UNIVERSITÀ DEGLI STUDI DELL'AQUILA
DIPARTIMENTO DI SCIENZE CLINICHE E BIOTECNOLOGICHE

Dottorato di Ricerca in Medicina Sperimentale

Curriculum in Neuroscienze di base e cliniche

XXXVII ciclo

Titolo della tesi

MODELLI STATISTICO-PSICOMETRICI NELLA REVISIONE E COSTRUZIONE DI STRUMENTI DI ASSESSMENT NEL DISTURBO DELLO SPETTRO DELL'AUTISMO: IMPLICAZIONI PER LA PRATICA CLINICA

SSD MEDS-24/A

Dottoranda

ILENIA LE DONNE

Coordinatrice del corso

Prof.ssa Mariagrazia Perilli

Tutor

Prof. Marco Valenti

Co-Tutor

Prof.ssa Monica Mazza

a.a. 2023/2024

Indice

Abstract 5

Introduzione 7

Parte I: Background Teorico 12

1. Il Disturbo dello Spettro dell'Autismo 13

1.1 L'evoluzione della categoria diagnostica 13

1.2 Sistemi di classificazioni 15

1.3 Dati di prevalenza 20

1.4 Eziologia e fattori di rischio 22

1.5 Teorie Cognitive e neuroscienze 24

1.6 Il Deficit di Cognizione Sociale 26

1.7 Presentazione clinica: sintomi precoci e comorbidità 30

1.8 Il processo diagnostico e valutativo 35

1.9 Interventi e prognosi 37

2. Metodi statistico-psicometrici nel setting clinico 40

2.1 Psicologia e Psicometria 40

2.2 Misurazione e Testing Psicologico 41

2.3 Psicometria nella ricerca psicologica clinica 43

2.4 Le proprietà psicometriche: Validità, Affidabilità e Dimensionalità 45

2.4.1 Item Response Theory: Modello di Rasch 46

2.4.2 Analisi fattoriale 50

2.4.3 Exploratory Graph Analysis (EGA) 54

2.5 L'Accuratezza Diagnostica 57

Parte II: Contributi Sperimentali 62

Introduzione agli studi 63

Studio 1 70

Diagnosi precoce nel disturbo dello spettro autistico: valutazione dell'accordo predittivo tra il Modulo Toddler e il Modulo 1 dell'Autism Diagnostic Observation Schedule, seconda edizione. 70

1.1 Scopo dello Studio 70

1.2 Metodo 70

Partecipanti 70

Misure 71

1.3 Analisi 72

1.4 Risultati 72

Matrice di classificazione 72

Modelli di Regressione Logistica 73

1.5 Discussione 73

1.6 Conclusioni 75

Studio 2 77

Applicazione dei metodi psicometrici nell'analisi dimensionale e nell'integrazione degli strumenti di valutazione nella diagnosi precoce del disturbo dello spettro autistico. 77

2.1 Scopo del lavoro 77

2.2 Metodo 77
Partecipanti 77
Misure 78

2.3 Analisi 78

2.4 Risultati 80

2.5 Discussione 83

2.6 Conclusioni 86

Studio 3 87

Autismo e test di attribuzione dell'intenzione: una valutazione non verbale con fumetti 87

3.1 Scopo del lavoro 87

3.2 Studio 1: Costruzione e validazione dell'IA-CST 87

3.2.1 Metodi 87
Procedura 87
Misure 88
Costruzione del IA-CST 89
Partecipanti 89

3.2.2 Analisi Statistiche 90

3.2.3 Risultati 92
Scala finale: IA-CST 95

3.3 STUDIO 2: Validità Clinica 96

3.3.1 Metodo 96
Partecipanti 96

3.3.2 Analisi 96

3.3.3 Risultati 97

3.5 Conclusioni 102

Studio 4 104

Valori normativi e ottimizzazione diagnostica di tre misure di cognizione sociale per la diagnosi di autismo e schizofrenia in un campione di adolescenti e adulti sani 104

4.1 Scopo del lavoro 104

4.2 Materiali e Metodi 104
Partecipanti 104
Procedure 106
Misure 106

4.3 Analisi statistiche 109
Modello IRT 109
Validità clinica 110

4.4 Risultati 111
IRT 111
Dati normativi dalla popolazione 114
Validità clinica 115

4.5 Discussione 119
Scale riviste nel quadro IRT 120
Validità clinica delle scale riviste 123
4.6 Conclusioni 128
Conclusioni generali 129
Appendice 132
Bibliografia 138
Lista abbreviazioni 173

Abstract

In recent years, research has been focusing on identifying reliable and valid assessment and evaluation methods to increase knowledge and symptom characterization of Autism Spectrum Disorder (ASD). Although there is well-established evidence of the need to develop valid and reliable instruments, with the introduction of measures there has also been the introduction of methodological problems and limited psychometric quality; for example, reviews of the literature report that only 7% of the studies reviewed included information on the psychometric properties of measures for ASD. Another aspect that should be taken into consideration when choosing instruments for use in clinical settings is the length and complexity of test administration.

This Ph.D thesis, therefore, aimed to highlight the importance of interdisciplinary approaches, such as the use of solid statistical and psychometric methodologies in the field of clinical psychology, particularly for ASD. This approach plays an important role when using tools for diagnosis or early identification of ASD-related symptoms. One example is the use of the Toddler Module of the Autism Diagnostic Observation Schedule (Second Edition; ADOS-2), which is used to estimate a range of risk for ASD in children aged 12 to 30 months. At present, there is no evidence about diagnostic agreement of Toddler Module with the other Modules of the ADOS-2 with which the diagnosis is, instead, placed; moreover, studies investigating its use in clinical practice are limited. The use of statistical and psychometric methodologies also plays an important role in developing and revising test instruments to investigate the core features of the condition, such as the construct of *Social Cognition* (SC), that complex set of cognitive functions and processes that underlie the ability to understand and act appropriately within social situations. The introduction of SC measures, which are valid and provide normative values, within diagnostic assessments for ASD is crucial not only in terms of accuracy of diagnosis but also for the implementation of individualized treatments that take into account the specific developmental stage of the subject and the abilities to be enhanced.

Based on these assumptions, four research studies have been conducted, in which statistical-psychometric methods are applied in the analysis, review, and development of diagnostic and SC measures. Specifically, we investigated the diagnostic agreement between two measures used in early diagnosis, namely the Toddler Module and Module 1 of ADOS-2; furthermore, through the use of psychometric networks, particularly using *Exploratory graph analysis* (EGA), we attempted to replicate the latent structure of the Toddler Module, with the aim to analyze the connection patterns between items, assessing the internal consistency of the Toddler Module, and identifying items that might not be in line with the general structure of the test.

Through the application of EGA, we have, in addition, constructed and proposed a new, easy-to-administer nonverbal test, the *Intentions Attribution -Comic Strip Task (IA-CST).* Psychometric properties and clinical validity were evaluated using the ROC curve, proving to be a valid test to be introduced in clinical settings, and in the assessment pathway for ASD, particularly for people with limited verbal abilities. With a view to use an interdisciplinary approach and to highlight the importance of integrating SC measures alongside the "gold standard" instruments in clinical practice, we reviewed three SC measures, *Advanced-Theory of Mind (A-ToM), Emotion Attribution Task (EAT)* and *Social Situation Test (SST),* according to the *Item Response Theory (IRT)* model.

The results obtained from this thesis have highlighted the importance and need for the use of statistical and psychometric methodologies in the field of clinical psychology, with the aim to develop solid psychometric tools, given the impact they can have on the individual, not only in accurate clinical framing, identifying strengths and weaknesses, but also in structuring individualized rehabilitative and habilitative interventions.

Introduzione

Negli anni sta acquisendo sempre più importanza l'utilizzo degli approcci interdisciplinari, come l'utilizzo di metodiche statistico-psicometriche nella pratica clinica psicologica e psicopatologica. Tale approccio deriva dalla necessità di sviluppare strumenti testistici e misurazioni rigorose, valide ed efficaci per condurre una ricerca di buona qualità, per studiare con precisione i costrutti psicologici e le loro relazioni, fornire un quadro accurato dello stato clinico delle persone e supportare gli operatori sanitari nella strutturazione di interventi basati su "*l'evidence based*", efficaci e mirati a promuovere il miglioramento della prognosi e della qualità della vita degli individui (Flake et al., 2022). Un esempio paradigmatico dell'utilizzo di strumenti sempre più specifici per studiarne le caratteristiche sintomatologiche e cliniche è rappresentato dal Disturbo dello Spettro dell'Autismo (*Autism spectrum disorder-ASD*). La crescente prevalenza di ASD, il suo impatto *long life* e la sua ampia eterogeneità clinica rappresentano una sfida significativa per la valutazione diagnostica, che richiede strumenti in grado di catturare sia le manifestazioni universali che le variabilità individuali all'interno dello spettro. La valutazione diagnostica, o di routine, nell'ASD, tipicamente, prevede la raccolta di informazioni da genitori o caregiver, l'osservazione del bambino, adolescente o adulto e la somministrazione di strumenti per la valutazione sintomatologica, funzionale, adattiva e cognitiva; in tale contesto l'applicazione di metodiche analitiche, con il fine di rendere più accurata e adattiva la valutazione, in termini di tempi di somministrazione e valutazione ma anche di ottimizzazione delle informazioni raccolte, ha permesso la costruzione e revisione di misure che siano sempre più applicabili nella pratica clinica.

Nella Parte I del presente elaborato viene riportata una descrizione dettagliata della condizione di ASD, partendo dall'analisi dei criteri diagnostici, passati e attuali, il processo valutativo, l'epidemiologia e le teorie cognitive presenti in letteratura, proposte per dare una spiegazione alle caratteristiche sintomatologiche della condizione. Verrà trattato il deficit di *Cognizione Sociale* (*Social Cognition- SC*), considerato il deficit *core* dell'ASD e della sua importanza nella valutazione diagnostica, spesso trascurata nei setting clinici. La *SC* è l'insieme delle abilità che permettono agli individui di comprendere, interpretare e rispondere ai comportamenti e alle emozioni altrui; essa comprende sia abilità di base, come l'elaborazione dei volti e la percezione del movimento biologico, sia abilità complesse, come la comprensione e l'interpretazione degli stati mentali propri e altrui (Teoria della Mente [ToM]), nonché la condivisione di stati fisiologici ed emotivi (empatia) (Beaudoin & Beauchamp, 2020; Kilford et al., 2016). La difficoltà dell'utilizzo delle misure di SC nella pratica clinica è data anche da una limitata disponibilità di strumenti con proprietà psicometriche adeguate e in grado di misurare le abilità

di SC (Morrison et al., 2019). Inoltre, è necessario rendere accessibili i valori normativi che siano indicativi della prestazione riportata dal soggetto (Mazza & Valenti, 2019). In tale background teorico ampio spazio verrà dedicato, appunto, all'importanza dell'utilizzo della psicometria, delle metodiche statistiche e analitiche che hanno permesso, nel tempo, la definizione del concetto di misurazione. Questo perché, la qualità dei test psicologici e delle misure di assessment ha un impatto non solo in termini di diagnosi clinica ma anche sulla valutazione di interventi sanitari, dei servizi e nella ricerca clinica. Tuttavia, l'efficacia dei test dipende dall'accuratezza, dall'affidabilità e dalla validità dei costrutti che essi misurano. Verrà ripercorsa la concettualizzazione di misurazione nella psicologia e di come la costruzione e la revisione di strumenti testistici richiedono un approccio rigoroso che integri metodologie psicometriche di base e innovative, come la *Teoria Classica dei Test* (CTT, Classical Test Theory) e la *Teoria della Risposta agli Item* (IRT, Item Response Theory). Verranno riportati i principi base dell'Analisi Fattoriale, che permette lo studio della dimensionalità e validità di uno strumento testistico fino ad arrivare a metodiche più innovative, che tentano di superare alcuni limiti delle tecniche tradizionali. Tra questi vi è l'introduzione dell'applicazione dell'Analisi dei *network* allo studio dei costrutti psicologici, come *l'Exploratory Graph Analysis* (EGA). Tali metodiche consentono di ottimizzare la misurazione delle dimensioni latenti associate all'ASD, migliorando al contempo la precisione diagnostica e l'applicabilità degli strumenti a diversi contesti culturali e clinici. Tuttavia, un altro importate aspetto da tenere in considerazione quando parliamo di metodiche statistico-psicometriche è valutare l'accuratezza diagnostica delle misurazioni applicate nella pratica clinica, intesa come la capacità dello strumento di discriminare tra condizioni cliniche e non. Verrà quindi trattata la capacità discriminativa e predittiva di un test e, di conseguenza, le proprietà di sensibilità e specificità. La relazione tra sensibilità e specificità viene rappresentata attraverso la *curva ROC* (*Receiver Operating Characteristic curve*), consentendo di analizzare le caratteristiche di uno strumento nel caso non sia possibile, o non abbia un significato clinico, esprimere i risultati in forma dicotomica.

Sulla base di tali presupposti, la Parte II è stata dedicata all'esposizione di quattro lavori di ricerca, il cui obiettivo comune è esplorare l'utilizzo delle metodiche statistico-psicometriche nella costruzione e revisione di strumenti diagnostici per l'ASD. I lavori si focalizzano su due aspetti principali, 1)l'impiego di tecniche avanzate per migliorare l'efficacia e la validità degli strumenti utilizzati nella valutazione per l'ASD e 2) le implicazioni pratiche di questi strumenti nella pratica clinica.

I primi due studi sono stati condotti con il fine di indagare l'utilizzo dell'Autism Diagnostic Observation Schedule, Seconda Edizione (ADOS-2; Lord et al., 2012), nella diagnosi e identificazione precoce di bambini con o a rischio ASD. L'ADOS-2 è considerato il "gold-standard" per la diagnosi di ASD e comprende diversi moduli adattati a diversi gruppi di età e livelli di sviluppo. Per la diagnosi precoce, i moduli ADOS-2 principalmente utilizzati sono il Modulo Toddler e il Modulo 1, specificamente progettati per valutare i bambini molto piccoli. Diversi studi hanno analizzato la sensibilità e la specificità degli algoritmi diagnostici e delle dimensioni dell'ADOS-2, in particolare per i Moduli 1, 2, 3 e 4 (de Bildt et al., 2011; Hong et al., 2021; Kamp-Becker et al., 2013; Medda et al., 2019). Per il Modulo Toddler, invece, gli studi sul suo utilizzo nella pratica clinica rimangono limitati (Hong et al., 2021; Lee et al., 2019). Inoltre, nel contesto Italiano non esistono studi che indagano l'accordo tra il Modulo Toddler e gli altri moduli. Così il primo studio, di tipo longitudinale, si è incentrato sulla valutazione dell'accordo diagnostico tra il Modulo Toddler e il Modulo 1, coinvolgendo 43 bambini, con un'età media di 25.27 mesi (3.6 SD) alla prima consultazione. Verificare l'accordo predittivo delle misure faciliterebbe la continuità diagnostica e l'identificazione precoce dei bambini a rischio di ASD. E' stato inoltre analizzato l'effetto predittivo dei due costrutti diagnostici condivisi dal Modulo Toddler e Modulo 1, ovvero l'Affetto Sociale (AS) e i Comportamenti ristretti e ripetitivi (CRR). La comprensione di come i due costrutti AS e CRR influenzano la diagnosi, consente di strutturare interventi mirati alle caratteristiche del disturbo con l'obiettivo di migliorare la prognosi.

Su questa scia, il secondo studio è stato condotto con l'obiettivo di replicare la struttura degli item del Modulo Toddler dell'ADOS-2 in una popolazione italiana di 91 bambini di età compresa tra i 12 e i 30 mesi, con l'obiettivo di identificare fattori latenti ed eventuali pattern o cluster di item correlati, mediante l'utilizzo dell'Exploratory Graph Analysis. Dall'analisi del costrutto e della dimensionalità del Modulo Toddler, sono stati analizzati i compiti strutturati presenti in quest'ultimo, i quali permettono di codificare gli item selezionati per l'algoritmo diagnostico. Sono stati confrontati gli item che si aggregano nelle dimensioni individuati dall'analisi EGA, con i compiti "principali" che ne permetterebbero la codifica. Questo approccio ci permette di differenziare gli item essenziali da quelli "ridondanti", facilitando un processo di valutazione adattivo ed efficiente per i bambini molto piccoli.

Il terzo e il quarto studio sono stati condotti con l'obiettivo di fornire evidenza dell'importanza della valutazione del costrutto di SC e del fornire strumenti validi e standardizzati da implementare nella pratica clinica.

Nello specifico, nel terzo studio viene proposto un nuovo strumento non verbale, *l'Intentions Attribution - Comic Strip Test* (IA-CST), per la valutazione della capacità di inferire le intenzioni altrui, prerequisito fondamentale della ToM. La ToM rappresenta una delle principali componenti del costrutto di SC e consiste nella capacità di comprendere e attribuire stati mentali e affettivi a sé stessi e agli altri. In particolare lo studio ha avuto un duplice obiettivo: (1) costruire e validare l'IA-CST e fornire i valori normativi di un ampio campione (N=261) di adolescenti e adulti; (2) confrontare la prestazione del gruppo ASD con i controlli TD appaiati per QI. Per valutare preliminarmente le prestazioni diagnostiche dell'IA-CST è stata eseguita una curva ROC. L'IA-CST ha dimostrato buone proprietà psicometriche, rappresentando uno strumento utile per identificare i deficit nell'attribuzione delle intenzioni.

Il quarto e ultimo studio riportato, è stato condotto partendo dalla consapevolezza dell'importanza di valutare le abilità di SC non solo nell'ASD ma anche in condizioni psicopatologiche che possono presentare importanti aree di *overlapping* con l'autismo, come ad esempio i disturbi dello spettro schizofrenico (*schizophrenic spectrum disorder - SSD*). Lo scopo dello studio è quello di fornire strumenti adatti a valutare gli aspetti della cognizione sociale che possano essere utilizzati in contesti clinici e di ricerca a supporto delle misure gold-standard. In questo caso, sono state analizzate le proprietà psicometriche, nell'ottica di revisionare tre misure di SC proposte dalla letteratura, ovvero compiti di Teoria della Mente Avanzato (*Advanced- Theory of Mind;* A-ToM), di Attribuzione dell'Emozioni (*Emotion Attribution Task;* EAT) e delle Situazioni Sociali (*Social Situations Task*; SST), utili per indagare rispettivamente le abilità cognitive e affettive della ToM e la comprensione delle norme sociali. Questo obiettivo è stato raggiunto con l'utilizzo dell'*Item Response Theory (IRT);* un vantaggio del modello IRT è che permette di stimare i parametri specifici degli item tenendo conto delle abilità degli individui e della difficoltà degli item. Con l'obiettivo di implementare tali strumenti lungo il percorso valutativo dell'ASD, il lavoro di ricerca presentato fornisce i valori normativi di un ampio campione di individui a sviluppo tipico; inoltre, è stata valutata la validità clinica delle tre misure riviste valutando le prestazioni di due popolazioni con un funzionamento atipico della cognizione sociale, ovvero ASD e SSD.

