificare e confrontare una versione standard di scala di risposta con una versione allungata creata appositamente per i fini della ricerca.

Spesso in letteratura si tende, erroneamente, a pensare che un metodo efficace per massimizzare la raccolta di informazioni nei questionari sia quello di utilizzare un numero inferiore di item (per non stancare il rispondente) e un numero maggiore di categorie di risposta. L'obiettivo di questo lavoro di ricerca è però quello di mostrare che non sempre l'utilizzo di una categoria di risposta ampia offre risultati più attendibili. Il presente studio vuole mostrare questo aspetto confrontando tra loro due diverse scale di risposta di uno stesso questionario.

Nel caso specifico in esame si è deciso di prendere in considerazione come strumento il questionario STAI, composto da 20 item che valutano i sentimenti di ansia generale e la predisposizione relativamente stabile a vedere le situazioni stressanti come minacciose. Una versione originale con 4 categorie di risposta (STAI-STD) e l'altra creata appositamente ai fini della ricerca con 8 categorie di risposta (STAI-LONG).

Per raggiungere gli obiettivi di ricerca sono state condotte le seguenti analisi: PA, EFA e CFA, e infine è stato applicato il PCM per analizzare le soglie.

In primo luogo, si è analizzata la dimensionalità delle due versioni dello STAI attraverso PA (Horn, 1965), EFA e CFA. Entrambe le versioni utilizzate della scala STAI sono

risultate essere bifattoriali. Le analisi hanno mostrato come vi siano due fattori soggiacenti alle scale: uno costituito dagli item con orientamento positivo e uno costituito dagli item con orientamento negativo. Il fattore negativo comprende 11 item e quello positivo 9.

Infine, dopo aver verificato la monodimensionalità delle scale come requisito dei modelli di Rasch, è stato possibile applicare il PCM.

Analizzando ogni singolo item si è visto che: nella standard le soglie sono tutte ordinate (tranne l'item6 "*Mi sento riposato*") e l'attendibilità è buona. Invece nella scala allungata, le soglie sono tutte disordinate. Il mancato ordine delle soglie (condizione denominata disordine delle soglie) fornisce la prova che la scala di valutazione non funziona come previsto. Tale disordine è stato osservato per tutti gli item della scala estesa a 8 punti. Questo suggerisce che la variabile latente è scarsamente definita nella mente degli intervistati, o che agli intervistati sono state presentate più categorie di quelle che sono in grado di percepire nel rispondere agli item.

Soffermandosi sull'interpretazione dell'attendibilità di queste due scale è possibile, confrontando i valori di R e di α di Cronbach, riscontrare un aumento di tali indici nella scala STAI-LONG rispetto ai valori della scala STAI-STD. Nello specifico i valori passano da una media di .90 a una di .92. Tale aumento potrebbe portare erroneamente a pensare che effettivamente la scala allungata consenta la raccolta di maggiore informazione. Ma questo miglioramento non è così elevato da poter giustificare un completo disordine di soglie. Dunque, risulta evidente che lo strumento nella sua versione allungata non misura il costrutto latente meglio di quella standard, ma al contrario misura male quello che dovrebbe misurare.

Analizzando più nel dettaglio l'ordine delle soglie nella scala STAI-LONG, si osserva che alcune categorie sono più difficili da interpretare rispetto alle altre. Le soglie τ_{1-2} e τ_{2-3} , ad esempio, sono quelle dove si osserva maggiore confusione e disordine. Queste soglie corrispondono alle categorie di risposta “mai” – “quasi mai” e “quasi mai – raramente”. Seguite poi dalla soglia τ_{4-5} che corrisponde all'intervallo “a volte” – “spesso”. Si evince dunque che la soglia “quasi mai” e “raramente” non è quasi mai interpretata correttamente. L'unica categoria di risposta che sembra essere interpretata quasi sempre in modo corretto è l'ultima τ_{7-8} , che corrisponde alla categoria “sempre”.