Nel complesso, i lavori presentati evidenziano l'importanza delle applicazioni delle metodiche statistico-psicometriche nel campo della salute, nella costruzione e revisione di strumenti di valutazione; inoltre, i risultati ottenuti hanno implicazioni importanti nella pratica clinica. Forniscono, infatti, un supporto all'utilizzo di strumenti di SC accanto ai test "gold standard", fornendo valori normativi e cut off, nonché importanti informazioni cliniche durante il percorso valutativo di ASD. Inoltre, l'utilizzo degli approcci interdisciplinari hanno implicazioni per

quanto riguarda la diagnosi e l'identificazione precoce di bambini con ASD, in particolare nel fornire una diagnosi affidabile e stabile fin dai 12 mesi, nell'ottimizzazione della raccolta delle informazioni, permettendo la strutturazione di interventi tempestivi e mirate sulle caratteristiche della condizione.

Parte I: Background Teorico

1. Il Disturbo dello Spettro dell'Autismo

1.1 L'evoluzione della categoria diagnostica

Il Disturbo dello Spettro dell'Autismo (*Autism Spectrum Disorder -ASD*) rappresenta un costrutto utilizzato per descrivere le persone che presentano una combinazione specifica di compromissioni comportamentali, sociali e adattive, divenendo un esempio paradigmatico nello studio di diversi processi cognitivi e nell'integrazione di approcci multidisciplinari.

Sebbene le prime descrizioni siano attribuite a Leo Kanner (1943) e Hans Asperger (1944), l'autismo ha attirato l'attenzione di ricercatori e clinici da circa 500 anni prima (Lord et al., 2020; Wing, 1997). La descrizione delle caratteristiche principali dell'autismo, infatti, nascono da miti e alimentato storie riguardanti i discepoli di San Francesco D'Assisi, in particolare, del fratello Juniper (Frith, 1989) o a Victor, il bambino che viveva nella foresta dell'Aveyron e cresciuto dai lupi (Lord et al., 2020; Wing, 1997). La concettualizzazione dell'autismo nella letteratura psichiatrica risale al 1867 in cui Henry Maudsley suggerì che i bambini con comportamenti strani potevano essere classificati come "*affetti da psicosi infantile*"; nella prima metà del XX secolo, altri iniziarono a descrivere bambini all'interno della psicosi infantile, come Sante De Santcis (1909) e Theodor Heller (1908) che descrivevano bambini con uno sviluppo apparentemente normotipico per un periodo per poi perdere il linguaggio e le abilità sociali (Wing, 1997).

Eugen Bleuler, psichiatra svizzero, nel 1911, introduce per la prima volta in ambito clinico la parola autismo (dal greco "*autòs",* sé stesso). Bleuer utilizzò il termine "autismo" per indicare un atteggiamento tipico della schizofrenia, contraddistinta da un "pensiero autistico" (o *dereistico*), ovvero una condizione in cui l'individuo appare distaccato dal mondo esterno e "ritirato in sé stesso" (Bleuler, 1911). È con lo psichiatra Leo Kanner (1943) che il termine autismo viene utilizzata come "etichetta diagnostica" per indicare una condizione clinica ad esordio precoce e che presenta principalmente da difficoltà nella comunicazione e nelle relazioni sociali. Kanner, infatti, definiva "*disturbo autistico da contatto affettivo"* un complesso di sintomi presentati da un gruppo di 11 bambini, tra i due e dieci anni; i bambini descritti da Kanner presentavano una significativa chiusura relazionale, linguaggio ripetitivo e non funzionale, resistenza a cambiamenti, nonché comportamenti stereotipati e interessi ristretti (Evans, 2013; Rosen et al., 2021). L'anno successivo, Asperger (1944) descrisse un gruppo di bambini e adolescenti, con caratteristiche simili a quelli individuati da Kanner, ovvero un'inappropriata interazione sociale, comportamenti schematici e ripetitivi, e assorti nei loro interessi circoscritti; tuttavia, i bambini descritti da Asperger presentavano buone abilità di

linguaggio e un livello intellettivo da normale a superiore. Fu in tale contesto che Asperger introdusse il termine "*psicopatia autistica".*

Un primo tentativo di definire un range di disturbi, compreso "l'autismo" di Kanner, in maniera più rigorosa, avvenne nel 1961, anno in cui Mildred Creak propose un set di criteri (i Nove punti) che definivano "*schizofrenia infantile*". La sovrapposizione tra autismo infantile, psicosi e schizofrenia infantile, caratterizzò la letteratura fino i primi anni '70 (Wing, 1997). La distinzione tra autismo e schizofrenia infantile non viene fatta nelle prime edizioni dei sistemi di classificazione diagnostica americani ed europei, come il Manuale Diagnostico e Statistico dei Disturbi Mentali (DSM) a cura dell'American Psychiatric Association (APA) e la Classificazione Statistica Internazionale delle Malattie e dei Problemi Sanitari Correlati (ICD; World Health Organization -WHO),

Un contributo sostanziale allo studio sull'autismo viene dai lavori scientifici effettuati da Israel Kolvin (1971) e Michael Rutter (1972). Nei loro lavori vengono descritti nel dettaglio le caratteristiche cliniche della condizione autistica, ovvero isolamento sociale e la resistenza a cambiamenti delle routine.

L'autismo è stato riconosciuto come categoria diagnostica, nel 1978, nella terza edizione del DSM (DSM-III; APA, 1980) e inserito tra i Disturbi Pervasivi dello Sviluppo (PDD), insieme all'*autismo atipico*, concettualizzato come "danneggiamento dello sviluppo sociale e comunicativo, con età di esordio prima dei 30 mesi di età". Nella revisione del DSM-III-R, (APA, 1987), l'autismo infantile viene sostituito da *disturbo autistico*; vi è, inoltre, un ampliamento dei criteri diagnostici e viene riconosciuta la natura pervasiva del disturbo e la necessità di considerare la diagnosi in una prospettiva evolutiva (Volkmar et al., 1988); la diagnosi non è più limitata alla prima infanzia, ma estesa anche dopo i 36 mesi, permettendo l'identificazione dei casi più lievi. Inoltre, nel DSM-III-R (APA, 1987) riporta una triade sintomatologica: compromissione dell'interazione sociale reciproca, della comunicazione e la presenza dei comportamenti e interessi ristretti, stereotipati e ripetitivi.

Nella successiva revisione del DSM-IV (APA, 1994) l'autismo è inserito tra i "*Disturbi pervasivi dello sviluppo",* considerando le due condizioni analoghe ed eterogenee per modalità di esordio e decorso con caratteristiche che riguardano la *triade* di sintomi. Nel DSM- IV e nella decima revisione dell'ICD (WHO, 1992), a fianco al Disturbo Autistico, vengono inserite altre categorie diagnostiche, come il disturbo di Asperger, il disturbo di Rett e il disturbo disintegrativo dell'infanzia che verranno mantenute anche nella revisione del DSM-IV-TR (APA, 2000). La *Sindrome di Asperger* è stata descritta per la prima volta da Lorna Wing

(1981), per descrivere persone che presentavano buone abilità di linguaggio ma che nella comunicazione e interazione sociale erano compromessi.

Wing, dunque, si rifà alla condizione clinica osservata anni prima da Hans Asperger, proponendo che la Sindrome di Asperger faccia parte di un continuum, o spettro, dell'autismo. Secondo Wing (1981, 1992), l'autismo di Kanner e la Sindrome di Asperger rappresentano manifestazioni dello stesso disturbo, differenziandosi per gravità (Wing, 1992). Wing, dunque, introduce anche il concetto di *spettro*, ritenendo che le diverse manifestazioni dell'autismo siano accumunate dalla presenza di compromissioni di tre aspetti principali: interazione sociale, comunicazione e immaginazione.

In linea con questa concettualizzazione, Doris Allen (1988) suggerisce che i bambini con caratteristiche autistiche costituiscano una popolazione eterogenea. Secondo Allen, le differenze cliniche derivano dalle manifestazioni dei deficit in tre ambiti principali: (1) sociale, affettivo e comportamentale; (2) linguaggio e comunicazione interpersonale; (3) attività, gioco e interessi ripetitivi e stereotipati. Inoltre, Allen conia il termine "*disturbo dello spettro autistico*", termine che troveremo attuali classificazioni diagnostiche di rifermento.

1.2 Sistemi di classificazioni

Dalle prime descrizioni fornite da Leo Kanner e Hans Asperger (1943 e 1944) il termine Autismo ha affrontato diverse definizioni operazionali, giungendo all'attuale classificazione presente nel DSM 5 (APA, 2013) e DSM 5-TR (APA, 2022) come *Disturbo dello Spettro dell'Autismo (Autism Spectrum Disorder - ASD). L*'autismo è classificato tra i "*Disturbi del neurosviluppo*", un insieme di condizioni caratterizzate da uno sviluppo atipico ed eterogeneo che si manifesta nelle prime fasi della vita. Queste condizioni, fortemente influenzate da un'eziologia genetica, modulata dall'interazione tra geni e ambiente, si distinguono per deficit che compromettono il funzionamento sociale, personale, scolastico o lavorativo (Tick et al., 2016; Mandy & Lai, 2016). Tra i disturbi inclusi in questa classificazione si annoverano il disturbo dello sviluppo intellettivo (o disabilità intellettiva), il disturbo della comunicazione, il disturbo da deficit di attenzione/iperattività, il disturbo specifico dell'apprendimento e i disturbi del movimento. Con l'introduzione del DSM-5, sono state apportate significative innovazioni nei criteri di classificazione diagnostica. Una delle principali novità riguarda il passaggio da un sistema multi-categoriale a una diagnosi unitaria dell'autismo, basata su dimensioni multiple (Lord et al., 2020). Le precedenti sottocategorie diagnostiche del DSM-IV-TR (APA, 2000) – disturbo autistico, disturbo di Asperger, disturbo pervasivo dello sviluppo non altrimenti specificato (PDD-NOS), disturbo di Rett e disturbo disintegrativo dell'infanzia – sono state

riunite, nel DSM-5 (APA, 2013, 2022), sotto il termine unitario di Disturbo dello Spettro Autistico (ASD), riflettendo la natura dello "spettro" della condizione (Lai et al., 2014; Rosen et al., 2021). Il concetto di spettro evidenzia la multidimensionalità del disturbo, le cui manifestazioni tengono conto della gravità della sintomatologia, del livello di sviluppo e dell'età cronologica dell'individuo (APA, 2013)

Un altro importante cambiamento che ha portato il DSM 5 (APA, 2013) è la scomparsa della triade sintomatologica e la proposta di un modello sintomatologico a due dimensioni, in cui vi è il raggruppamento in un unico dominio delle categorie "comunicazione" e "interazione sociale"; il dominio dei comportamenti e interessi ripetitivi viene mantenuto come separato (Rosen et al., 2021). Le peculiarità sensoriali, inoltre, vengono incluse nel dominio dei comportamenti e interessi ristretti e ripetitivi, sottolineandone la pervasività (Lord et al., 2020; Rosen et al., 2021).

Dunque, i criteri del DSM 5, e della recente revisione DSM 5-TR (APA, 2022) prevedono i seguenti criteri:

A. Deficit persistenti della comunicazione sociale e dell'interazione sociale in molteplici contesti, come manifestato dai seguenti fattori, presenti attualmente o nel passato:

1. Deficit della reciprocità socio-emotiva, che vanno, per esempio, da un approccio sociale anomalo e dal fallimento della normale reciprocità della conversazione; a una ridotta condivisione di interessi, emozioni o sentimenti; all'incapacità di dare inizio o di rispondere a interazione sociali;

2. Deficit dei comportamenti comunicativi non verbali utilizzati per l'interazione sociale, che vanno, per esempio, dalla comunicazione verbale e non verbale scarsamente integrata; ad anomalie del contatto visivo e del linguaggio del corpo o deficit della comprensione e dell'uso dei gesti; a una totale mancanza di espressività facciale e di comunicazione non verbale;

3. deficit dello sviluppo, della gestione e della comprensione delle relazioni, che vanno, per esempio, dalle difficoltà di adattare il comportamento per adeguarsi ai diversi contesti sociali; alle difficoltà di condividere il gioco di immaginazione o di fare amicizia; all'assenza di interesse verso i coetanei.

B. Pattern di comportamento, interessi o attività ristretti, ripetitivi, come manifestato da almeno due dei seguenti fattori, presenti attualmente o nel passato:

1. Movimenti, uso degli oggetti o eloquio stereotipato o ripetitivi, (ad esempio, stereotipie motorie semplici, mettere in fila giocattoli o capovolgere oggetti, ecolalia, frasi idiosincratiche);

2. Insistenza nella *sameness* (immodificabilità), aderenza alla routine priva di flessibilità o rituali di comportamento verbale o non verbale (ad esempio, estremo disagio davanti a piccoli cambiamenti, difficoltà nelle fasi di transizione, schemi di pensiero rigidi, saluti rituali, necessità di percorrere la stessa strada o mangiare lo stesso cibo ogni giorno);

3. Interessi molto limitati, fissi che sono anomali per intensità o profondità, (ad esempio, forte attaccamento o preoccupazioni nei confronti di oggetti insoliti, interessi eccessivamente circoscritti o perseverativi);

4.Iper- o iporeattività in risposta a stimoli sensoriali o interessi insoliti verso aspetti sensoriali dell'ambiente (ad esempio, apparente indifferenza a dolore/temperatura, reazione di avversione nei confronti di suoni o consistenze tattili specifici, annusare o toccare oggetti in modo eccessivo, essere affascinati da luci o da movimenti).

C. I sintomi devono essere presenti nel periodo precoce dello sviluppo (ma possono non manifestarsi pienamente prima che le esigenze sociali accedano le capacità limitate, o possono essere mascherati da strategie apprese in età successiva).

D. I sintomi causano compromissione clinicamente significative del funzionamento in ambito sociale, lavorativo o in altre aree importanti.

E. Queste alterazioni non sono meglio spiegate da Disabilità Intellettiva (disturbo dello sviluppo intellettivo) o da ritardo globale dello sviluppo. La disabilità intellettiva e il disturbo dello spettro dell'autismo spesso sono presenti in concomitanza; per porre diagnosi di comorbilità di disturbo dello spettro dell'autismo e di disabilità Intellettiva, il livello di comunicazione sociale deve essere inferiore rispetto a quanto atteso per il livello di sviluppo generale.

Il DSM 5 e DSM 5-TR richiedono, nel porre la diagnosi, l'utilizzo di specificatori, quali:

- Con o senza compromissione intellettiva associata
- Con o senza compromissione del linguaggio associata
- Associato ad una condizione medica o genetica nota o ad un fattore ambientale
- Associato ad un altro disturbo del neurosviluppo, mentale o comportamentale
- Con catatonia

È previsto, inoltre, l'uso di specificatori, strumenti utili per identificare le caratteristiche individuali e valutare il livello di gravità della condizione, che può variare a seconda del contesto e del tempo. Sono stati introdotto tre livelli di gravità, definiti in base alle compromissioni nelle abilità di comunicazione sociale (Criterio A) e nei comportamenti ristretti e ripetitivi (Criterio B). Questo permette di stabilire l'intensività del supporto di cui ha bisogno una persona (Tabella 1). Il livello di gravità "lieve", Livello 1, prevede che la persona con ASD

presenti alterazioni qualitative nelle abilità di comunicazione sociale, accompagnate da rituali occasionali, comportamenti ripetitivi e interessi stereotipati e rigidi. Il Livello 2 corrisponde a una gravità "moderata", con alcune abilità socio-comunicative presenti, ma limitate al punto da ostacolare significativamente l'interazione sociale; sono inoltre evidenti interessi ristretti e comportamenti stereotipati. Infine, il livello di gravità più elevato, Livello 3, richiede un "supporto molto significativo" ed è caratterizzato da gravi deficit nella comunicazione verbale e non verbale, nelle interazioni sociali, e da rigidità comportamentale che compromette gravemente il funzionamento quotidiano.

Tabella 1. Livelli di gravità secondo il DSM 5 (APA, 2013) e DSM 5-TR (APA, 2022)

Livello di Gravità	**Comunicazione Sociale**	**Comportamenti Ristretti e Ripetitivi**
Livello 3 "è necessario un supporto molto significativo"	Gravi deficit delle abilità di comunicazione sociale verbale e non verbale causano gravi compromissioni del funzionamento, avvio molto limitato delle interazioni sociali e reazioni minime alle aperture sociali da parte degli altri. Per esempio, una persona con un eloquio caratterizzato da poche parole comprensibili, che raramente avvia interazione sociali e, quando lo fa, mette in atto approcci insoliti per soddisfare esigenze e risponde solo ad approcci sociali molto diretti.	Inflessibilità di comportamento, estrema difficoltà nell'affrontare il cambiamento, o altri comportamenti ristretti/ripetitivi interferiscono in modo marcato con tutte le aree del funzionamento. Grande disagio/difficoltà nel modificare l'oggetto dell'attenzione o l'azione.
Livello 2 "è necessario un supporto significativo"	Deficit marcati delle abilità di comunicazione sociale verbale e non verbale; compromissioni sociali visibili anche in presenza di supporto; avvio limitato delle interazioni sociali; reazioni ridotte o anomale alle aperture sociali da parte di altri. Per esempio, una persona che parla usando frasi semplici, la cui interazione è limitata a interessi ristretti e particolari e che presenta una comunicazione non verbale decisamente strana.	Inflessibilità di comportamento, difficoltà nell'affrontare i cambiamenti o altri comportamenti ristretti/ripetitivi sono sufficientemente frequenti da essere evidenti a un osservatore casuale e interferiscono con il funzionamento in diversi contesti. Disagio/difficoltà nel modificare l'oggetto dell'attenzione o l'azione.
Livello 1 "è necessario un supporto"	In assenza di supporto, i deficit della comunicazione sociale causano notevoli	L'inflessibilità di comportamento causa interferenze significative con il

	compromissioni. Difficoltà ad avviare le interazioni sociali, e chiari esempi di risposte atipiche o infruttuose alle aperture sociali da parte di altri. L'individuo può mostrare un interesse ridotto per le interazioni sociali. Per esempio, una persona che è in grado di formulare frasi complete e si impegna nella conversazione bidirezionale con gli altri, e i cui tentativi di fare amicizia sono strani e in genere senza successo.	funzionamento in uno o più contesti. Difficoltà nel passare da un'attività ad un'altra. I problemi nell'organizzazione e nella pianificazione ostacolano l'indipendenza.

A differenza delle precedenti classificazioni, in cui riconducevano la manifestazione sintomatologica nei primi 3 anni di vita, il criterio C del DSM 5, riconosce i primi sintomi durante il periodo precoce dello sviluppo ma che la sintomatologia potrebbe non manifestarsi pienamente fino a quando le richieste sociali non eccedono le capacità dell'individuo. Con tale cambiamento viene riconosciuta la natura evolutiva dell'autismo, permettendo l'identificazione della sintomatologia anche in età adolescenziale e adulta (Lord et al., 2020).

Un altro sistema di classificazione rilevante a livello internazionale è l'ICD (International Classification of Diseases), riconosciuto come il principale sistema medico-diagnostico globale. A differenza di altri approcci, l'ICD ha come obiettivo principale la categorizzazione delle condizioni cliniche piuttosto che la diagnosi specifica (Doernberg & Hollander, 2016; Rutter, 2011). Nella decima edizione (ICD-10; WHO, 1992), l'autismo è incluso nel Capitolo V, dedicato alle *Sindrome e disturbi psichici e comportamentali* (F80-89). In questa sezione vengono classificati disturbi con insorgenza dei sintomi nell'infanzia, caratterizzati da compromissioni o ritardi nello sviluppo di funzioni correlate alla maturazione biologica del sistema nervoso centrale e da un decorso continuo. L'autismo è specificato nella sottocategoria *Sindromi da alterazione globale dello sviluppo psicologico;* sindromi che condividono caratteristiche comuni, tra cui anomalie qualitative nelle interazioni sociali e nella comunicazione, unite a un repertorio di interessi e attività limitato, stereotipato e ripetitivo. Tra le sottocategorie sono comprese le seguenti diagnosi:

- Autismo infantile (F84.0)
- Autismo atipico (F84.1)
- Sindrome di Rett (F84.2)
- Disturbo disintegrativo dell'infanzia di altro tipo (F84.3)
- Sindrome iperattiva associata a ritardo mentale e movimenti stereotipati (F84.4)

- Sindrome di Asperger (F84.5)
- Altre sindromi da alterazione globale dello sviluppo psicologico
- Sindrome non specificata da alterazione globale dello sviluppo psicologico (F84.9).

Con l'introduzione dell'ICD-11 (WHO, 2018), entrato in vigore il 1° gennaio 2022, sono state introdotte importanti modifiche nella classificazione dell'autismo, seguendo una linea simile a quella del passaggio dal DSM-IV al DSM-5. L'autismo è ora inserito nella categoria *Disturbi mentali, comportamentali o dello sviluppo neurologico* (06), con la specifica denominazione di *Disturbo dello spettro autistico* (6A02) all'interno dei *Disturbi dello sviluppo neurologico*. L'ICD 11 (WHO, 2018) propone sette forme di autismo, differenziate in base alla presenza di compromissione intellettiva e alterazioni linguistiche:

- Disturbo dello spettro autistico senza disturbo dello sviluppo intellettivo e con lieve o nessuna alterazione del linguaggio funzionale (6A02.0);
- Disturbo dello spettro autistico con disturbo dello sviluppo intellettivo e con compromissione lieve o assente del linguaggio funzionale (6A02.1);
- Disturbo dello spettro autistico senza disturbo dello sviluppo intellettivo e con linguaggio funzionale alterato (6A02.2);
- Disturbo dello spettro autistico con disturbo dello sviluppo intellettivo e con alterazione del linguaggio funzionale (6A02.3);
- Disturbo dello spettro autistico con disturbo dello sviluppo intellettivo e con assenza di linguaggio funzionale (6A02.5).

Analogamente al DSM-5 (APA, 2013), l'ICD-11 suddivide la sintomatologia dell'autismo in due principali domini:

- Deficit persistenti nella capacità di avviare e sostenere l'interazione sociale reciproca e la comunicazione sociale.
- Modelli di comportamento, interessi o attività limitati, ripetitivi e inflessibili.

Tuttavia, rispetto al DSM-5, l'ICD-11 non include tra i criteri diagnostici le caratteristiche di iper- o iporeattività sensoriale o gli interessi sensoriali insoliti (Fuentes et al., 2020; Rosen et al., 2021).

1.3 Dati di prevalenza

Negli ultimi anni si è osservato un incremento generalizzato della prevalenza di ASD. Tale fenomeno può essere attribuito a diversi fattori, come una maggior conoscenza della condizione nella popolazione generale e dei clinici e l'aumento dei servizi diagnostici, come anche il

cambiamento dei criteri diagnostici che hanno permesso l'identificazione di persone ASD (Fombonne, 2015; Scattoni et al., 2023). Gli studi epidemiologici internazionali, attualmente, riportano una prevalenza dell'1,5% nei paesi sviluppati, con un incremento dei casi di autismo senza disabilità intellettiva (Valenti et al., 2019). Tuttavia, le stime variano in base alla metodologia utilizzata, come dimostrato dalle diverse prevalenze rilevate in Europa: dallo 0,63% in Danimarca e Svezia all'1,16% nel Regno Unito (Scattoni et al., 2023). In Italia, recenti studi hanno evidenziato una variabilità nella prevalenza. Ad esempio, a Catania la prevalenza è stimata allo 0,05% (Ferrante et al., 2015), mentre in Emilia-Romagna e Piemonte i dati, ottenuti da sistemi informativi regionali, indicano rispettivamente lo 0,44% e lo 0,48% tra i bambini di 6-10 anni (ELEA, 2016; NPI.net, 2016). Nel Centro Italia, uno studio condotto a Pisa ha riportato una prevalenza dell'1,15% nei bambini di 7-9 anni (Narzisi et al., 2018); in Abruzzo la prevalenza cumulativa è stata stimata dello 0,95% (352 casi su 36.938 abitanti), con un picco dell'1,19% per i nati tra il 2001 e il 2003 (Valenti et al., 2019) considerando la popolazione tra i 0 e 18 anni. Questi dati sono in linea con le stime europee, che indicano una prevalenza dell'autismo intorno all'1% (Valenti et al., 2019). Nello studio di Valenti e colleghi (2019), è stato utilizzato il "Registro Epidemiologico dei Deficit Psicologici, Neurologici e Sensoriali" (Madonna et al., 1998; Valenti et al., 2019) e viene mostrata una stabilità temporale della prevalenza cumulativa e una riduzione dell'età alla prima diagnosi. Di recente è stato condotto uno studio su scala nazionale, promosso dal Ministero della Salute e coordinato dall'Osservatorio Nazionale per l'Autismo dell'Istituto Superiore di Sanità, con l'obiettivo di stimare la prevalenza di ASD nei bambini tra i 7 e 9 anni, raccogliendo dati dal 24 febbraio 2016 al 23 febbraio 2018. Lo studio (Scattoni et al., 2023) ha utilizzato una combinazione di dati di registro e di screening della popolazione generale, rilevando una prevalenza di ASD di circa 1 bambino su 77 tra i bambini di 7-9 anni che hanno frequentato le scuole delle regioni coinvolte nello studio (Lombardia, Lazio, Sicilia), e un rapporto maschio/femmina di 4,4:1. Questi risultati sono in linea con altri studi condotti negli Stati Uniti (4,5:1), e Paesi Europei (Francia sudorientale 4,0:1; Islanda 4,4:1; Danimarca 3,9:1) (Baio et al., 2018; Delobel-Ayoub et al., 2020).

Negli Stati Uniti la prevalenza è cresciuta significativamente negli ultimi 20 anni, passando da 0,67% nel 2000 (1 su 150), a 2,3% nel 2018 (uno su 44) a 2,8% bambini di 8 anni (uno su 36) nel 2020 (Maenner et al., 2023). Per quanto riguarda l'età adulta esistono pochi studi epidemiologici. Di recente, uno studio condotto dal Center for Disease Control and Prevention (CDC) riporta una prevalenza stimata del 2,21% tra adulti dai 18 anni in su negli Stati Uniti e come per le stime dell'ASD nei bambini, la prevalenza è più alta negli uomini rispetto alle

donne (3,62% negli uomini e 0,86% nelle donne) (Dietz et al., 2020). Sebbene il tasso di ASD sembri in aumento, questa tendenza potrebbe essere influenzata dai dati provenienti da registri amministrativi, senza chiari riscontri di variazioni nella comunità tra il 1990 e il 2010 (Lord et al., 2020). In generale, la prevalenza di ASD è più alta nei maschi rispetto alle femmine con un rapporto stimato di 4:1 (Lord et al., 2020; Scattoni et al., 2023; Valenti et al., 2019). Tuttavia, i risultati non sono omogenei: alcuni studi riportano un rapporto maschi-femmine ancora più alto (8-9:1) negli individui con capacità intellettive nella media o sopra la media (Brugha et al., 2011; Fuentes et al., 2020), mentre negli studi basati su accertamento dei casi si stima viene riportato un rapporto più basso (3,7% nei bambini e 1,5% nelle bambine) (Zeidan et al., 2022; Hirota et al., 2023). Questo potrebbe essere ricondotto anche ad una sottodiagnosi nelle femmine, che spesso ricevono una diagnosi più tardiva o non vengono riconosciute affatto, in particolare nelle forme ad alto funzionamento (Begeer et al., 2013; Lai et al., 2014). Le donne, infatti, tendono a "mascherare" meglio i deficit di interazione e comunicazione sociale (Dean et al., 2017; Fuentes et al., 2020; Lai et al., 2017; Ratto et al., 2018).

1.4 Eziologia e fattori di rischio

Le ipotesi sull'eziologia dell'autismo sono molteplici e ancora non definite. Dalle prime teorizzazioni della metà del ventesimo secolo, di stampo psicoanalitico, in cui l'autismo veniva considerato come conseguenza della freddezza emozionale materna (*"madre frigorifero"*) e del "*desiderio dei genitori che il bambino non esista"*(Bettelheim, 1967; Greydanus & Toledo-Pereyra, 2012; Gyawali & Patra, 2019), ciò che ha ricevuto il sostegno scientifico sono, invece, le ipotesi neurobiologiche concomitanti e la proposta di Kanner di una "*condizione innata"* e "*incapacità di formare contatti affettivi con le persone"* (Lai et al., 2014).

Attualmente, l'autismo è considerato una condizione con un'origine multifattoriale, risultante dalla complessa interazione tra fattori genetici e ambientali (Corrales & Herbert, 2011; Lai et al., 2014). Gli studi provenienti da dati sulla popolazione, hanno stimato un'influenza dell'81% dei fattori genetici associati al rischio di ASD, mentre i fattori ambientali influenzerebbero dal 14% al 22% (Hirota et al., 2023). Sono stati principalmente gli studi sui gemelli a supportare l'ipotesi dell'ereditabilità dell'ASD, dimostrando che i gemelli monozigoti hanno una concordanza di circa il 64- 91% (Bailey et al., 1995; Lai et al., 2014; Takumi et al., 2020; Tick et al., 2016), rispetto a una concordanza dello 0-51% nei dizigoti (Hallmayer et al., 2011; Ronald & Hoekstra, 2011; Rosenberg et al., 2009; Taniai et al., 2008). Anche i fratelli di bambini con ASD presentano un rischio di ricevere una diagnosi che varia dal 2 al 20% (Ozonoff et al., 2011; Sandin et al., 2014). Esistono, tuttavia, una varietà di fattori genetici e

ambientali associati con ASD ma nessuno è specifico nello sviluppo della sintomatologia (Hirota et al., 2023). A oggi, sono stati individuati più di 100 geni e regioni genomiche correlati all'autismo (Lord et al., 2020), tra cui regioni cromosomiche come 2q37, 7q31 e 22q11. Anomalie frequenti includono delezioni e duplicazioni del cromosoma 16p11.2, nonché duplicazioni di origine materna nella regione 15q11-q13 (Duyzend et al., 2016; Lord et al., 2020; Takumi et al., 2020). L'influenza genetica è ulteriormente corroborata dall'associazione dell'ASD con patologie genetiche specifiche in circa il 10% dei casi. Tra queste, si annoverano la sindrome dell'X fragile, la sclerosi tuberosa, la sindrome di Rett e la neurofibromatosi (Fuentes et al., 2020; Lai et al., 2014).

Sono state evidenziate, inoltre, delle caratteristiche neurobiologiche che possono essere associate con l'ASD. L'utilizzo di metodiche di neuroimaging, elettrofisiologiche e genetica molecolare, hanno fatto nascere l'idea che l'autismo sia caratterizzato da una connettività neurale atipica, piuttosto che da un insieme di regioni cerebrali atipiche (Lai et al., 2014). Esistono studi che hanno evidenziato una diminuzione della connettività fronto-occipitali e da un iper-connettività parietale-occipitale (Lai et al., 2014), nonché un aumento volume cerebrale riscontrabile precocemente, tra i 6 e i 24 mesi di età (Lai et al., 2014). Sembrerebbe, inoltre, che ci siano differenze neuroanatomiche nel corso dello sviluppo sia nella materia grigia (Ippocampo, amigdala e precuneo) che nella materia bianca (fascicolo arcuato), come anche una riduzione del corpo calloso (Frazier & Hardan, 2009; Valenti et al., 2020).

In letteratura vengono, inoltre, riportate alterazioni a carico dei sistemi della serotonina e dell'acido γ-amminobutirrico (GABA), e nello specifico in termini di un iperserotoninemia e una compromissione dei circuiti GABAergici (Lai et al., 2014; Takumi et al., 2020). Sono stati presi in considerazione anche il ruolo giocato dai sistemi dell'ossitocina e della vasopressina, data la loro relazione con i comportamenti sociali, e il ruolo degli androgeni ed estrogeni come fattori di rischio e di protezione nell'insorgenza dell'autismo (Lai et al., 2014).

Per quanto riguarda i fattori ambientali, la letteratura riporta l'età avanzata dei genitori, nello specifico è stato riportato un aumento del rischio di ASD con un'età materna dai 35 anni in su e un aumento del 21% per ogni aumento di 10 anni dell'età paterna (Hirota et al., 2014; Lord et al., 2020). Vengono descritti anche lo stile di vita materno, tra cui l'obesità prima e durante la gravidanza, ipertensione gestazionale e l'utilizzo di farmaci (tra cui l'utilizzo dell'acido valproato in gravidanza) (Christensen et al., 2013; Hirota et al., 2023; Lord et al., 2020). Inoltre, ad aumentare il rischio di ASD sono stati studiati fattori quali basso peso alla nascita, parto pretermine e sanguinamento uterino, come anche l'intervallo tra una gravidanza e l'altra (Hirota et al., 2023; Lord et al., 2020; Newschaffer et al., 2007). Inoltre, l'ipotesi di una possibile

relazione tra ASD e vaccini a base di Thimerosal e anti-morbillo- parotite-rosolia o vaccinazioni ripetute è stata smentita da un gran numero di studi (DeStefano et al., 2013; Hirota et al., 2023; Sanchack & Thomas, 2016).

1.5 Teorie Cognitive e neuroscienze

A differenza delle prime teorizzazioni di stampo psicoanalitico formulate per fornire una spiegazione "all'enigmatica condizione", con l'avvento dell'era cognitiva, la ricerca sull'autismo si è orientata a fornire spiegazioni basate su evidenze e sperimentazioni (Rajendran & Mitchell, 2007). A fornire le prime evidenze furono i lavori di Utha Frith (1970, 1971), Margot R. Prior (1979), Beat Hermelin e Neil O'Conner (1970), che studiarono le funzioni cognitive quali la percezione, la memoria e il linguaggio. Da tali studi, seppur con risultati contrastanti, si iniziò a ipotizzare che le persone con autismo presentassero atipie nelle abilità percettive, di ipo- e ipersensibilità per stimoli visivi e uditivi (Prior, Gajzago, & Knox, 1976), probabilmente derivanti da una compromissione primaria nell'attenzione selettiva (Frith, 1970, 1971), come anche disturbi del linguaggio, suddividendo i bambini con autismo che acquisivano spontaneamente il linguaggio, quelli che lo acquisivano in ritardo e quelli che non lo acquisivano affatto (Ricks & WIng, 1976). Ad oggi le prospettive cognitive sull'autismo possono essere raggruppate in base a domini di interesse ma, per natura, connesse tra di loro (Lai et al., 2014) (Tabella 2).

Tabella 2. Domini cognitivi nella ricerca sull'autismo riportati da Lai et al., 2014.