Se ci soffermiamo sugli item della scala STAI-LONG vediamo che l'item che presenta il miglior ordine delle soglie è il 5 “*Mi sento un fallito*”. Al contrario gli item che funzionano meno bene sono il 10 “*Sono felice*”, il 14 “*Prendo decisioni facilmente*”, e il 18 “*Vivo le delusioni con tanta partecipazione da non poter togliermele dalla testa*”, solo per citarne alcuni. Gli item che possono essere definiti come molto difficili sono il 7 “*Io sono calmo, tranquillo e padrone di me*” e il 20 “*Divento teso e turbato quando penso alle mie attuali preoccupazioni*”. Questi item hanno delle risposte che non corrispondono per nulla a quelle attese.

Se osserviamo invece gli item della scala STAI-STD vediamo che vi è un solo item che non funziona bene, ovvero l'item 6 “*Mi sento riposato*”. Nello specifico, si presenta una sola soglia disordinata in corrispondenza delle soglie τ_{1-2} (“quasi mai” – “a volte”) e τ_{2-3} (“a volte” – “spesso”).

Questi risultati offrono una chiara risposta al nostro quesito: ovvero che una scala di risposta più ampia non necessariamente offre una misura migliore e più informativa, ma

anzi, come si evince in questo caso, potrebbe influire negativamente sulla qualità della misurazione.

In questo lavoro di ricerca risulta essere evidente che la scala STAI-LONG offre una misurazione peggiore rispetto a quella della STAI-STD, in quanto crea confusione nei rispondenti. Questa alla luce del fatto che la versione allungata presenta tutte soglie disordinate, mentre la versione standard una sola. Emerge dunque l'importanza di non dare per scontato il funzionamento della scala di rating, ma di verificarlo empiricamente.

4.4.1 LIMITI

Il presente lavoro di ricerca presenta alcuni limiti. Uno di questi è il fatto che il campione a cui è stato somministrato il questionario non si può definire omogeneo in termini di genere, in quanto la percentuale delle donne è maggiore rispetto a quella degli uomini sia nella versione standard che in quella allungata.

Un ulteriore limite, intrinsecamente connesso alla somministrazione dei questionari self-report, è che i rispondenti potrebbero mentire su quello che provano o sulla propria situazione attuale.

Inoltre, all'interno delle analisi statistiche condotte, risultano esserci alcuni item della scala allungata (gli item 5, 7 e 13) che nell'analisi fattoriale mostrano dei *cross loading* (come si è visto in precedenza nella **Tabella 2**) e quindi potrebbero risultare dubbi.

4.4.2 SVILUPPI FUTURI

Questo lavoro ha permesso di mettere in luce alcuni aspetti su cui riflettere sia riguardanti lo strumento in sé, sia il testing più in generale.

Nel presente lavoro di ricerca è emerso che il questionario STAI presenta una bifattorialità della scala. È dunque possibile individuare degli sviluppi futuri che potrebbero verificare

ed approfondire quanto emerso dalle analisi condotte. Ad esempio, si potrebbe verificare se questa bifattorialità sia il frutto di un effetto di metodo nella costruzione degli item (come la presenza di *item reverse*, piuttosto che la presenza di componenti differenti dello stesso costrutto).

Per quanto riguarda la ricerca di per sé, un possibile sviluppo futuro potrebbe riguardare la sensibilizzazione degli sviluppatori di test rispetto a questa tematica che è nota ma spesso trascurata. Aumentare la consapevolezza e richiamare all'attenzione su questo possibile "bias" in quegli strumenti che sono costituiti da scale di risposta ampie.

Ancora, uno sviluppo futuro potrebbe essere quello di analizzare uno strumento che anziché avere delle categorie di risposta, abbia una sorta di continuum numerico in cui collocarsi tra un minimo e un massimo. In questo caso non si avrebbe più il problema di definire in quale categoria ci si riconosce. Sarebbe interessante verificare come si comporterebbe lo STAI in questa situazione, non avendo più categorie ma un continuum.