	Caratteristiche comportamentali	**Costrutti cognitivi (psicologici)**
Cognizione Sociale e percezione sociale	Interazione sociale e comunicazione sociale atipiche	Sguardo e contatto visivo; percezione delle emozioni; elaborazione dei volti; percezione del movimento biologico; attenzione e orientamento sociale; motivazione sociale; elaborazione della ricompensa sociale; comunicazione non verbale; imitazione; empatia e simpatia affettiva; attenzione congiunta; gioco di finzione; teoria della mente o assunzione di prospettiva mentale; cognizione autoreferenziale; alessitimia (difficoltà a comprendere e descrivere le proprie emozioni); consapevolezza metacognitiva.
Funzioni Esecutive	Comportamento ripetitivo e stereotipato; interazione sociale e comunicazione sociale atipiche.	Flessibilità cognitiva; pianificazione; controllo inibitorio; spostamento dell'attenzione; monitoraggio; memoria di lavoro

Elaborazione dell'informazione bottom-up e top-down (locale e globale)	Elaborazione sensoriale-percettiva idiosincratica; eccellente attenzione ai dettagli; interessi limitati e comportamenti ripetitivi; interazione sociale e comunicazione sociale atipiche.	Funzionamento percettivo globale vs locale (elaborazione sensoriale-percettiva di basso livello superiore); coerenza centrale (preferenza globale vs locale); sistematizzazione (spinta a costruire sistemi basati su regole, capacità di comprendere sistemi basati su regole, conoscenza di sistemi fattuali)

L'ipotesi di un deficit delle Funzioni Esecutive (*Executive Dysfunction*) (Ozonoff et al., 1994; Pennington et al., 1997) deriva dalla somiglianza nella sintomatologia autistica con quella associata a specifiche lesioni cerebrali. Alcuni degli aspetti che le persone con ASD presentano sono il bisogno di specifiche routine, la difficoltà di spostare l'attenzione, nel controllo degli impulsi e tendenza a perseverare, sintomi simili a quelli mostrati da soggetti con una sindrome disesecutiva (Dysfunction executive symptoms – DES; Baddeley & Wilson, 1988). Le funzioni esecutive rappresentano un termine ombrello per racchiudere i più alti processi cognitivi che permettono all'individuo di controllare comportamenti e azioni per raggiungere gli obiettivi, come ad esempio inibire risposte inappropriate, pianificare, monitorare la perfomance e spostare o mantenere l'attenzione (Ozonoff et al., 1991; Denkla, 1996). Una compromissione delle funzioni esecutive potrebbe spiegare la presenza dei comportamenti stereotipati e interessi ripetitivi, ma anche i deficit nella comunicazione sociale nell'autismo. Gli studi di imaging hanno evidenziato il coinvolgimento di sistemi sottostanti al deficit esecutivo nelle persone con ASD, come la corteccia frontale e prefrontale (PFC), parietali e striatali (Dichter, 2012; Lai et al., 2014; Philip et al., 2012). Sebbene l'ipotesi di un deficit delle funzioni esecutive spieghi alcune delle caratteristiche dell'autismo, presenta alcuni limiti. In primo luogo, non tutti gli individui con autismo presentano difficoltà esecutive o le possono presentare in diverse funzioni. In letteratura sono presenti pochi studi che hanno tentato di determinare la prevalenza dei deficit esecutivi nell'autismo, ma con risultati contrastanti; le difficoltà nei compiti esecutivi sono stati riscontrati dal 50% al 96% del gruppo ASD rispetto al gruppo di controllo (Pellicano et al., 2006; Ozonoff et al., 1991); alcuni studi hanno rilevato un deficit nei compiti di pianificazione e di flessibilità cognitiva (Bennetto, Pennington, & Rogers, 1996; Ozonoff & Jensen, 1999; Ozonoff et al., 1991), mentre altri hanno riscontrato come queste difficoltà siano correlate più al quoziente intellettivo (QI) piuttosto che all'autismo in sé (Hill, 2004; Mari et al., 2003). Inoltre, un deficit nelle funzioni esecutive non è specifico dell'autismo; è comunemente riportata in altre condizioni neuropsichiatriche, anche se con modelli diversi, come ad esempio nella schizofrenia, il disturbo da deficit attentivo e iperattività (ADHD).

Tuttavia, gli studi esistenti in letteratura che hanno tentato di differenziare i disturbi sulla base delle prestazioni in compiti di EF hanno dato risultati contrastanti e incoerenti (Rajendran & Mitchell, 2007).

Nel 1989, Frith propose la Teoria della Coerenza Centrale Debole (*Weak Central Coherence* - WCC, Frith, 1989, 2003; Frith & Happé, 1994; Happé, 1999), riferendosi alla preferenza dell'individuo, in particolare con ASD, per i dettagli locali rispetto all'elaborazione globale. La Coerenza Centrale rappresenta un processo, dominio generale, che permette di integrare l'informazione a differenti livelli per raggiungere un significato. La teoria WCC spiegherebbe alcune caratteristiche non sociali e sociali nell'autismo, come l'attenzione per i dettagli, dimostrato con l'esecuzione di compiti delle figure incorporate (*Embedded Figures Task* – EFT) e nel test "disegno a blocchi" (Baron-Cohen, 2004). La teoria descrive come gli individui a sviluppo tipico elaborino le informazioni estraendo il significato globale, mentre le persone con autismo presentano, appunto, una debole, se non assente, spinta alla coerenza globale (Frith & Happè, 1994). Tale teoria spiegherebbe la buona abilità attentiva per i dettagli, l'elaborazione e discriminazione sensoriale-percettiva e la reattività sensoriale idiosincratica (cioè l'iper-reattività o l'iporeattività agli input sensoriali o l'interesse insolito per le caratteristiche sensoriali dell'ambiente) nell'autismo (Lai et al., 2014). Alcuni studi riporterebbero come le basi neurali coinvolte siano distribuite e dipendenti dal compito somministrato, tuttavia, esisterebbe un coinvolgimento delle cortecce sensoriali primarie a discapito delle aree associative e frontali coinvolte nel controllo top-down (Samson et al., 2012); viene riportata, inoltre una maggior sincronizzazione dei circuiti parieto-occipitali (Minshew & Keller, 2010). La teoria WCC ha evidenziato anch'essa alcuni limiti, ottenuti dai risultati contrastanti di molti studi (Rajendran & Mitchell, 2007), portando, nel corso del tempo ad un'evoluzione della teoria. In origine Frith (1989) concepiva la teoria come un deficit nell'elaborazione globale, riformulata invece come un'elaborazione locale superiore; l'ipotesi della WCC è cambiata in modo tale da non considerare le persone con autismo come aventi un deficit, piuttosto la WCC è considerato uno stile cognitivo (Happé, 1999). Infine, la teoria non cerca più di spiegare tutti gli aspetti dell'autismo, ma piuttosto è vista come una parte della cognizione nell'autismo (Happé & Frith, 2006).

1.6 Il Deficit di Cognizione Sociale

Da circa 30 anni, numerosi studi hanno supportato l'ipotesi che le difficoltà nella comunicazione ed interazione sociale osservata nell'ASD siano spiegate da un deficit della Cognizione Sociale (*Social Cognition – SC*) (Happè & Frith, 2014; Lai et al., 2014; Mazza et

al., 2017). Attualmente, il deficit di SC è considerato uno dei sintomi *core* dell'ASD; anche il DSM-5 e il DSM 5-TR (2022) indica la compromissione persistente della comunicazione e dell'interazione sociale, nonché le difficoltà nell'espressione e nel riconoscimento delle emozioni, tra le caratteristiche essenziali del disturbo.

La SC rappresenta un processo cognitivo basato sulla capacità di processare le informazioni sociali provenienti dall'ambiente esterno. Si tratta di un complesso di abilità che consentono all'individuo di anticipare il comportamento degli altri e modulare, di conseguenza, le proprie azioni (Mazza et al., 2010; Bishop-Fitzpatrick et al., 2017; Pino et al., 2017). I principali studi riguardanti la SC si sono focalizzati su una delle componenti principali, la Teoria della Mente (*Theory of Mind - ToM*). La ToM, conosciuta anche come processo di mentalizzazione, rappresenta la capacità di comprendere gli stati mentali ed emozionali propri ed altrui, incluse credenze, intenzioni, bisogni ed emozioni e riguarda la capacità di capire che gli altri potrebbero avere punti di vista diversi dal proprio (Premack & Woodruff, 1978; Mazza et al, 2017; 2020; Ziv, Hadad, Khateeb, & Terkel-Dawer, 2014). Tale capacità sono state il tema centrale della letteratura che ha tentato di caratterizzare i deficit di SC nell'ASD (Alcalá- López et al., 2018; Baron-Cohen et al., 1985, 1995; Mazza & Valenti, 2019). Furono Baron-Cohen, Leslie e Frith (1985), che in uno studio pioneristico hanno dimostrato che i bambini con ASD presentavano una difficoltà nella capacità di riconoscere lo stato mentale degli altri in compiti di falsa credenza, come il noto test di Sally e Anne. Quando i bambini con ASD venivano confrontati con i bambini a sviluppo tipico in età prescolare e bambini con sindrome di Down, fallivano nei test di falsa credenza; tale evidenza suggerisce che le difficoltà di ToM osservate nell'ASD non dipendano da un ritardo generale dello sviluppo, ma rappresenterebbero un deficit specifico nell'acquisizione della ToM (Baron-Cohen, 1989). In linea teorica, dunque, la ToM rappresenta un costrutto dominio-specifico, in linea con l'ipotesi della modularità della mente (Baron-Cohen, 1991; Mazza & Valenti, 2019). Secondo l'approccio modulare (Fodor, 1983), la ToM, in quanto abilità innata, sarebbe costituita da diversi moduli coinvolti nell'elaborazione di informazioni, automaticamente e seguono percorsi predeterminati. Sulla scia di questa prospettiva, diversi studi hanno confermato lo sviluppo atipico che avverrebbe nell'ASD, non solo nell'espressione comportamentale, ma anche per lo sviluppo di precursori alla base dell'interazione sociale, quali, il contatto visivo, l'attenzione congiunta, il gioco di finzione, il processamento di movimenti biologici e dei volti e la motivazione sociale (Lai et al., 2014).

È in tale contesto che viene formulata la teoria della "*cecità mentale"* o "*mindblindness"*, che propone come nel'ASD la ToM è compromessa con livelli di gravità variabili (Baron-Cohen et al., 1995). Tale ipotesi suggerisce una compromissione selettiva nella "lettura della mente",

rappresentando, dunque, il sintomo centrale della condizione clinica (Baron-Cohen et al., 1985, Baron-Cohen, 1995; Mazza& Valenti, 2019). Sebbene l'ipotesi di un deficit a carico della ToM abbia dominato la ricerca sull'autismo, ulteriori studi basati sull'utilizzo dei compiti di falsa credenza hanno fornito risultati contrastanti (Boucher, 2012). Alcuni studi hanno evidenziato come bambini e adulti con ASD, con abilità intellettive nella media o superiori alla media, superavano i compiti di falsa credenza, con prestazioni simili ai gruppi di controllo (Bowler, 1992; Dahlgren & Trillingsgaard, 1996; Grossman, Klin, Carter e Volkmar, 2000; Happé, 1994; Ozonoff et al. 1991; Tager-Flusberg & Sullivan, 1994). Tuttavia, molti di loro fallivano nei compiti di falsa credenza di secondo ordine, suggerendo che, sebbene abbiano sviluppato una ToM di livello inferiore, presentano un ritardo specifico nell'acquisizione di abilità di ToM più complesse (Baron-Cohen, 1989, 1991; Mazza & Valenti, 2019). La conferma dell'ipotesi secondo cui vi sia un ritardo nell'acquisizione delle abilità di ToM viene evidenziata da studi successivi (Pino et al., 2017, 2018; Mazza et al., 2017; Broekhof et al., 2015), evidenziando come la mente autistica non sia completamente "cieca" di fronte agli stati mentali altrui, ma presenterebbe tempi di acquisizione più lenti rispetto alla popolazione generale (Mazza& Valenti, 2019). La ToM in quanto abilità evolutiva, si sviluppa lungo un continuum e seguendo delle tappe ben precise (Happè & Frith, 2014; Pino et al., 2017; Mazza & Valenti, 2019). Momento cruciale nello sviluppo della ToM si verifica intorno i 3-5 anni, quando i bambini acquisiscono la capacità di attribuire false credenze e comprendono che gli stati mentali, come credenze o intenzioni altrui, possono essere diverse dalla realtà (Mazza et al., 2017; Mazza & Valenti, 2019; Pino et al, 2020). Secondo Happè e Frith (2014), una volta completato lo sviluppo delle abilità di SC, essa diventa una complessa rete di funzioni o "network", costituita da processi o "nodi" che a loro volta si ramificano in sottoprocessi, tutti interdipendenti e interconnessi.

Esistono altre prospettive teoriche che hanno cercato di spiegare le difficoltà di ToM nell'ASD, come la Theory Theory (TT) (Gopnik & Wellman, 1994). A differenza della teoria modulare, la TT sostiene che le abilità di ToM siano costituite da una serie di concetti (come credenze, desideri e intenzioni), rappresentando una teoria causale di come gli stati mentali di un individuo influenzano il suo comportamento (Apperly, 2008). Adottando questa "teoria", un individuo può fare inferenze, formulare spiegazioni e prevedere gli stati mentali e i comportamenti degli altri (Apperly, 2008; Carruthers, 1996; Mazza & Valenti, 2019; Wilkinson & Ball, 2012).

La TT descrive il bambino come un “piccolo scienziato” e collega le difficoltà di ToM ad un deficit di tipo concettuale; dunque, le difficoltà nel leggere gli stati mentali sono il risultato della mancanza di una teoria iniziale su altre persone o di una compromissione nel processo di formazione di tale teoria (Gopnik et al., 2000). Nell caso dell’ASD si osserva una carenza di “conoscenza”, ovvero le persone con autismo non possiederebbero un numero adeguato di preposizioni mentali e, probabilmente, non sarebbero in grado di formularne correttamente (Mazza & Valenti, 2019).

Un’alternativa all’approccio teorico della TT, è il modello dell’inferenza bayesiana che ha tentato di spiegare le difficoltà sociali e non delle persone con ASD (Chambon et al., 2017; Palmer et al., 2017; Zalla & Kormann, 2018; Mazza & Valenti, 2019). Questi modelli considerano le "credenze" come distribuzioni di probabilità, rappresentazioni probabilistiche di uno specifico stato del mondo, che verrebbero aggiornate in base all'esperienza (ai dati osservati): i nuovi dati modificano una credenza precedente trasformandola in una credenza a posteriori (Haker et al., 2016; Mazza & Valenti, 2019). Molto spesso le informazioni che arrivano dagli scenari sociali, sono complesse e ambigue; le precedenti conoscenze, dunque, possono facilitare a discriminare le informazioni sensoriali in arrivo attraverso il recupero, dalla memoria episodica, di informazioni sociali rilevanti (Zalla & Kormann, 2018). Le persone con ASD incontrano difficoltà nel dedurre le intenzioni degli altri quando queste non sono immediatamente visibili, ma richiedono un'interpretazione basata sul contesto, suggerendo che le difficoltà potrebbero essere legate a una ridotta capacità di integrare le informazioni sensoriali presenti con quelle precedentemente acquisite. Tale prospettiva aiuterebbe a comprendere come gli adulti con ASD siano in grado di superare i compiti di ToM, mantenendo, tuttavia, compromissioni nelle abilità sociali (Mazza & Valenti, 2019).
La Teoria della Simulazione (*Simulation Theory - ST*; Goldman, 1989; Gordon, 1986), rappresenta un terzo approccio teorico allo studio della ToM e parte presupposto che gli esseri umani comprendono gli stati mentali altrui “impersonando” e automaticamente “imitando” i loro stati mentali. Abilità di immaginazione, simulazione mentale e presa di prospettiva, rappresentano solo una parte di capacità messe in atto nelle interazioni sociali, in cui non vengono fatte delle inferenze teoriche (Shanton & Goldman, 2010; Mazza & Valenti, 2019). Secondo la ST, le persone con ASD presenterebbero una difficoltà nell’attribuire stati mentali a causa di una compromissione delle funzioni esecutive, fondamentali per i processi di simulazione (Bermúdez, 2006). Esistono diverse ipotesi su una possibile relazione tra funzioni esecutive e abilità di ToM, sebbene siano contrastanti (Pennington et al., 1997; Rajendran & Mitchell, 2007; Russel, 1997): alcuni studi ipotizzano che il funzionamento esecutivo facilitano

quello sociale (Hughes, 1998), altri che la compromissione della funzione esecutiva può avere un "effetto domino" sullo sviluppo della ToM (Jones et al., 2018; Pellicano, 2010; Russell et al., 1999) o sull'attenzione congiunta (McEvoy et al., 1993), che a loro volta influenzano lo sviluppo sociale (Bottema-Beutel et al., 2018). Secondo tale prospettiva, le persone con ASD presenterebbero una difficoltà nell'immaginare scenari alternativi e nel proiettare sé stessi in una situazione diversa da quella presente; infatti, la simulazione richiede la capacità di "fingere di essere un'altra persona", un'abilità che presuppone la distinzione tra le cose reali e le cose non reali (Frith, 1996). In linea con questa prospettiva, le persone con autismo non sarebbero in grado di inibire la propria prospettiva, con la conseguente compromissione del processo di simulazione (Goldman, 2006).

All'interno della prospettiva teorica della simulazione, vi è la scoperta dei "neuroni Specchio" che suggeriscono una chiave di lettura alternativa alle difficoltà dell'autismo. E' stato ipotizzato che i neuroni specchio siano alla base della comunicazione sociale e che permettono la comprensione dei comportamenti altrui (Keller et al., 2011; Mazza & Valenti, 2019). Queste particolari cellule cerebrali si attivano sia quando si compie un'azione sia quando si osserva compierla da qualcun altro; esse sono considerate coinvolte in diverse componenti della SC, tra cui empatia e l'elaborazione di stati emozionali (Dumas et al., 2014; Gallese, 2007; Gallese et al., 2013). Molti studiosi hanno elaborato l'ipotesi degli "specchi infranti" (Ramachandran & Oberman, 2006), secondo cui un'alterazione del sistema specchio potrebbe essere responsabile di alcuni dei sintomi chiave dell'autismo: i deficit sociali, comunicativi e d'imitazione (Altschuler et al., 2000; Saffin & Tohid, 2016; Williams et al., 2001). L'ipotesi che sintomi principali dell'ASD siano riconducibili ad una compromissione dei neuroni specchio ha suscitato controversie all'interno della comunità scientifica; come accaduto per altre prospettive teoriche, esistono risultati contrastanti in letteratura: alcuni studi evidenziano un preservato funzionamento dei neuroni specchio (Carpenter et al., 2001; Dinstein et al., 2008; Hatt et al., 2009; Sebanz et al., 2005), altri suggeriscono un miglioramento dei meccanismi di rispecchiamento in età adulta (Bastiaansen et al., 2011).

1.7 Presentazione clinica: sintomi precoci e comorbidità

La presentazione sintomatologica ASD dipende dall'età, il livello di linguaggio acquisito, che va da assenza della comunicazione verbale ad un linguaggio fluente, dal sesso, nonché dalle abilità cognitive (Hirota et al., 2023). Nonostante l'eterogeneità nella presentazione comportamentale le caratteristiche chiave della condizione riguardanti la comunicazione e

l'interazione sociale e comportamenti sensoriali, ripetitivi e ristretti possono essere riconosciute da clinici esperti (Lord et al., 2020). Negli ultimi 10 anni la conoscenza sui patterns di insorgenza dei sintomi ASD si è estesa, data l'importanza dell'intervento precoce e della relazione tra diagnosi e trattamento. In generale i sintomi dell'autismo si manifestano in maniera graduale e, sebbene la diagnosi di ASD viene fatta intorno ai 2 e 3 anni di età, la presenza di sintomi potrebbe essere evidente già a 12-18 mesi di età (Barbaro & Dissanayake, 2017). Tuttavia, nonostante la consapevolezza dei segni precoci dell'autismo sia aumentata, solo il 42% dei bambini con ASD riceve una valutazione iniziale prima dei 3 anni (Baio, 2018; Benedetto et al., 2021). In letteratura l'insorgenza dei sintomi comportamentali dell'ASD è concettualizzata attraverso tre modelli: 1) il modello ad esordio precoce, in cui i bambini mostrano ritardi e deviazioni nello sviluppo sociale e comunicativo nelle prime fasi della vita; 2) il modello regressivo, in cui i bambini sviluppano varie abilità come previsto in un periodo iniziale e poi sperimentano un declino significativo e 3) un modello di stagnazione (o plateau) dello sviluppo (Ozonoff & Iosif, 2019; Shumway et al., 2011). Quest'ultimo modello è caratterizzato da abilità iniziali intatte, ma il bambino non riesce a compiere i progressi attesi (Ozonoff & Iosif, 2019). Oltre alla variabilità nell'insorgenza dei sintomi, il ritardo nella diagnosi potrebbe essere dovuto anche alla questione della stabilità diagnostica nei bambini identificati prima dei 3 anni, considerando che molti clinici sono riluttanti a stabilire una diagnosi prima di questa età (Ozonoff et al., 2015). Questo problema è stato evidenziato da diversi studi che hanno riscontrato che la maggior parte dei bambini e dei ragazzi con diagnosi di disturbo dello spettro autistico tra i 18 e i 48 mesi di età ha continuato a manifestare la sintomatologia nei follow-up successivi; tuttavia, una percentuale di bambini con punteggi di gravità dei sintomi più severi ha mostrato una riduzione della gravità dei sintomi, passando a un anno di distanza a una fascia di punteggio più lieve (Benedetto et al., 2021). In generale vi è un comune accordo in letteratura far rientrare tra le *red flags* nei primi 2 anni di vita (Tabella 4) la compromissione nell'acquisizione o regressione delle abilità linguistiche e dei gesti communicativi, del *pointing* dichiarativo*,* la scarsa responsività sociale, come lo scarso contatto visivo e mancata risposta al nome, nonché la presenza di comportamenti atipici (movimenti stereotipati come il *flapping hands*, interesse eccessivo per alcuni oggetti e l'utilizzo ripetitivo di parole o frasi).

Tabella 4. Segni comportamentali precoci (*red flags)*

Assenza delle millestones di sviluppo affettivo
• Evitare o non mantenere il contatto oculare

• Mancata risposta al nome a 9 mesi • Ridotta espressione facciale delle emozioni dai 9 mesi • Scarso divertimento condiviso con il caregiver • Assenza di giochi interattivi dai 12 mesi • Scarso o assente utilizzo dei gesti • Mancata condivisione di interessi con altri • Scarsa o assente imitazione di altri o assenza di finzione • Assenza del pointing dichiarativo dai 18 mesi
Comportamenti Atipici
• Allineare i giochi in un determinato modo e rabbia quando l'ordine viene cambiato • Utilizzo ripetitivo di parole e frasi • Movimento delle dita, delle mani o del corpo in maniera atipica (sfregare le dita, flapping, dondolio del corpo, girare su se stesso) • Mostrare eccessivo interesse per particolari oggetti • Avere un interesse ossessivo per certi oggetti o attaccamento per oggetti insoliti • Mostrare un insolita reazione per stimoli sensoriali (arrabbiarsi per un'etichetta di abbigliamento, evitare di mangiare cibi con certe consistenze) • Mostrare un forte interesse e ricercare stimoli sensoriali insoliti (strizzare gli occhi o mostrare flapping per certe luci, strofinare eccessivamente certe consistenze, leccare o odorare oggetti)

Sebbene la sintomatologia ASD si evidenzi, generalmente, nelle prime fasi dello sviluppo, già prima dei 12 mesi, ci sono casi in cui la prima valutazione, o la prima diagnosi, avviene in adolescenza o in età adulta. Questo può accadere a bambini che non presentano un ritardo nell'acquisizione delle principali tappe di sviluppo e potrebbero giungere ai servizi clinici con la comparsa di problemi comportamentali, come difficoltà nel seguire le istruzioni a causa di un intenso interesse per le attività preferite o in presenza di condizioni psichiatriche concomitanti (Hirota et al., 2023). La diagnosi tardiva riguarda, in particolare, persone con un funzionamento cognitivo nella norma, o superiore, che sono riusciti a sviluppare comportamenti compensativi o strategie di *coping*, almeno in alcuni contesti di vita quotidiana. Tuttavia, tali strategie possono risultare insufficienti quando le richieste sociali superano le capacità adattive della persona (APA, 2022). Inoltre, è molto comune che le persone con ASD presentino anche condizioni del neurosviluppo o psichiatriche concomitanti, con una prevalenza circa del 70%, rispetto alla popolazione generale (Lai et al., 2014; Simonoff et al., 2008).

Tra le principali condizioni mediche che si riscontrano vi è l'epilessia (8-30%), sebbene i tassi di prevalenza varino a secondo dell'età, del sesso e di ulteriori condizioni mediche associate; l'esordio, di solito, avviene nella prima infanzia o nel periodo adolescenziale (Lai et al., 2014). I disturbi gastrointestinali si riscontrano dal 9 al 70%, e i sintomi più riportati sono la costipazione cronica, dolore addominale, diarrea, reflusso gastroesofageo, gastrite e infiammazione intestinale (Chaidez et al., 2013; Mannion et al., 2013). Vengono riportati, inoltre, disfunzioni del sistema immunitario (oltre il 38%) (Grigorenko et al., 2008; Heuer et al., 2008; Okada et al., 2007) e un'elevata comorbilità con allergie e malattie autoimmuni come il diabete di tipo 1 e morbo di Crohn (Chen et al., 2013). Anche i disturbi del sonno rappresentano una comorbilità elevata (50-80%) (Lai et al., 2014; Mannion & Leader, 2013). Infine, tra le sindromi genetiche (circa il 5%), le principali sono X fragile, la sclerosi tuberosa, la sindrome di Down, la sindrome di Rett, la fenilchetonuria e la sindrome di Angelman (Hodges et al., 2020; Lai et al., 2014).

Alcuni dei disturbi del neurosviluppo si possono manifestare in età scolare nelle persone con ASD (Lord et al., 2020). Il disturbo da deficit attentivo e iperattività (ADHD) e la disabilità intellettiva sono tra i disturbi più comuni nell'ASD rispetto alla popolazione generale, rispettivamente 28% e 23% nell'ASD contro il 7% e 0.7% negli individui senza ASD (Hirota et al., 2023; McKenzie et al., 2016; Mutluer et al., 2022). Il rapporto tra ADHD e ASD è molto complesso, in quanto presentano molteplici caratteristiche comuni, che possono rendere difficile la diagnosi differenziale: le due condizioni sembrano condividere aspetti genetici e comportamentali (Grzadzinski, et al., 2011; Ronald et al., 2008), tra cui difficoltà attentive, comunicative e sociali, impulsività, iperattività e comportamenti ristretti (Leitner, 2014; Rao & Landa, 2014). Nel DSM 5 (APA, 2013) viene esplicitamente riconosciuto che alcune persone ASD possono presentare pattern di sintomi paralleli compatibili con la diagnosi di ADHD (Mannion & Leader, 2013).
Condizioni particolarmente frequenti, che possono presentarsi in comorbilità con l'ASD, sono i disturbi da tic (14-38%), la sindrome di Tourette (circa il 6,5%), e disturbi motori (oltre il 79% dei casi), come anche un ritardo dello sviluppo psicomotorio, catatonia, deficit di coordinazione e pianificazione del movimento, difficoltà nella deambulazione e nell'equilibrio (Lai et al., 2014). Poiché le stereotipie motorie rientrano tra le principali caratteristiche dell'ASD (APA, 2022), la diagnosi aggiuntiva di disturbo da movimento stereotipato non viene posta quando tali comportamenti sono meglio spiegati dalla presenza di ASD; è possibile porre entrambe le diagnosi quando le stereotipie causano autolesionismo e necessitano di intervento mirato (APA, 2013).

Una questione molto importante è data da una maggiore probabilità di ricevere diagnosi errate di psicosi, o altre condizioni cliniche, negli adolescenti e adulti ad alto funzionamento, o Livello 1, che arrivano ai servizi di salute mentale causa di disturbi psichiatrici concomitanti, o al contrario, per una diagnosi differenziale (Keller et al., 2015; Lai & Baron-Cohen, 2015; Mazza et al., 2022). Tra i disturbi psichiatrici più frequenti nella popolazione ASD riguardano il disturbo ossessivo/compulsivo (DOC), disturbi psicotici come la schizofrenia, disturbi depressivi, disturbi d'ansia e di personalità (Bertelli et al., 2015; Hirota et al., 2023; Lai et al., 2019; Postorino et al., 2017). La co-occorrenza di disturbi psichiatrici in adolescenza e in età adulta contribuiscono ad una maggior compromissione della qualità di vita e del rischio suicidario (Hirota et al., 2023;o'Halloran et al.,2022). La prevalenza del DOC in concomitanza con una diagnosi di ASD è stimata dal 9% al 22% (Hollocks et al., 2019; Hossain et al., 2020; Lai et al., 2019) e la presenza della sintomatologia ossessiva-compulsiva è condivisa da entrambe le condizioni, portando spesso ad una diagnosi errata. Nonostante la sovrapposizione tra ASD e DOC, la diagnosi differenziale si basa sulla presenza o assenza delle difficoltà precoci nella comunicazione sociale, caratteristica tipica dell'ASD. Inoltre, mentre nell'ASD il comportamento ossessivo-compulsivo è vissuto in maniera egosintonica, nel DOC questo comportamento è percepito come egodistonico. Nel momento in cui vengono soddisfatti i criteri di entrambe le condizioni, è possibile effettuare la diagnosi di entrambi i disturbi in comorbilità APA, 2013; Cadman et al., 2015). Per quanto riguarda la prevalenza dei disturbi d'ansia e depressivi, la probabilità di sviluppare tale sintomatologia è quasi dell'80% e dal 2,5-47,1% rispettivamente (Keller & Bari, 2019; Hudson et al., 2019; Lai et al., 2019). La presenza di disturbi d'ansia compromette il funzionamento psicosociale, aumentando la gravità dei sintomi dell'ASD, come quelli legati all'evitamento sociale, alle interazioni sociali e ai disturbi del sonno; ciò è molto più probabile in adolescenti e adulti con ASD ad alto funzionamento in quanto maggiormente consapevoli delle loro difficoltà (Davis et al., 2011; Montes & Halterman, 2006; Sharma et al., 2018). Allo stesso modo la depressione può peggiorare la sintomatologia, specialmente in termini di ritiro sociale; inoltre, può condurre al suicidio o alla catatonia, compromettendo in modo importante l'indipendenza quotidiana (Keller & Bari, 2019).

Una delle relazioni più complesse è quella esistente tra psicosi, ed in particolare con la schizofrenia, e ASD (Mazza et al., 2022), in quanto rappresentano due condizioni che condividono caratteristiche neurobiologiche e cognitive (Bechi et al., 2019; Bertelli et al., 2015; Deste et al., 2020; Mazza et al., 2021b). Una delle caratteristiche che rendono la diagnosi differenziale complessa è la sovrapposizione sintomatologica: i disturbi dello spettro della

schizofrenia e l'autismo condividono nello specifico deficit a carico della cognizione ed interazione sociale. Tuttavia, la diagnosi differenziale viene generalmente supportata dalla presenza di comportamenti ripetitivi e limitati tipici dell'ASD, mentre i sintomi positivi della schizofrenia, come allucinazioni e deliri, non sono riscontrabili nell'autismo (APA, 2013). Anche la relazione tra ASD e disturbi di personalità (PD) appare complessa. La letteratura stima un'incidenza fino al 60% della concomitanza di un PD negli adulti con ASD (Keller, 2019; Lugnegård et al., 2011; Strunz et al., 2014). I più frequenti sono i disturbi di personalità classificati nei cluster A (schizoide e schizotipico) e C (ossessivo-compulsivo ed evitante); questo potrebbe derivare da un'inadeguata diagnosi differenziale, una vera comorbilità oppure *overlapping* di specifiche definizioni comportamentali-cognitive.

Nella valutazione di adolescenti e adulti con un sospetto di ASD, la diagnosi differenziale, o l'identificazione di eventuali comorbilità rappresenta una delle sfide cliniche. L'utilizzo di test standardizzati ed efficaci che possono supportare il clinico nel processo diagnostico, in fase di sceerning o valutativo, determinerebbe anche la strutturazione di specifici interventi riabilitativi (Mazza & Valenti, 2019).

1.8 Il processo diagnostico e valutativo

In generale, la diagnosi di ASD consiste in una combinazione di raccolta di informazioni anamnestici dai genitori e/o caregivers, l'osservazione diretta della persona e l'utilizzo di strumenti diagnostici validi e standardizzati. La raccolta delle informazioni provenienti dalle figure che ricoprono un ruolo significativo nella vita della persona consiste in un indagine sistematica della storia dello sviluppo, delle capacità cognitive, funzionali e adattive, delle abilità di linguaggio espressive e ricettive, nonché del background sociale e familiare. L'assessment diagnostico, dunque, dovrebbe coinvolgere un èquipe multidisciplinare che preveda la presenza di figure con esperienza clinica e con competenze nell'ambito dei disturbi del neurosviluppo (Fuentes et al., 2020). L'American Academy of Pediatrics raccomanda che i bambini siano sottoposti ad un precoce screening a partire dai 18-24 mesi, data la relazione che un'identificazione precoce permette l'avvio e la strutturazione di interventi altrettanto precoci (Hyman et al., 2020; Lai et al., 2014). Tra gli strumenti di screening, ideati per la rilevazione di sintomi precoci dell'autismo, tra cui anomalie sensoriali e motorie, atipie nel contatto oculare o assenza del sorriso sociale e dell'attenzione congiunta, vi è il Modified Checklist for Autism in Toddlers (M-CHAT), un questionario somministrato ai genitori, ideato per identificare bambini dai 16-30 mesi di vita (Robins et al., 2001). L'M-CHAT è stato validato in ambito clinico e mostra una buona specificità per l'ASD, tuttavia, ha una dubbia sensibilità, con il

conseguente rischio di misdiagnosi. Tuttavia, viene raccomandato dalle linee guida di non basare la diagnosi su un unico strumento, ma bambini che risultano positivi per un eventuale manifestazione sintomatologica ASD dovrebbero essere sottoposti ad una valutazione complessiva. Per quanto riguarda gli adulti, non vi sono evidenze che supportino uno screening per ASD in persone asintomatiche e in letteratura sono presenti pochi strumenti utilizzabili. A tal proposito è stato ideato *l'Autism Spectrum Quotient* (AQ; Baron-Cohen et al., 2001b; Woodbury-Smith et al., 2005), un questionario self-report somministrabile dai 16 anni in su per persone con sospetto ASD e che non presentino una disabilità intellettiva e che indagano 5 diverse aree: abilità sociali, capacità di spostare l'attenzione, attenzione ai dettagli, comunicazione e immaginazione. Uno strumento di supporto è la *Ritvo Autism Asperger Diagnostic Scale-Revised* (RAADS-R; Ritvo et al., 2011), intervista destinata a persone dai 18 anni in su, compilata dal clinico insieme al soggetto; questa permette di valutare aspetti inerenti a difficoltà nelle relazioni sociali e la presenza di interessi ristretti, comportamenti sensoriali e motori.

Tra gli strumenti considerati gold-standard per la diagnosi *l'Autism Diagnostic Observation Schedule – Second Edition* (ADOS-2; Lord et al., 2012) e *l'Autism Diagnostic Interview-Revised* (ADI-R; Lord et al., 1994). L'ADOS- 2 rappresenta uno strumento osservativo semi-strutturato che permette di indagare gli aspetti di comunicazione, interazione sociale, comportamenti ristretti e ripetitivi, mediante attività strutturate dirette alla stimolazione dei comportamenti legati all'ASD (Colombi et al., 2013; Lord et al., 2012). Lo strumento è costituito da 5 diversi moduli, selezionati in relazione alle abilità di linguaggio e all'età. Il Modulo Toddler e il Modulo 1 sono destinati ai bambini con abilità di linguaggio che vanno dall'assenza completa fino a semplici frasi, rispettivamente somministrabili a bambini 12 ai 30 mesi e 31 mesi di età o più gradi; il Modulo 2 viene somministrato a bambini che si esprimono attraverso frasi ma non sono ancora verbalmente fluenti; il Modulo 3 è destinato a bambini e giovani adolescenti che presentano un linguaggio fluente; infine, il Modulo 4, può essere somministrato a tardoadolescenti e adulti verbalmente fluenti. L'ADI-R (Lord et al., 1994), invece, è un'intervista semistrutturata rivolta ai *caregiver* che permette di raccogliere informazioni relative allo sviluppo della persona a partire dai primi anni di vita, al fine di determinare se il suo percorso di sviluppo e le caratteristiche del suo comportamento soddisfino o meno i criteri per la diagnosi di ASD. Le 93 domande strutturate indagano principalmente aree del linguaggio e della comunicazione, lo sviluppo sociale e il gioco, gli interessi e i comportamenti e comportamenti quali aggressività, comportamenti autolesivi ed eventuali tratti epilettici. Esistono, inoltre, strumenti utilizzati dal clinico a supporto della diagnosi come la

Childhood Autism Rating Scale-Second Edition (CARS2*;* Schopler et al., 2010), una scala che identifica l'ASD a partire dai 2 anni di età attraverso l'osservazione diretta del comportamento, consentendone una valutazione in termini di intensità (da lieve a grave), durata, frequenza e particolarità. Le due versioni della scala, attualmente utilizzabili nella pratica clinica, sono la *standard version* (CARS-ST) e la *high functioning version* (CARS-HF); quest'ultima indicata per individui verbalmente fluenti, di età superiore ai 6 anni, con un quoziente intellettivo > 80 (Schopler et al., 2010; Jurek et al., 2021).

Al di là della valutazione sintomatologica, nel processo diagnostico vengono considerati anche aspetti cognitivi e adattivi, con il fine di fornire un profilo per giungere ad un giudizio clinico. In particolare bisogna ricordare l'utilizzo delle scale Wechsler (per bambini e adulti) per la stima di un Quoziente Intellettivo (QI) in casi di persone che presentino un linguaggio verbale, e la Leiter International Performance scale – terza edizione (Leiter-3) utilizzabile per pazienti con compromissioni degli aspetti cognitivi e verbali, permettendo di valutare un indice generale di funzionamento cognitivo. Altri strumenti che possono essere utilizzati per definire un profilo funzionale della persona con ASD così da giungere ad una diagnosi e allo sviluppo di piani terapeutici, vi è la Vineland Adaptive Behavior Scale (VABS-II; Sparrow et al., 2005). Quest'ultima è un'intervista semistrutturata da somministrare ai genitori o caregivers della persona, utile per indagare il funzionamento adattivo quotidiano e stilare un profilo sulle abilità di comunicazione, delle autonomie personali e comunitarie, di socializzazione e motorie.