4.5 CONCLUSIONI

L'elaborato nasce con l'obiettivo di mostrare il funzionamento delle *rating scales* verificandolo empiricamente. Per svolgere questo lavoro si è partiti da un'attenta analisi della letteratura psicologica sullo sfondo teorico dei tratti di personalità e del costrutto di ansia. In primo luogo, sono state analizzate e descritte le teorie della personalità che si sono susseguite nel tempo. È stato poi approfondito il costrutto d'interesse, ovvero l'ansia. Una volta definito il costrutto, sono state elencate le diverse classificazioni dei disturbi d'ansia, i principali sintomi e alcune delle loro implicazioni sulla vita delle persone. Dopodiché sono stati descritti alcuni metodi di misurazione del costrutto

dell'ansia: il *Beck Anxiety Inventory* (BAI), l'*Ender Multidimensional Anxiety Scales* (EMAS) e lo *State-Trait Anxiety Inventory* (STAI).

In seguito, è stato necessario un importante approfondimento in merito alla misurazione in psicologia, procedura essenziale nel processo diagnostico e nella ricerca più in generale. Dopo alcune definizioni, e un breve cenno storico sui primi tentativi di misurazione in psicologia, sono state descritte le norme metriche necessarie per la costruzione e il corretto uso dei test psicologici. Sono quindi stati introdotti i concetti di affidabilità e validità fondamentali per definire la bontà di uno strumento e ne sono stati brevemente descritti i metodi di calcolo.

Infine, è stata descritta l'*Item Response Theory* per meglio comprendere alcune delle teorie sulla costruzione di strumenti di assessment psicologico. Questa teoria presenta un grande potere innovativo e delle ottime potenzialità applicative. Sono stati passati in rassegna alcuni dei modelli IRT, come quello di Rasch ed alcune sue estensioni.

Tutte queste informazioni reperite dalla letteratura sono state di fondamentale importanza nel condurre la ricerca e perseguire gli obiettivi. Attraverso l'applicazione del *Partial Credit Model* è stato possibile verificare come una versione allungata dello STAI, creata appositamente ai fini della ricerca, offra una misurazione peggiore rispetto alla scala standard, in quanto crea confusione nei rispondenti. I risultati ottenuti consentono dunque di confermare l'ipotesi iniziale che non sempre l'utilizzo di una categoria di risposta ampia offra risultati più attendibili. Questo perché lo sforzo nel collocarsi in una tra tante categorie di risposta potrebbe essere maggiore rispetto a quello di rispondere ad un item in più, come appunto si è visto in questo lavoro analizzando l'ordine delle soglie Rasch-Andrich.

Il presente lavoro di ricerca offre un importante spunto di riflessione sul testing psicologico e al tempo stesso permette di arricchire il materiale di ricerca presente in letteratura. Il testing psicologico è presente quotidianamente nelle vite dei professionisti che prendono decisioni sulla vita delle persone. Motivo per il quale è di fondamentale importanza riflettere e prestare la massima attenzione nella costruzione delle scale psicologiche, al fine di realizzare buoni strumenti, per offrire il miglior servizio possibile alle persone che lo richiedono. Nella costruzione dei test psicologici è quindi un dovere avere cura nel costruire gli item, ma anche ragionare accuratamente sulle scale di risposta.

Ulteriori riflessioni su quanto emerso potrebbero riguardare anche altri contesti, come ad esempio quello del marketing. Si pensi ad esempio ad una situazione in cui si vuole raccogliere l'opinione di consumatori rispetto ad un prodotto/servizio tramite delle interviste. È estremamente importante non fare troppa confusione sulle opzioni di risposta, perché altrimenti la rilevazione potrebbe risultare inefficace per il business.

Concludendo, possiamo affermare che sono stati raggiunti gli obiettivi del presente lavoro di ricerca. Più precisamente, l'obiettivo di questo lavoro era quello di sottolineare l'importanza di porre attenzione non solo alla scelta e alla creazione degli item, ma anche alla creazione delle adeguate categorie di risposta. Questa importante riflessione ci deve ricordare che per ottenere una rilevazione valida non è sufficiente avere degli ottimi item, ma è necessario anche avere delle categorie di risposta adeguate.