1.9 Interventi e prognosi

L'importanza di arrivare ad una diagnosi certa e valida è data dalla possibilità dell'avvio e strutturazione di un intervento. L'obiettivo della terapia è il miglioramento del funzionamento individuale, della qualità di vita, nonché della prognosi. Sebbene l'autismo sia radicato nella biologia, la maggior parte degli interventi efficaci finora sono di tipo comportamentale ed educativo; i farmaci hanno avuto solo un ruolo minore (Hirota et al., 2023; Lai et al., 2014). Infatti, non esistono evidenze di efficacia della terapia farmacologica sui sintomi nucleari dell'ASD, ossia non migliorano la comunicazione sociale, anche se sono in progresso sperimentazioni di farmaci che hanno come target diversi sistemi, come l'ossitocina, agenti colinergici e glutamminergici. Tra i farmaci più usati, comunque, ci sono gli antipsicotici, risultando utili nel controllare i comportamenti ripetitivi e l'aggressività. Gli inibitori selettivi della ricaptazione della serotonina sono utilizzati per attenuare i tratti ossessivi ed eventuali sintomi depressivi. Lo scopo della terapia, dunque, è diretto ai sintomi bersaglio per un migliore

adattamento generale e quindi ad una maggior partecipazione ai programmi terapeutico-riabilitativi (Lai et al., 2014).

Per quanto riguarda gli approcci comportamentali ed educativi, esistono approcci globali i quali hanno come target un ampio range di abilità, quali cognitive, linguistiche, sensomotorie e comportamentali adattativi, attraverso programmi intensivi a lungo termine. Essi sono raggruppati in tecniche ABA-based (*Applied Behaviour Analysis*) e Structured Teaching. L'approccio ABA, ideato da Iva Lovaas, si basa su strategie comportamentali e si è dimostrato che un intervento precoce e intensivo sembri agevolare lo sviluppo intellettivo, comunicativo, e adattativo. Lo Structured Teaching origina dal *Treatment and Education of Autistic and related Communication handicapped Children* (TEACCH), e prevede la strutturazione dell'ambiente e delle attività che possono essere comprese dall'individuo; utilizza gli interessi e le abilità visive dello stesso per migliorare le capacità più deboli e per innescare un apprendimento. Questa tecnica, dunque, mira al supporto e al miglioramento delle iniziative del soggetto. Vi sono approcci, inoltre, che si focalizzano su specifici domini cognitivi e comportamentali che potrebbero essere utili nel breve termine. Riguardo agli interventi che promuovono l'acquisizione di abilità sociali e che promuovono il riconoscimento emotivo, teoria della mente e imitazione, ad oggi non ci sono sufficienti evidenze sull'efficacia (Lai et al., 2014).

In letteratura sono presenti studi che sono stati condotti prima della diffusione dei programmi d'intervento precoce ed hanno mostrato che il 58-78% degli adulti con ASD presentavano una minor vita indipendente e di occupazione e un maggior bisogno dei servizi di salute mentale (Hirota et al., 2023; Howlin et al., 2004; Howlin et al., 2013; Lai et al., 2014). Molto spesso si dimentica che l'ASD rappresenta una condizione che ha un impatto long-life sullo sviluppo e l'adattamento di un individuo. Nonostante la crescente competenza dei servizi pubblici di salute mentale nella diagnosi precoce e nella strutturazione di interventi per migliorare il trattamento della persona ASD, la transizione dall'età evolutiva all'età adulta rappresenta una delle principali criticità della rete dei servizi nella cura delle persone con ASD (Lai et al., 2014; Valenti et al., 2023). Il termine "transizione" si riferisce generalmente a un processo pianificato e mirato che affronta i bisogni dei giovani ASD (dai 16 ai 18 anni) in modo olistico nel passaggio da un sistema di assistenza sanitaria orientato all'infanzia a uno orientato all'età adulta (King et al., 2020). Il numero di servizi per ASD diminuisce dall'infanzia all'età adulta e continua a diminuire fino alla tarda età, mentre aumenta il ricorso ai farmaci e ai servizi di ospedalizzazione (Keller et al., 2020; Howlin & Magiati, 2017). Inoltre, il periodo che segna il passaggio dall'infanzia all'età adulta è caratterizzato da profondi cambiamenti neurobiologici e

comportamentali. Durante questa fase, adolescenti e giovani adulti affrontano quelle che vengono definite le “sfide dello sviluppo”, tra cui il passaggio all'istruzione secondaria, la conclusione del percorso scolastico, l'ingresso nel mondo del lavoro, l'acquisizione dell'autonomia personale e lo sviluppo di amicizie e relazioni intime (Valenti et al., 2023). Questi cambiamenti che avvengono durante lo sviluppo sono resi difficili anche dai deficit di interazione sociale, che diventano più evidenti con il raggiungimento dell'età adulta, un'età in cui una persona dovrebbe raggiungere una maggiore indipendenza. L’intensità e la necessità di supporto vengono attribuite sulla base del DSM 5 (APA, 2013) e vengono differenziate in 3 livelli di gravità (Livello 1, 2, 3). Il livello di gravità basato sull'intensità dei bisogni di supporto fornisce un quadro del funzionamento della persona con ASD, tenendo conto di variabili quali i problemi comportamentali, il funzionamento cognitivo e adattivo, le comorbidità e le variabili legate al contesto e all'ambiente (Mehling et al., 2016; Valenti et al., 2023). L’esito della fase di transizione nelle persone con ASD ha un impatto rilevante sui caregivers che rappresentano il primo sistema di supporto continuo. Diversi studi dimostrano che i caregiver di adulti con ASD hanno più sintomi depressivi, peggiori condizioni di salute e minori capacità funzionali, con un impatto negativo sulla qualità della vita (Wong et al., 2020).

Di conseguenza, è evidente la necessità di competenze e risorse per fornire una valutazione clinica accurata e adeguata all'età adulta, nonché servizi di intervento appropriati per le persone con ASD di tutte le età da una prospettiva di sviluppo (Valenti et al., 2019).

2. Metodi statistico-psicometrici nel setting clinico

2.1 Psicologia e Psicometria

Nel 1918, lo psicologo Edward Lee Thorndike scriveva "*conoscere bene qualcosa implica conoscerne la quantità",* tuttavia, ancora oggi la psicometria non viene vista da molti professionisti nel campo della psicologia come la *via* attraverso la quale approcciare i diversi oggetti di ricerca e/o intervento (Sartori, 2008). In generale la psicometria viene definita come disciplina fondamentale della psicologia che si occupa della misurazione dei fenomeni psicologici attraverso strumenti standardizzati. La psicometria, in realtà, può essere intesa come l'anello mancante tra la psicologia, intesa come scienza dei comportamenti umani e dei costrutti mentali sottesi ai suddetti comportamenti, e la statistica, come disciplina in grado di fornire un supporto alla psicologia ad inferire, dal comportamento osservato e quantificato, i costrutti sottostanti (Sartori, 2008).

Il termine psicometria deriva dalle parole greche "*psiché*" (anima) e "*metria*" (misura), riflettendo così l'obiettivo di quantificare dimensioni astratte della mente umana.

Il primo a introdurre il concetto di psicometria e di misura in psicologia è stato Christian Wolff nei suoi due lavori "*Psychologia empirica"* (1732) e "*Psychologia rationals"* (1734), per poi comparire in maniera sporadica per tutto il Settecento. Tuttavia, si può dire che l'indagine psicometrica ha avuto due punti di partenza: 1) la scoperta di leggi generali che mettono in relazione il mondo fisico con il comportamento osservabile e 2) l'obiettivo di esplorare e testare alcune ipotesi sulla natura delle differenze individuali utilizzando i test psicologici (Anunciação, 2018; Craig, 2017; Furr & Bacharach, 2008). Circa un secolo dopo da Wolff, il filosofo e psicologo tedesco Johann Hebart (1824, 1825) iniziò ad affermare che i ragionamenti psicologici dovevano essere di natura matematica ed è considerato il creatore di modelli matematici della mente nel suo campo della psicofisica. Ciò influenzò i lavori di Wilhelm Wundt e del suo laboratorio a Lipsia, a cui viene attribuita la fondazione della scienza della psicologica (Carpenter, 2005; Craig, 2017). È in questo periodo che Francis Galton (1822-1911) fu il primo ad applicare la statistica alla psicologia e viene definito il "padre della psicometria". Il lavoro di Galton, interessato in particolare alla misurazione dell'intelligenza, così come Alfred Binet (1857-1911) in Francia, influenzò il lavoro di James McKeen Cattell (1860-1944), che coniò il termine *test mentale*. Cattell fu il primo psicologo a scrivere di psicometria nel 1886 con una tesi intitolata "*Psychometric Investigation*", in cui studiò quello che oggi conosciamo come *effetto Stroop*. Herbart, tramite Wundt, e Galton, tramite Cattell,

rappresentano i pionieri dei moderni test psicologici (Craig, 2017). È nel XX secolo che venne fondata la *Psychimetric Society* da Louis Thurstone (1936) e la psicometria inizia ad essere vista come una scienza a sé.
L'evoluzione e il lavoro della misurazione degli "oggetti" psicologici, inizia a prendere spazio nella psicologia clinica, che con il perfezionamento delle misure di valutazione delle capacità cognitive, si focalizzerà alla misurazione di altri elementi di funzionamento psicologico. Nel 1936, con la nascita della pubblicazione del volume *Psychometrika*, si afferma la quantificazione in psicologia e la sua misurazione. Si può dire che l'evoluzione del concetto della psicometria, e dei tentativi di applicarla, si intreccia con la questione sempre aperta della quantificazione psicologica.

2.2 Misurazione e Testing Psicologico

"Compito fondamentale di tutte le attività scientifiche è la *misurazione*…" scriveva Elena Zambianchi (1994,13) nel suo *Elementi di Psicofisica.* È la misurazione, dunque, che caratterizza le discipline scientifiche e la psicologia non ne è da meno. La misurazione consente di formulare leggi o confermare teorie con cui vengono spiegati, giustificati e previsti fenomeni empirici (Luce & Krumhansl, 1988; Sartori, 2008). Con la misura si rende dunque esplicita attraverso una funzione la relazione tra un'entità non numerica (ad es. il peso, una caratteristica di personalità o un'abilità) e una quantità numerica che la definisce (Giannini et al., 2014).
Tuttavia ancora oggi il problema della misurazione in psicologia suscita dibattiti tra coloro che sono a favore della quantificazione e che ritengono che sia possibile utilizzare per lo studio del comportamento umano i cosiddetti metodi quantitativi, e coloro che ritengono invece i fenomeni psicologici non misurabili. Almeno fino all'Ottocento la psicologia sottostava al "pregiudizio" secondo cui la *psiche*, essendo immateriale non può essere misurata. È grazie ai lavori pioneristici di Weber (1834), Fechner (1860), Helmholtz (1867), Wundt (1873,74) e Galton (1883) e altri autori di tutto il Novecento che la *psiche* sia diventata, in qualche modo, misurabile; la psicometria, in particolare, cerca di giungere ad una quantificazione valida e attendibile, nonché di esprimere relazioni tra le variabili di pertinenza tramite le rappresentazioni matematiche (Cristante & Mannarini, 2003; Sartori, 2007; 2008). Tuttavia, nonostante gli sforzi, è da notare che in psicologia non esiste nulla che può essere paragonato ad un'unità campione che possa essere utilizzata come unità di misura della *psiche*. Il problema della misura delle caratteristiche psicologiche viene affrontato ormai da lungo tempo e ancora oggi rappresenta un campo di indagine e di dibattito.

Le persone differiscono per abilità, comportamenti, tratti e i primi tentativi di misurare e analizzare tali diversità sembra risalire già dal 2200 a.C. in Cina (Sartori, 2008; Thorndike, 1990). Successivamente, anche filosofi e medici greci, come Ippocrate e Galeno, si interessarono alle differenze individuali, sviluppando una classificazione su quattro "tipi psicologici", associati alla predominanza di un particolare *umore* (fluido): sangue, bile, atrabile e flemma (Brunas-Wagstaff, 1999; Sartori, 2008). Tali tentativi, proseguiti anche nel corso del Novecento, come ad esempio le tipologie di personalità individuate da Sheldon (1942) sulla base della morfologia, testimoniano la volontà di operare una classificazione delle persone sulla base della presenza/assenza di caratteristiche psicologiche. Si può dire che con Galton, e i suoi studi sull'intelligenza, si assiste ad una sistematizzazione della ricerca sulle caratteristiche psicologiche degli individui. Galton con i suoi lavori ha contribuito a rendere credibile l'idea che, attraverso procedure rigorose, oggettive e standardizzate, si possano ottenere risultati solidi anche in psicologia; ha, inoltre, introdotto i procedimenti per il calcolo della regressione e correlazione (quest'ultima ripresa da Karl Pearson nel 1896), ed è a lui che si fa risalire anche il metodo delle libere associazioni, ripreso in seguito da Jung e la scuola psicoanalitica (Sartori, 2008). In questo framework di ricerca sulla possibilità di differenziare le persone sulla base dell'intelligenza vengono svolti alcuni dei lavori che cambieranno l'ambito della psicologia; nel 1894, Cattel apre il Laboratorio di Psicologia alla Columbia University e ampliò gli strumenti dei suoi predecessori chiamandoli "test mentali"; nel 1905, Alfred Binet e Theodore Simon pubblicarono in Francia il primo vero test di intelligenza per bambini, con 30 *item* di difficoltà crescente; nel 1908 in occasione della revisione del test viene introdotto il concetto di *età mentale* e, nel 1911, il test viene esteso per la popolazione adulta; sei anni più tardi, la diffusione del metodo viene adattato alla popolazione americana da Lewis Madison Terman della Stanford University, con l'introduzione del concetto di Quoziente Intellettivo (QI), per poi giungere al 1960 con la nascita del test d'intelligenza Stanford-Binet (Sartori, 2008). Nello stesso periodo in cui si lavorava alla costruzione del test d'intelligenza, lo psicologo inglese Charles Edward Spearman propose una teoria sull'intelligenza che fornirà la base scientifica per l'introduzione di una metodologia statistica più utilizzata nella costruzione di un test psicologico, ovvero l'analisi fattoriale, messa a punto, nel 1938, da Louis Leon Thurstone.

L'assessment in psicologia, tuttavia, inizia a definirsi intorno i primi anni del XX secolo, con appunto, lo sviluppo dei test per la valutazione delle caratteristiche quali l'intelligenza e la personalità. Tra il 1930 e 1940 sono stati introdotti strumenti, come il Test di Rorschach (Klopfer, 1936), il Minnesota Multiphasic Personality Inventory (MMPI) (Hathaway & McKinley, 1940). Ma è nel Secondo dopoguerra che l'assessment nella psicologia diventa

centrale nella pratica psicologica, tanto che solo nel 1940 furono somministrati 60 milioni di test a 20 milioni di soldati e civili (Goldenberg, 1983). D'altra parte, durante gli anni '70, il rapporto assessment e psicologia inizia a diminuire nei setting di salute mentale, assistendo ad una drastica riduzione dell'uso delle tecniche tradizionali di valutazione e allo sviluppo dell'attività psicoterapeutica (Butcher et al., 2006). Uno dei principali dubbi evidenziati dalla comunità psicoterapeutica, in particolare con la diffusione del paradigma analitico, era appunto la quantificazione dei costrutti psicologici, ritenendo riduttive le procedure di raccolta e di analisi di dati numerici sulle funzioni normali e patologiche dei soggetti in trattamento. Tale prospettiva cambierà con l'avvento e l'affermarsi del cognitivismo (Giannini et al., 2014). Anche in Italia l'utilizzo di test nella pratica clinica sta diventando sempre più centrale, rilevandosi strumenti di misura in grado di supportare decisioni e programmare gli interventi. Come esistono diverse definizioni di misurazione, ce ne sono altrettante per test psicologico; in generale un test (o reattivo) psicologico può essere definito come una situazione standardizzata in cui il comportamento di una persona, adeguatamente campionato, viene prima osservato e poi descritto per mezzo della misura, anch'essa standardizzata e oggettiva (Anastasi & Urbina, 1997; Boncori, 1993; Sartori, 2008). Attraverso l'applicazione di un test psicologico, con caratteristiche metriche di validità e attendibilità, è possibile la rilevazione delle caratteristiche possedute in maniera diversa dagli individui, come anche confrontore due prestazioni diverse di uno stesso individuo, o ancora la prestazione del singolo soggetto e la norma statistica del suo campione di riferimento (Anastasi & Urbina 1997; Sartori, 2008). In conclusione, l'applicazione di un test psicologico, valido, attendibile e standardizzato su un ampio campione rappresentativo della popolazione cui il test è destinato permetterebbe di raggiungere l'obiettivo desiderato dai precedenti autori citati, la classificazione delle persone sulla base del grado con cui possiedono una determinata caratteristica psicologica (Sartori, 2008).

2.3 Psicometria nella ricerca psicologica clinica

La psicometria clinica è uno strumento indispensabile in molteplici campi della ricerca e dell'intervento psicologico e psichiatrico, utile ai fini della valutazione diagnostica, alla programmazione e alla valutazione di trattamento psicologico e farmacologico, al monitoraggio dei processi di sviluppo tipici e atipici. E' ormai noto come anche istituzioni, come l'Istituto Superiore di Sanità in Italia, stiano adottando sempre più politiche di "pratiche basate sull'evidenza" (*evidence-based practice)*, richiedendo ai professionisti che i diversi trattamenti siano giustificabili secondo evidenze accumulate dalla ricerca a livello nazionale e internazionale.

Se uno dei principali utilizzi della psicometria nel contesto clinico è quello di stabilire se una persona presenta una diagnosi, un problema o un'attitudine in un determinato tempo, dall'altro l'obiettivo principale è stabilire in che misura la persona possiede la qualità da testare (Fairfax, 2017) e tale giudizio è basato su qualità empiriche (di validità e attendibilità). Tuttavia, ogni qualità, condizione o attitudine e ogni altra variabile psicologica sono considerate come latenti e non osservabili (Barrett, 2003; Fairfax, 2017). Ancora oggi esiste una diatriba sull'orientamento degli psicologi pro e contro la misurazione, tra chi ritiene possibile una misurazione in senso fisico e chi invece la ritiene una riduzione della complessità "delle qualità umani a semplici numeri" (Cristante & Mannarini, 2003; Ercolani & Perugini, 1997; Sartori, 2008). Ma per rispondere alla questione se le caratteristiche psicologiche esistano, in una certa quantità e se si possono misurare, basti pensare che nella vita di tutti i giorni le persone descrivono se stesse e le altre in termini di caratteristiche psicologiche interiori (come ad esempio nell'essere *più* o *meno* intelligenti o estroversi); inoltre, la descrizione o il confronto delle caratteristiche psicologiche (non visibili) tra persone diverse avviene attraverso l'osservazione del comportamento, del modo di fare e reagire alle diverse situazioni (Sartori, 2008). I comportamenti, dunque, diventano gli indicatori della presenza di qualcosa che non può essere visto direttamente. Anche la concezione esistente, in ambito psicoterapeutico, che sia possibile la modificazione non solo del comportamento manifesto ma anche delle caratteristiche psicologiche in sé, implica l'esistenza di quest'ultime come *cose* misurabili(Sartori, 2008). Questo aspetto apre le strade ad un'altra questione: dal momento che la caratteristica psicologica esiste come *costrutto*, per giungere ad una sua definizione è necessario l'utilizzo di cose esperibili che siano in relazione con il costrutto stesso, ovvero gli indicatori. In base alla scelta di indicatori si giungerebbe ad una definizione diversa del costrutto di cui si vuole ottenere una misura valida e attendibile.

L'importanza di una buona qualità psicometrica nella ricerca psicologica clinica permetterebbe una misurazione dei costrutti clinici. In precedenza, abbiamo riportato come la psicometria potrebbe essere considerata come l'anello di giunzione tra la psicologia e la statistica; così, se volessimo parlare in termini statistici, una scarsa qualità psicometrica può portare ad una sottostima di effetti importanti ma anche ad una sovrastima di effetti "*banali"*(Furr, 2020). Dunque, una scarsa qualità di misurazione può portare ad ambiguità o interpretazioni inaccurate dei risultati ottenuti dal *testing* psicologico e, di conseguenza, a conclusioni erronee sulle ipotesi di partenza. Ciò si applica sia se si parla di studi sperimentali o osservazionali, sia nell'utilizzo delle misurazioni per effettuare assessment cognitivi, utilizzo di questionari e interviste e nelle diagnosi cliniche.

2.4 Le proprietà psicometriche: Validità, Affidabilità e Dimensionalità

La qualità dei test psicologici e delle misure di assessment ha un impatto non solo in termini di diagnosi clinica ma anche sulla valutazione di interventi sanitari, dei servizi e nella ricerca clinica.

Generalmente, il concetto di validità di un test può essere definito come "*il grado i cui il test misura davvero ciò che si propone di misurare"* (Anastasi &Urbina, 1997). Tuttavia, la definizione di validità dei test si è evoluta per tutto il Ventesimo secolo, tanto da aprire dibattiti sulla stessa definizione. La definizione appena riportata, infatti, solleva questioni riguardo al fatto che un test non può essere considerato valido sempre e comunque; è possibile avere test che siano validi per un certo tipo di popolazione (ad esempio, bambini, adulti, adolescenti, o maschi e femmine), in determinati contesti. Le diverse definizioni che si sono susseguite negli anni dai diversi ricercatori e psicometristi, inoltre, ha portato alla concettualizzazione della validità come "costrutto unitario" (*Unitary Concept of Validity*), ovvero come caratteristica unitaria di un test, quando*"...l'appropriatezza, la sensatezza e l'utilità delle inferenze basate sui punteggi dei test sono inseparabili*…" (Messick, 1995; Sartori, 2008). Ad oggi, tuttavia, il modello più usato è il "*Trinitarian Model"* che prevede la valutazione combinata delle validità di contenuto, criterio e costrutto, sebbene l'American Psychological Association (APA; 1954) ha individuato quattro tipi di validità: validità di contenuto, di costrutto, concorrente, predittiva. Così, riprendendo la definizione riconosciuta anche come APA standard, la definizione di validità riguarda non solo la proprietà di un test ma anche l'interpretazione dei punteggi del test, dunque viene definita come "*la considerazione più importante nella valutazione di un test. Il concetto si riferisce all'appropriatezza, alla sensatezza e all'utilità delle specifiche inferenze fatte a partire dai punteggi del test. Si definisce validazione il processo attraverso il quale vengono accumulate evidenze a supporto di tali inferenze*". Il concetto di validità è legato a quello di attendibilità, tanto che alcuni autori sostengono che "la validità richiede attendibilità" (Sartori, 2008). Con attendibilità si riferisce alla precisione con cui i punteggi di un test rappresentano un costrutto psicologico e, più in generale, ci si riferisce al tentativo di stabilire se una misura "produce gli stessi risultati in prove ripetute", o ancora, da un punto di vista statistico, che le misure siano libere dall'errore causale e siano stabili nel tempo. Poiché i ricercatori non conoscono i punteggi veri o l'errore degli individui, non sapranno mai l'affidabilità dei punteggi osservati di una misura; tuttavia, esistono metodi per la stima dell'attendibilità, tra cui il metodo test-retest e di coerenza interna (Furr, 2020). Mentre il metodo test-retest stima l'attendibilità attraverso la coerenza di una stessa misura nel tempo, il

metodo della coerenza interna stima la coerenza tra differenti parti di una misura (split-half e alpha di Cronbach). L'alpha di Cronbach è il metodo più utilizzato nella stima dell'attendibilità e della coerenza interna di una misura e si basa (1) sul grado in cui gli item di una misura sono coerenti con gli altri e (2) il numero degli item (Furr, 2020). Tra i metodi esistenti per la valutazione dell'attendibilità vi è anche l'utilizzo dell'omega di McDonald (1999) basata sul metodo della analisi fattoriale. Un altro aspetto fondamentale da tenere in considerazione per una buona qualità di misurazione è la questione sulla dimensionalità di una misura. La dimensionalità si riferisce al numero e alla natura dei costrutti psicologici che influenzano le risposte ad una misura (Furr, 2018; 2020). Alcune misure possono riflettere un unico costrutto psicologico o presentare una dimensionalità più complessa, con due o più costrutti. Questo aspetto è importante per due motivi; il primo è che ha implicazioni per lo scoring della misura, in quanto ogni punteggio potrebbe riflettere una dimensione, quindi ogni punteggio si basa su un set di item che hanno una struttura unidimensionale (se un punteggio si basa su item multidimensionali non avrebbe un chiaro significato psicologico) (Ziegler & Hagemann, 2015; Furr, 2020). Il secondo aspetto, inerente all'importanza teorica, è che la dimensionalità permette la comprensione del costrutto rappresentato da una misura. Nella pratica di ricerca come nella clinica gli aspetti di validità, attendibilità e dimensionalità rappresentano i punti di partenza nello sviluppo, nella valutazione o revisione di una misura.

2.4.1 Item Response Theory: Modello di Rasch

I modelli statistici, che descrivono le relazioni tra le risposte agli item e le dimensioni psicologiche che il test intende misurare, permettono la valutazione delle proprietà psicometriche di una misura, della sua validazione o revisione. In particolare, sono due i modelli psicometrici principali nella costruzione di test psicologici: la Teoria Classica dei Test (TCT) e l'Item Response Theory (IRT). La TCT ha dominato la ricerca scientifica fino al 1980 e può essere fatta risalire al lavoro di Charles Spearman all'inizio del XX secolo. La TCT, denominata anche come "Teoria dell'Errore", si basa sulla concettualizzazione che i punteggi ottenuti da una misura possa essere scomposto in due componenti, una *vera* e una di *errore* (Craig, 2017). Inoltre, il punteggio vero non è direttamente osservabile e può essere inferito dal punteggio vero senza la componente di errore; quest'ultima può essere causale o sistematico. L'errore causale è presente nelle singole misure e non è prevedibile; l'errore sistematico agisce in maniera costante su tutti i soggetti e strumenti di misura, ed è in parte controllabile. Nel framework della TCT il concetto di attendibilità con cui uno strumento permette di valutare l'abilità del soggetto è legata al numero delle prove; al crescere del numero degli item cresce

l'attendibilità dello strumento. Secondo la TCT, inoltre, un individuo in possesso di una maggiore abilità dovrebbe rispondere correttamente a un maggior numero di domande. Gli item, dunque, vengono selezionati sulla base della loro difficoltà e capacità di discriminare tra i soggetti, definendo la difficoltà di un item come la proporzione di soggetti, che in un campione specifico, risponde correttamente all'item in questione (Barbaranelli & Natali, 2005).

Nella TCT, l'abilità di un soggetto è definita esclusivamente dal punteggio ottenuto al test, mentre la difficoltà di un *item* è determinata dalla proporzione di soggetti che rispondono correttamente all'*item* rispetto al totale. Dunque, se un test fosse particolarmente semplice, i punteggi ottenuti dai soggetti risulterebbero molto alti, suggerendo alti livelli di quella abilità; al contrario se il test fosse composto da item difficili si otterrebbe l'interpretazione inversa. Inoltre, anche la stima della difficoltà degli item dipende dal livello di abilità del campione a cui è stato somministrato (MacDonald & Paunonen, 2002). Uno dei principali limiti della TCT è l'impossibilità di separare le caratteristiche del soggetto da quelle del test stesso (Craig, 2017). Questo comporta difficoltà nel confrontare le abilità dei soggetti sottoposti a diversi test e nel confrontare le caratteristiche degli item di un test somministrato a gruppi di soggetti diversi (Lucidi, 2013).

Un framework alternativo nella concettualizzazione e valutazione dell'attendibilità e validità è rappresentato dal modello dell'*Item Response Theory* (IRT), conosciuto anche come *Teoria del tratto Latente*, utilizzato in particolare nello sviluppo, validazione e revisione di misure in molte aree della psicologia, inclusa la psicologia clinica (Furr, 2020). L'IRT si tratta di una famiglia di modelli sviluppati intorno gli anni '50-'60 basati sui lavori del matematico danese George Rasch (1960) e dai lavori di Fred Lord (1968) e Alan Birnbaum (Birnbaum, 1968; Bock, 1997; Lord & Novick, 1968).

In generale, i modelli IRT hanno l'obiettivo di calcolare la probabilità che un soggetto risponda in un determinato modo ad un item, riflettendo in questo modo il livello di abilità posseduto dal soggetto (la caratteristica latente) e delle proprietà degli item; nei modelli IRT, dunque, viene inclusa la componente "persona" e "item" (Baylor et al., 2011; Furr, 2020). Rispetto alla TCT, con l'IRT l'attendibilità viene considerata come una variazione di un insieme di punteggi, ovvero un item o una scala potrebbero effettivamente differenziare tra persone che presentano punteggi bassi sul costrutto rilevante, ma potrebbe non differenziare le persone con punteggi più elevati (Furr, 2020). I modelli IRT utilizzano le informazioni sui singoli item, detti parametri, per spiegare la relazione tra il modo in cui il soggetto risponde a ciascun item e i livelli dei tratti latenti posseduti dal soggetto. I parametri utilizzati sono rappresentati dalla *difficoltà di un item*, il *livello di discriminazione di un item* e il *guessing effect* (la possibilità

che il soggetto indovini a caso). Sulla base del numero di parametri degli item e il numero di opzioni di risposta che vengono considerati, vengono distinti anche diversi modelli IRT. I modelli IRT possono essere grossolanamente classificati in modelli logistici a 1 Parametro (1PL), logistici a 2 Parametri (2PL), logistici a 3 Parametri (3PL). Il modello di Rasch, noto anche come modello logistico a un parametro (1PL), è uno dei più semplici e diffusi nella teoria della risposta all'item (Rasch, 1960). Esso considera la *difficoltà dell'item* l'unico parametro per spiegare le relazioni tra l'item, il tratto latente e la risposta del soggetto all'item (Baylor et al., 2011). Il Modello di Rasch è applicabile a item dicotomici e assume che gli item abbiano lo stesso il livello di discriminazione, ovvero che tutti gli item abbiano lo stesso potere discriminante, mentre il parametro di *Guessing* non viene considerato. In questo modello la probabilità di rispondere correttamente all'item dipende dal tratto latente del soggetto e dalla difficoltà dell'item; se quest'ultimo supera il livello di abilità posseduta dal soggetto, la probabilità di rispondere correttamente è inferiore al 50%. Al contrario, per soggetti con un tratto latente superiori al valore di difficoltà di un item, avranno una maggiore probabilità di rispondere correttamente all'item. La difficoltà dell'item è quel punto dell'intervallo di tratti in cui l'item è più efficace nel discriminare tra i livelli di tratti (cioè, è più probabile che i clienti con livelli di tratti diversi rispondano all'item in modo diverso) (Baylor et al., 2011). Tale relazione, tra l'abilità della persona (θ), la difficoltà dell'item (β) e la probabilità di una risposta corretta, può essere rappresentata graficamente ed è nota come *Item Charateristic Curve* (ICC). La scala sull'asse delle ascisse (x) rappresenta sia l'abilità del soggetto sia la difficoltà dell'item per il tratto stimato dallo strumento; l'asse delle ordinate (y) rappresenta la probabilità di risposta corretta all'item. Le ICC per ogni item hanno forma a S e dimostra che all'aumentare dell'abilità (asse delle x), aumenta la probabilità di rispondere correttamente all'item (asse delle y). Teoricamente l'abilità (θ) va da $-\infty$ a $+\infty$, ma nelle applicazioni è solitamente compresa tra -3 e + 3 (Baylor et al., 2011; Kline, 1996).[1] Nel Modello di Rasch il parametro che influenza la ICC è il livello di difficoltà dello stesso.

[1] L'equazione matematica del modello 1PL è $P_{ni} = [e^{(\theta n - \beta i)}]/[1 + e^{(\theta n - \beta i)}]$. Il modello 1-PL stima che la probabilità di una risposta corretta da parte della persona n all'item i (P_{ni}) è una funzione della differenza tra l'abilità della persona (rappresentata da θ) e la difficoltà dell'item (rappresentata da β). Il simbolo e è la base del logaritmo naturale. All'aumentare della differenza tra l'abilità della persona e la difficoltà dell'item (cioè, θ-β), aumenta la probabilità di una risposta corretta. Al diminuire di questa differenza, diminuisce la probabilità di rispondere correttamente (Baylon et al., 2011).

Nell'utilizzo dei modelli IRT e del Modello Rasch nella costruzione e revisione di scale, bisogna rispettare alcune assunzioni di base, quali la *unidimensionalità, l'indipendenza locale, monotonicità* e l'assenza del *fattore velocità* (Barbaranelli & Natali, 2005). L'assunto di unidimensionalità si riferisce che gli item di uno strumento misurano un solo tratto o costrutto; se vengono inclusi item che rappresentano costrutti diversi verranno introdotti variabili diverse che confondono la stima del costrutti di interesse (Baylor et al., 2011). Con indipendenza locale intendiamo che gli item sono indipendenti o non correlati tra loro, e che la dipendenza tra gli item è determinata solo dal tratto d'interesse. In linea di principio, ogni volta che la risposta a un item dipende o è correlata alla risposta a un altro item, l'ipotesi di indipendenza locale è violata (Baylor et al., 2011). Il principio di monotonicità negli IRT assume, invece, che la probabilità di rispondere correttamente all'item aumenta monotonicamente all'aumentare del livello di abilità; tutti gli errori commessi sono attribuibili alla mancanza di conoscenza dell'argomento e non ad altri fattori, come la mancanza di tempo; se così non fosse, potrebbe esser violata l'assunzione di monodimensionalità, che esclude assunto l'influenza del fattore velocità sulla prestazione. Un aspetto distintivo degli IRT è la possibilità di calibrare uno strumento, affinchè misuri con maggior precisione un determinato livello di abilità ritenuto di particolare interesse. Ciò è reso possibile dal fatto che l'errore standard di misura non è costante lungo l'intera gamma di possibili valori del tratto che si va a misurare, ma varia in funzione del livello di tratto. Dunque è possibile ridurre il numero di item da somministrare per stimare adeguatamente un'abilità senza compromettere la precisione della stima dell'abilità, limitando così bias come l'affaticamento del soggetto o gli effetti di apprendimento dovuti a una sequenza troppo lunga di item. Uno dei vantaggi dell'applicazione dei modelli IRT, dunque, è la costruzione di misure più brevi e più attendibili (Pilkonis et al., 2011).

Un concetto di questi modelli da tenere in considerazione è la funzione informativa dell'item, *Item Information Function* (IIF), che esprime la precisione con cui l'item misura l'abilità in un preciso livello di tratto. Questa funzione fornisce un'idea della posizione sulla scala delle abilità dove l'item è più informativo; nel modello di Rasch il punto del tratto in cui l'informazione dell'item è massima corrisponde al livello di difficoltà dell'item stesso.

Attualmente l'analisi degli item può avvalersi di software specializzati che forniscono statistiche sull'adattamento degli item al modello. Questo processo è cruciale per verificare se le risposte dei soggetti seguono l'andamento previsto dall'ICC; infatti, se le risposte ottenute dai soggetti non seguono tale andamento, ovvero soggetti con livelli di abilità più elevati mostrano una probabilità maggiore di rispondere correttamente a un item, il modello IRT non è in grado di stimare il tratto (Baylor et al., 2011). Per questo motivo, dopo aver verificato che

gli item rispettino le assunzioni di unidimensionalità e indipendenza locale, una prima analisi è la valutazione dell'adattamento degli item al modello. L'adattamento dell'item viene valutato calcolando la differenza tra l'ICC osservato (dalle risposte effettive ottenute nel processo di sviluppo dello strumento) e l'ICC previsto (teorico).

Lo scarso adattamento dell'item può verificarsi a causa di distorsioni, come l'ambiguità dell'item o che l'item mostri un funzionamento differenziale tra i sottogruppi (DIF). Secondo il modello IRT le risposte agli item devono essere basate sull'abilità e non da un modello di risposta generato da caratteristiche del soggetto come età, genere, stato socioeconomico. Mediante l'utilizzo del DIF (tra cui il test statistico più utilizzato è il Mantel-Haenszel test), si analizza se i parametri dell'item siano gli stessi tra i sottogruppo della popolazione, dopo aver controllato le differenze di abilità. Le analisi DIF permettono di identificare e rimuovere item che possono introdurre distorsioni.

In conclusione gli approcci IRT si può costruire uno strumento in grado di misurare in maniera precisa e attendibile il livello di abilità di interesse, sulla base di un numero ridotto di item, specifici per il costrutto target.