BIBLIOGRAFIA

- Abbott-Jones, A. (2021). A quantitative study identifying the prevalence of anxiety in dyslexic students in higher education. *Research Journal of Education*, 42-51.
- Allport, G. O. (1936). *Trait-names: A Psycho-lexical Study, Volume 47, Edizione 1*. Psychological Review Company.
- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.
- Allport, G. W. (1962). The general and the unique in psychological science. *Journal of personality*, 30(3), 405-422.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-57.
- Andrich, D. (1982). An index of person separation in latent trait theory, the traditional KR 20 index and the Guttman scale response pattern. *Educational research and perspectives*, 9:95-104.
- Anselmi, P. F. (2018). Comparison of four common data collection techniques to elicit preferences. *Quality & Quantity*, 52(3), 1227–1239. .
doi:<https://doi.org/10.1007/s11135-017-0514-7>
- Antonelli, E. R. (2017). *La valutazione della personalità attraverso i test. Appunti di Teorie e tecniche dei test di personalità*. Venezia: Upsel Domeneghini Editore.
- APA. (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, suppl 51, part.2, pp 1-38.
- APA. (1995). *Diagnostic and statistical manual of mental disorders (DSM-IV)*. Washington, D.C : American Psychiatric Association .
- APA. (2016, January 28). *Causes and Effects of Anxiety*. Retrieved from Particular exciting experiments in Psychology (APA PeePs):
<https://www.apa.org/pubs/highlights/peeps/issue-62>
- APA. (2023, Marzo 16). *APA Dictionary of Psychology*. Retrieved from American Psychiatric Association:
https://dictionary.apa.org/anxiety?_ga=2.163970175.1172828578.1678891776-859166977.1614286034
- APA, A. P. (2014). *Manuale diagnostico e statistico dei disturbi mentali quinta edizione DSM-5*. (M. Biondi, Ed.) Milano: Raffaello Cortina Editore.
- Ashton, M. (2023). *Individual differences and personality*. Londra: Elsevier.
- Ashton, M. C. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychological Review*, 11, 150-166.
- Ashton, M. C. (2023). Chapter 3 - Personality Structure Classifying Traits. In M. C. Ashton, *Individual Differences and Personality (Fourth Edition)* (pp. 63-87). Academic Press.
- Baker, F. (2001). *The basics of item response theory. Second Edition*. Washington, DC: ERIC clearinghouse on assessment and improvement .

- Baumgarten, F. (1937). Die charaktereigenschaften. *The Journal of Nervous and Mental Disease*, 86 (1), 113.
- Beck, A. E. (1988). An inventory for measuring clinical anxiety: psychometric properties. *J Consult Clin Psychol.*, 56:893-7.
- Beck, A. S. (1991). Relationship between the Beck Anxiety Inventory and the Hamilton Anxiety Rating Scale with anxious outpatients. *J Anx Disord*, 5:213–23.
- Bennet, E. (1946). The anxiety state. *POST-GRADUATE MEDICAL JOURNAL Psychiatrist West End Hospital for Nervous Diseases; Consulting Physician Tavistock Clinic, London.*, 375 – 378.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- Bentler, P. M. (1995). *EQS structural equations program manual*. . Encino, CA: Multivariate Software.
- Bock, R. a. (2010). Factor analysis of categorical item responses. In M. N. Ostini, *Handbook of Polytomous Item Response Theory Models* (pp. 155–184). Florence, KY: Lawrence Erlbaum.
- Bollen, K. A. (1981). Pearson's r and coarsely categorized measures. *American Sociological Review*, 46(2), 232–239.
- Boncori, L. (2006). *I test in psicologia. Fondamenti teorici e applicazioni*. Bologna : Il Mulino.
- Boncori, L. (2006). *I test in psicologia. Fondamenti teorici e applicazioni*. Bologna: Il Mulino.
- Browne, M. W. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: SAGE.
- Cai, L. C. (2016). Item Response Theory. *Annual Review of Statistics and its applications*, 297-321.
- Campbell, D. T. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. . *Psychological Bulletin*, 56, 81-105.
- Caprara, G. C. (2002). *Personalità. Determinanti, dinamiche, potenzialità*. Cortina Raffaello.
- Caprara, G. V.-B.-B. (2005). *Big Five Questionnaire*. Firenze: Giunti O.S. .
- Cattell I.H, S. R. (1958). The nature of anxiety: A review of thirteen multivariate analysis comprising 814 variables. *Psychological reports*, 4, 351-388.
- Cattell, R. B. (1947). Confirmation and clarification of primary personality factors. *Psychometrika*, 12, 197–220.
- Colledani, D. A. (2022). Rasch Models in the Analysis of Repgrid Data. *Journal of Constructivist Psychology* , 35:2, 605-625.
- Cronbach, L. J. (1970). *Essentials of psychological testing*. New York: Harper and Row, 3th ed.
- Donovan, G. M. (2000). Personality and job performance: The Big Five revisited. . *Journal of Applied Psychology*, 85, 869-879.

- DuBois, P. (1970). *A history of psychological testing*. Boston: Allyn & Bacon.
- Embretson, S. e. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Emilien, G. D. (2002). Normal and pathological anxiety. In Basel, *Anxiety Disorders*. Verlag: Birkhäuser. doi: https://doi.org/10.1007/978-3-0348-8157-9_1
- Endler, N. &. (2001). State and trait anxiety revisited. *Anxiety disorders* 15, 231-245.
- Endler, N. ., (1962). An S-R inventory of anxiousness. *Psychological Monographs*, 76,17.
- Endler, N. E. (1991). Endler Multidimensional Anxiety Scales (EMAS). *Psychological Bulletin*, 83, 956-974.
- Endler, N. M. (1976). *Interactional psychology and personality*. New York: Hemisphere.
- Eysenck, H. (1957). *The dynamics of anxiety and hysteria*. NY: Praeger.
- Eysenck., H. J. (1960). *The structure of human personality*. Londra: Methuen.
- Faraci, P. (2009). L'Item Response Theory nella costruzione degli strumenti . *ITEMS - La newsletter del testing psicologico Giunti O.S.*, 1-6.
- Fischer, W. (1970). *Theories of anxiety*. NY & London: Harper & Row.
- Freud, S. (1917). Introductory lectures 011 psychoanalysis 1975. In S. Freud, *The standard edition of the complete psychological works of Sigmund Freud*. London: Hogarth Press.
- Freud, S. (1924). *A general introduction to psychoanalysis*. New York: Washington square press.
- Freud, S. (1926). Inhibitions, symptoms and anxiety. 1975. In S. Freud, *The standard edition of the complete psychological works of Sigmund Freud*. . London: Hogarth Press.
- Furr, R. M. (1998). A multimodal analysis of personal negativity . *Journal of Personality and Social Psychology*, 74, 1580-1591.
- Galimberti, U. (2005). *Enciclopedia di psicologia*. Milano: Garzanti.
- Galton, F. (1884). Measurement of character. *Fortnightly Review*, 36, 179-185.
- Ghadirian, A. M. (1981, July). Clinical Diagnosis of Anxiety. *Can Fam Physician* Vol.27, 1123-1126.
- Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler, *Review of Personality and Social Psychology* Vol. 2 (pp. 141-165). Beverly Hills : Sage.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. In APA, *American Psychologist* (pp. Vol. 48, 1, 26-34). doi:10.1037//0003-066x.48.1.26.
- Gottfredson, M. R. (1990). *A general theory of crime*. . Standford, CA: Standford University Press.
- Gray, J. (1982). *The Neuropsychology of anxiety: an enquiry into the functions of the septo-hippocampal system*. Oxford: Oxford University Press.

- Hambleton, R. e. (1985). *Item response theory: Principles and applications*. . Boston: Kluwer-Nijhoff.
- Harvey, R. H. (1999). Item response theory. *The Counseling Psychologist Vol. 27*, 353-383.
- Ho, R. G. (2012). An empirical evaluation of the slip correction in the four parameter logistic models with computerized adaptive testing. *Applied Psychological Measurement*, 2, 75-87.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- John, O. P.-P. (1997). *La scienza della personalità*. . Milano: Raffaello Cortina.
- Jokela, M. J. (2020). Personality, disability-free life years, and life expectancy: Individual participant meta-analysis of 131,195 individuals from 10 cohort studies. *Journal of Personality*, 596-605.
- Julian, J. (2011). Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). *Arthritis care & research*, 63 Suppl 11(0 11), S467–S472.
- Kenneth H. Craik, R. H. (1993). *Fifty Years of Personality Psychology*. New York: Plenum Press.
- Kirkegaard, S. (1844). *The concept of dread. 2nd ed.* Princeton: Princeton university press.
- Kline, P. (1993). *The handbook of psychological testing*. London: Routledge.
- Kraepelin, E. (1892). *Über die Beeinflussung einfacher psychischer Vorgänge durch einige Arzneimittel*. Jena: Fischer.
- Lee, K. M. (2009). Similarity and assumed similarity in personality reports of well-acquainted persons. *Journal of Personality and Social Psychology*, 460-472.
- Linacre, J. M. (1999). Category disordering (disordered categories) vs. threshold disordering (disordered thresholds). *Rasch Measurement Transactions*, , 13(1), 675. .
- Linacre, J. M. (2002a). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*,, 3(1), 85–106.
- Linacre, J. M. (2002b). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16(2), 878.
- Lord, F. N. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Mair, P. H. (2016). *Package ‘eRm’*. Vienna, Austria: R Foundation.
- Malouff, J. M. (2010). The Five-Factor Model of personality and relationship satisfaction of intimate partners: A meta-analysis. *Journal of Research in Personality*, 124-127.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. . *Psychometrika*, 47,149–174. Retrieved from <http://dx.doi.org/10.1007/BF02296272>
- Maydeu-Olivares, A. F. (2017). Goodness of fit in item factor analysis: Effect of the number of response alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*,, 24(4), 495–505.

- McCrae, R. R. (2003). *Personality in adulthood: A Five-Factor Theory perspective (2nd ed.)*. New York: Guildford Press.
- Muraki, E. (1992). A generalized partial credit model: application of an EM algorithm. *Appl. Psychol. Meas.*, 16:159-76.
- Muthén, B. O. (2012). *Mplus Version 7: User's guide*. Los Angeles, CA: Author.
- Muthén, L. K. (2016). *Mplus. Version 7.4[Computersoftware]*. Los Angeles, CA: Muthén & Muthén.
- Nolen-Hoeksema, S. F. (2014). *Atkinson & Hilgard's Introduction to Psychology - 16th edition*. Cengage Learning EMEA.
- Novikova, I. (2013). Trait, Trait Theory. In K. D. Keith, *The Encyclopedia of Cross-Cultural Psychology, First Edition* (pp. 1-2). John Wiley & Sons, Inc.
- Ohman, A. (1993). Fear and anxiety as emotional phenomena: clinical phenomenology, evolutionary perspectives, and information-processing mechanisms. In M. L. Haviland, *Handbook of the emotions*. NY & London : The Guilford Press.
- Oliver P. John, R. W. (n.d.). *Gordon Allport. Father and Critic of the Five-Factor Model*.
- Pallant, J. T. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British journal of clinical psychology*, 46, 1-18.
- Paulhus, D. L. (1991). Measurement and control of response bias. In P. R. J. P. Robinson, *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.
- Pedrabissi, L. &. (1989). *Inventario per l'ansia di "Stato" e di "Tratto": Nuova versione italiana dello Stai forma Y: Manuale*. Firenze: Organizzazioni speciali.
- Pedrabissi, L. &. (1997). *I test psicologici*. Bologna: Il Mulino.
- Penta, M. A. (2008). *Analisi di Rasch e questionari di misura*. Milano: Springer.
- Pletzer, J. L. (2019). A meta-analysis of the relations between personality and workplace deviance: Big Five versus HEXACO. *Journal of Vocational Behavior*, 112, 369-383.
- Poropat, A. E. (2011). A meta-analysis of the Five-Factor Model of personality and academic performance. *Psychological Bulletin*, 135, 322-338.
- R Core Team . (2021). R: A language and environment for statistical computing [Computer software]. Vienna, Austria. Retrieved from <http://www.Rproject.org/>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment test*. Danish Institute for Educational Research. Reprinted, 1980. Chicago: The University of Chicago Press.
- Reise, P. S. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, 27-48.

- Revelle, W. &. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. . *Multivariate Behavioral Research*, 14(4), 403–414. doi:https://doi.org/10.1207/s15327906mbr1404_2
- Revelle, W. (2021). *Procedures for psychological, psychometric, and personality research (Version 2.1.9)*. . Retrieved from <https://cran.r-project.org/web/packages/psych/psych.pdf>
- Roberts, T. B. (2004). Conscientiousness and health-related behaviors: A meta-analysis of the leading behavioral contributors to mortality. *Psychological Bulletin*, 130, 887-919.
- Rubini, V. (1975). *Basi teoriche del testing psicologico*. Bologna: Patron.
- Rubini, V. (1984). *Test e misurazioni psicologiche*. Bologna: Il Mulino.
- Sanavio, E. S. (1997). *I test di personalità. Inventari e questionari*. Bologna: Il Mulino.
- Smith, E. V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3(2), 205–231.
- Spielberger C.D., G. R. (1968). *State-Trait anxiety inventory form X*. Palo Alto, California: Consulting Psychologists press.
- Spielberger, C. (1976). The nature and measurement of anxiety. In D.-G. R. Spielberger CD, *Cross-cultural anxiety* (pp. 3-12). Washington DC: Hemisphere publishing corporation.
- Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory STAI*. . Consulting Psychologists Press.
- Stanley, J. C. (1972). *Educational and Psychological Measurement and Evaluation*. . Englewood Cliffs, N.J: Prentice-Hall Inc. .
- Steer, R. R. (1993). Structure of the computer-assisted Beck Anxiety Inventory with psychiatric inpatients. *J Pers Assess.*, 60:532-42.
- Strongman, K. T. (1995). Theories of Anxiety. *New Zealand Journal of Psychology Vol.24 No.2*, 4-10.
- Sullivan, H. S. (1953). *The interpersonal theory of psychiatry*. NY: Norton.
- Taherdoost, H. (2019). What is the best response scale for survey and questionnaire design; Review of different lengths of rating scale / attitude scale / Likert scale. , . *International Journal of Academic Research in Management*, 8(1), 1–10.
- Tambelli, R. (2017). *Manuale di psicopatologia dell'infanzia* . Bologna: Il Mulino .
- Thissen, D. &. (1985). A response model for multiple choice items. *Psychometrika*, 49, 501-519.
- Thurstone, L. L. (1934). The vectors of mind. *Psychological Review* , 41, 1-32.
- Thurstone, L. L. (1953). *Thurstone Temperament Schedule*. Chicago: Science Research Associates.
- Treccani. (2023, Marzo 16). *Vocabolario online*. Retrieved from [Treccani.it: https://www.treccani.it/vocabolario/ansia/](https://www.treccani.it/vocabolario/ansia/)

- Tupes, E. a. (1961). Recurrent Personality Factors based on Trait Ratings. *USAF ASD Tech. Rep. No. , Lackland Airforce Base, TX: US Air Force.*, (pp. 61-97).
- Weiss, D. Y. (1991). Item Response Theory. In R. Z. Hambleton, *Advances in Educational and Psychological Testing: Theory and Applications. Evaluation in Education and Human Services Series, vol 28*. Dordrecht: Springer.
- Williams, J. E. (2010). Anger/hostility and cardiovascular disease. In . In G. S. M. Potegal, *International handbook of anger* (pp. 435-447). New York : Springer.
- Wothke, W. (1995). Covariance components analysis of the multitrait-multimethod matrix. . In P. E. (Eds.), *Personality research, methods, and theory: A festschrift honoring Donald W. Fiske* (pp. pp. 125–144). Lawrence Erlbaum Associates, Inc.
- Wright, B. &. (1999). *Measurement essentials (2nd ed.)*. Wide Range, Inc.
- Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, vol XIV, n. 2*, 97-116.
- Wright, B. D. (1982). *Rating scale analysis: Rasch measurement.* . Mesa Press.
- Wright, B. D. (1992). Combining (collapsing) and splitting categories. . *Rasch Measurement Transactions,* , 6(3), 233–235.
- Yuan, K. H. (2000). Three likelihood-based methods for mean and covariance structure analysis with non-normal missing data. . *Sociol. Methodol.*, 30, 165-200.
- Zammuner, V. (1998). *Tecniche dell'intervista e del questionario.* . Bologna: Il Mulino .