2.4.2 Analisi fattoriale

Come precedentemente descritto, la valutazione della dimensionalità di un test è un passo importante nello sviluppo, o revisione, degli strumenti che vengono poi utilizzati nella ricerca e nella pratica clinica. Indagare il numero di fattori latenti o di dimensioni alla base dei dati multivariati è un aspetto importante nella costruzione e nella validazione degli strumenti in psicologia (Golino et al., 2020; Timmerman & Lorenzo-Seva, 2011). La determinazione del numero di fattori, o della unidimensionalità di uno strumento, permette di rimuovere item che rappresentano costrutti diversi ed è rilevante anche nella costruzione di teorie psicologiche, poiché alcune aree (ad esempio, la personalità e l'intelligenza) si basano molto sull'identificazione di strutture latenti per comprendere l'organizzazione dei tratti umani (Garcia-Garzon, Abad, & Garrido, 2019b). *L'Analisi Fattoriale* costituisce un importante strumento di indagine della validità di un test e permette di stabilire se gli item utilizzati per la misura di una determinata caratteristica la rappresentino e, in caso contrario di modificare o eliminare gli item che hanno a che fare poco o nulla con quella caratteristica (Sartori, 2008). L'Analisi Fattoriale, inoltre, permette di esplorare le relazioni tra gli item e trovare quelli più strettamente correlati tra loro. Gli item che sono unidimensionali (misurano tutti lo stesso costrutto) si caricheranno su un singolo fattore che rende conto delle relazioni osservate tra gli item.

In base agli obiettivi, ai metodi e alle tecniche che si intende utilizzare, è possibile distinguere in Analisi Fattoriale *Esplorativa* (*Explorative Factor Analysis – EFA)* e *Confermativa (Confermative Factor Analysis – CFA)*. L'EFA viene utilizzata quando il ricercatore non ha ipotesi o un modello teorico di partenza riguardo la dimensionalità della misura e per conseguire due obiettivi principali, la spiegazione e la riduzione di dati (Floyd & Widaman, 1995; Furr, 2020). Per spiegazione dei dati si intende l'identificazione delle dimensioni sottostanti di un dominio valutato da uno strumento; così, uno strumento viene analizzato in fattori per identificare dimensioni separabili (i costrutti teorici) all'interno del dominio (Floyd & Widaman, 1995). Le dimensioni che derivano dall'analisi rappresenteranno le sottoscale di una misura. Per quanto riguarda la riduzione dei dati, l'insieme di variabili viene combinato in indici sintetici; ovvero l'obiettivo principale è la ponderazione delle variabili misurate in modo che un ampio numero di variabili correlate possa essere ridotto.

In generale per raggiungere tali obiettivi, l'Analisi Fattoriale utilizza la matrice delle correlazioni o delle covarianze delle variabili misurate, per identificare le variabili latenti, o fattori, che spiegano le covarianze tra le variabili (Floyd & Widaman, 1995). Dall'analisi si ottengono i *caricamenti dei fattori*, che altro non sono i pesi di regressione che permettono di prevedere le variabili misurate dalle variabili latenti, e le correlazioni tra le variabili latenti. Inoltre, è possibile stimare la varianza di ciascuna variabile: la varianza comune, o varianza associata alle variabili latenti, stimata in base alla varianza condivisa con altre variabili misurate nell'analisi, e la varianza unica, che è una combinazione di varianza affidabile specifica della variabile misurata e varianza casuale di errore nella variabile (Floyd & Widaman, 1995). Per quanto riguarda le tecniche, l'EFA comporta una serie di passaggi chiave che riguardano la scelta del metodo di estrazione dei dati, determinare il numero di fattori, la rotazione dei dati e la comprensione delle associazioni tra i fattori (Floyd & Widaman, 1995; Furr, 2020). Per la stima e l'estrazione dei dati, esistono diversi metodi psicometrici, come il metodo degli assi principali (*principal axis factoring – PAF* or *least square*) utilizzato quando i dati sono continui e seguono una distribuzione normale; altri metodi disponibili sono la massima verosomiglianza (*maximum likelihood*), minimi quadrati ponderati (*weighted least squares*) e minimi quadrati non ponderati (*unweighted least squares*). In generale, la scelta ottimale dipende dalla natura dei dati (ordinali o continui; distribuzione normale o non normale) (Furr, 2020).

Il secondo step riguarda l'identificazione dei fattori, o dimensioni, e possono essere utilizzati diversi metodi, tra cui quello di esaminare gli autovalori; altri metodi riguardano la valutazione di quanti fattori sono necessari per spiegare la varianza comune. Vengono riportati diverse percentuali di varianza totale spiegata: Strainer (1994) suggerisce che i fattori dovrebbero

spiegare almeno il 50% della varianza totale, mentre Floyd e Widaman (1995) suggeriscono almeno l'80% di varianza comune stimata. Viene raccomandato, inoltre, l'uso degli "scree plot" in cui si tracciano gli autovalori ordinati della soluzione fattoriale e si cerca l'"angolo" per determinare il numero di fattori da mantenere. Tra i metodi statistici ampiamente utilizzati nel determinare il numero di dimensioni vi sono l'analisi parallele (Hayton, Allen & Scarpello,2009), la Minimum Average Pertial Test (MAP; Velicer, 1976) e la Very Simple Structure method (VSS; Revelle & Rocklin, 1979).

La rotazione dei dati ha come scopo di produrre risultati che siano chiari e interpretabili, ovvero che raggiungono la "struttura semplice". Tra i differenti tipi di rotazione è possibile distinguere in rotazioni ortogonali e oblique; la rotazione ortogonale (ad esempio varimax) tenta di non far correlare i fattori, al contrario la rotazione obliqua (tra cui promax, oblim) prova a far correlare i fattori.

Infine, l'ultimo step, l'analisi delle connessioni tra fattori e item, permette di comprendere il significato psicologico dei fattori; il significato dei fattori viene definito dal caricamento dell'item su di esso. Dunque, quando la "struttura semplice" è rilevata, ogni item presenterà una forte associazione (load) su un fattore e ogni fattore presenterà molti item fortemente associati con esso. Gli item che presentano un forte carico su un fattore saranno aggregati per formare un punteggio di scala; quest'ultimo verrà interpretato in termini di dimensione psicologica. Tuttavia, può accadere che l'EFA riveli item con una forte associazione su più fattori, altri con una debole associazione su tutti i fattori, o anche fattori che presentano solo due o tre item. È questo il caso in cui vi è la motivazione a riesaminare il numero di fattori o l'eliminazione degli item che mostrano associazioni deboli (Furr, 2020).

La CFA, al contrario della pura esplorazione, viene utilizzata per confermare un ipotesi a priori, chepuò essere basate su teorie o su risultati di precedenti studi empirici. Nella CFA, dunque, si inizia con una precisa idea della dimensionalità di una misura; il modello viene definito dal numero di fattori, le connessioni tra fattori e item, in termini di quali item sono e non sono associati a ciascun fattore, e l'esistenza di associazioni "non-zero" tra i fattori, nel caso di modelli di multidimensionali. La soluzione fattoriale viene valutata in base alla sua capacità di riprodurre la matrice di covarianza del campione delle variabili misurate (Brown & Moore, 2012; Furr, 2020). Dunque, la CFA , a differenza dell'EFA, richiede una solida base empirica o concettuale per guidare la specificazione e la valutazione del modello fattoriale. Di conseguenza, l'EFA viene spesso utilizzata nelle fasi iniziali del processo di sviluppo della scala e di validazione del costrutto, mentre la CFA viene utilizzata nelle fasi successive, quando la struttura sottostante è stata stabilita su basi empiriche e teoriche precedenti (Brown & Moore,

2012). L'analisi dei fattori può rilevare i costrutti attesi, insieme ad altri fattori inaspettati; se possono essere interpretati in maniera significativa, la loro identificazione può contribuire alla comprensione del dominio di ricerca e all'interpretazione dei punteggi della misura (Floyd & Widaman, 1995). Tra i diversi scopi per cui la CFA viene utilizzata vi è la valutazione della validità di costrutto degli strumenti, la quale è sostenuta se la struttura fattoriale della scala è coerente con i costrutti che lo strumento intende misurare; la validità di costrutto è compromessa nel momento in cui l'analisi fattoriale non individua i costrutti sottostanti che spiegano una varianza sufficietne delle variabili considerate, o se i costrutti che vengono individuati non sono coerenti con le aspettative (Floyd & Widaman, 1995).

A seguito della specificazione del modello, nella CFA è possibile stimare i parametri di adattamento del modello, tra questi ricordiamo gli indici di fit e la stima dei parametri, come il caricamento dei fattori (factor loadings) e le correlazioni interfattoriali. Gli indici di fit riflettono il grado complessivo di coerenza delle risposte effettive alla misura con il modello specificato e, in generale, una prima preoccupazione è che i parametri rientrino in intervalli accettabili. L'adeguatezza generale del modello viene valutata tramite indici di adattamento "assoluti", che indicando la bontà del modello di per sé; l'indice più comunemente usato è il test Chi quadro (o Chi square), che valuta la significatività delle covarianze non spiegate, ovvero le covarianze tra le variabili misurate che non sono spiegate dal modello (Floyd & Widaman, 1995; Furr, 2020). In generale se la statistica del chi-quadro non è significativa, il modello fornisce una rappresentazione adeguata dei dati; tuttavia, tale indice risulta essere sensibile alla dimensione campionaria e da modelli che contengono molte variabili, di conseguenza anche quantità molto piccole di covarianza residua possono essere significative. Un approccio alternativo è l'uso del test chi-quadro close fit proposto da Browne e Cudeck (1993), sostenendo che non ci si deve aspettare che i modelli si adattino perfettamente ai dati, ma piuttosto che forniscano un adattamento relativamente vicino ai dati (Floyd & Widamn, 1995).

Altri indici assoluti di adattamento includono il rapporto chi-quadrato/degrees of freedom e l'indice di bontà dell'adattamento (GFI) (Joreskog & Sorbom, 1989). Tra gli indici di cui ne viene raccomandato l'uso (Marsh, Balla e McDonald, 1988) vi è l'indice di Tucker-Lewis (TLI), indice di adattamento relativo indipendente dalla dimensione del campione, il Comparative Fit Index (CFI), la Standardized Root Mean Square Residual (SRMR), e la Root Mean Square Error of Approximation (RMSEA). Se considerati insieme, questi indici forniscono una valutazione più conservativa e affidabile dell'adattamento del modello e dagli studi di simulazione sui criteri di cut off sono state fornite delle "linee guida" per considerare

accettabile l'adattamento dei dati al modello: valori di SRMR prossimi a .08 o inferiori; valori di RMSEA prossimi a .06 o inferiori; valori di CFI e TLI prossimi a .95 o superiori (Brown & Moore, 2012; Hu & Bentler, 1999). Applicazioni più avanzate della CFA permettono una serie di modellazione, come analizzare l'invarianza della misura, nel tempo o tra sottogruppo di popolazione, o l'analisi di fattori di ordine superiore, come l'analisi fattoriale di second'ordine (Brown, 2006; Brown & Moore, 2012).

2.4.3 Exploratory Graph Analysis (EGA)

Sebbene l'analisi fattoriale rappresenta uno dei metodi statistici più utilizzati nelle scienze psicologiche e comportamentali, di recente sono state proposte metodiche analitiche innovative. Le diverse tecniche che sono state sviluppate dagli anno '60 per stimare il numero di dimensioni sottostanti nei dati psicologici, hanno mostrato negli studi di simulazioni i limiti di ognuna delle tecniche (Garrido et al., 2013; Lubbe, 2019), inoltre altre sfide che vanno oltre la stima del numero delle dimensioni, vi sono la rotazione dei dati e l'interpretazione dei carichi fattoriali (Sass & Schmitt, 2010).

Partendo dal lavoro di Van Der Maas e colleghi (2006), la letteratura recente sta proponendo la modellazione di dati psicologici attraverso modelli di rete, *network*; l'assunto di base è che un insieme di dati che corrisponde a un modello generale di fattori può essere simulato utilizzando un modello di rete connesso (Golino & Epskamp, 2017). Nello specifico, nel campo della psicologia sta rivestendo un ruolo particolare la stima di *network non diretti*, i cosiddetti Markov Random Field (Lauritzen, 1996). Nel framework dei *network*, le variabili causali sono rappresentati da "nodi" che sono collegate da bordi o link che indicano il livello di interazione tra queste variabili. Questi modelli si concentrano sulla stima delle relazioni dirette tra le variabili osservate. Tali modelli si sono dimostrati molto promettenti in diversi campi psicologici, come la psicopatologia, la formazione degli atteggiamenti, la ricerca sulla qualità della vita e la psicologia dello sviluppo (Golino & Epskamp, 2017). Per rispondere alle esigenze nello studio e la stima di un *network* su dati psicologici sono nati i *network psicometrici* (Epskamp et al., 2015). I *network* vengono utilizzati per stimare la relazione tra più variabili, fornita dal *modello grafico Gaussiano* (GGM) (Lauritzen, 1996), che stima la distribuzione congiunta delle variabili casuali (cioè i nodi della rete) modellando l'inverso della matrice varianza-covarianza (Epskamp et al., 2017). I nodi (ad esempio, gli item del test) sono collegati da "bordi" (*edges*), che indicano la forza dell'associazione tra le variabili (Epskamp & Fried, 2018). Le *edges* rappresentanto tipicamente coefficienti di correlazione parziale (Epskamp & Fried, 2018), indicando, dunque, l'associazione che rimane tra due variabili dopo il controllo

di tutte le altre variabili; le *edges* assenti, invece, rappresentano correlazioni parziali nulle, indicando variabili condizionatamente indipendenti (Epskamp & Fried, 2018; Epskamp et al., 2017). Inoltre, quando una sezione di rete mostra molti nodi connessi tra loro viene detta *cluster*. I *cluster* sono di particolare interesse per la psicometria, poiché rappresenterebbero un modello a variabili latenti, o, a seconda delle ipotesi sulla struttura causale sottostante, l'influenza dovuta alle variabili latenti si manifesterà nelle strutture di rete come tali cluster cui tutti i nodi interagiscono tra loro (Golino & Epskamp, 2017). Applicando tale prospettiva alla letteratura della psicopatologia, si può sostenere che i *cluster* di nodi rappresentano i sintomi di una condizione, o ancora, nello studio sui tratti di personalità, questi possono essere presentate come cluster in un *network* stimato.

Recentemente, Golino e Epskamp (2017) hanno proposto un approccio innovativo per identificare le dimensioni dei costrutti psicologici dalla prospettiva della psicometria di rete, definito *Exploratory Graph Analysis* (Christensen et al., 2019; Golino & Demetriou, 2017; Golino & Epskamp, 2017; Golino, Moulder, et al., 2020; Golino, Shi, et al., 2020).

L'EGA combina il modello GGM con un algoritmo di clustering per reti peonderate (Walktrap; Pons & Latapy, 2006) per valutare la dimensionalità degli item nei costrutti psicologici; tale combinazione ha presentato un'accuratezza superiore rispetto ai metodi tradizionali basati sull'analsi dei fattori, nel recuperare il numero di dimensioni (o fattori) simulati. Viene inoltre utilizzato l'approccio EBICglasso (Epskamp & Fried, 2018), il quale minimizza la funzione di log-likelihood penalizzata e seleziona il modello migliore utilizzando il criterio di informazione bayesiano esteso (EBIC) (Chen & Chen, 2008). In un recente lavoro di Golino e colleghi (2020), è stata confrontata l'EGA con diversi tipo di metodi di analisi dei fattori, come ad esempio l'analisi parallela, evidenziando che l'utilizzo del modello di rete GGM ottiene un'accuratezza complessiva dell'87,91% nella stima del numero delle dimensioni simulate.

Il framework EGA presenta diversi vantaggi rispetto ai metodi più tradizionali (Golino, Shi, et al., 2020). A differenza dei metodi dell'EFA, l'EGA non richiede un metodo di rotazione per interpretare i fattori di primo ordine stimati. Sebbene la scelta del metodo di rotazione sia raramente discussa nella letteratura sulla validazione, ha conseguenze significative per la validazione, come ad esempio, per la stima dei caricamenti dei fattori (Sass & Schmitt, 2010). L'EGA, inoltre, indentifica automaticamente gli item nei fattori, o dimensioni, al contrario nell'EFA il ricercatore deve analizzare la matrice di caricamento dei fattori. Infine, un altro vantaggio è la rappresentazione a rete, che permette di illustrare il modo in cui gli item si relazionano all'interno e tra le dimensioni.

Dall'introduzione dei *network psicometrici*, nel framework dell'EGA sono stati utilizzati altri approcci di stima, come il Triangulated Maximally Filtered Graph (TMFG) (Christensen et al., 2018; Massara et al., 2016) o la Dynamic Exploratory Graph Analysis (DynEGA), per trattare i dati longitudinali intensivi, combinando le tecniche dei sistemi dinamici con la psicometria di rete (Golino et al., 2022).

Per stimare la stabilità dei *network psicometrici* e, appunto, il modello stimato mediante l'EGA, Christensen e Golino (2021a) hanno sviluppato un approccio bootstrap, parametrico e non parametrico (resampling) per indagare la stabilità degli item e delle dimensioni stimate dall'EGA. A tal proposito è possibile ottenere una stima sulla *stabilità dell'item*, ovvero stimare la frequenza con cui un item si replica nella dimensione EGA designata, e della *coerenza strutturale*, la stima della frequenza con cui una dimensione viene esattamente replicata. La procedura parametrica inizia con la stima della matrice di correlazione empirica, utilizzata poi come matrice di covarianza per generare dati da una distribuzione normale multivariata. La procedura di *resampling* viene attuata ricampionando con una sostituzione dal set di dati empirici.

Per quanto riguarda la stabilità dell'item ha un range che va da 0 (perfettamente instabile) e 1 (perfettamente stabile), tuttavia, Christensen e Golino (2021) suggeriscono di considerate un item con stabilità accettabile e coerente con la dimensione con una soglia ≥ 0.70. Di conseguenza tale approccio permette di determinare gli item che funzionano male. Per quanto riguarda valutare l'adattamento del modello, l'EGA non utilizza gli indici di fit tradizionali delle analifi fattoriali; in alternativa è stato sviluppato il Total Entropy Fit Index (TEFI) (Golino, Moulder, et al., 2020), indice che si basa sull'entropia di Von Neumann (1927), sviluppata per quantificare la quantità di disordine in un sistema e "intreccio" tra due sottosistemi (Preskill, 2018). L'indice TEFI è una misura relativa di adattamento che può essere utilizzata per confrontare due o più strutture dimensionali. La struttura dimensionale con il valore TEFI più basso indica un migliore adattamento ai dati (Golino et al., 2020).

Un'altra metrica che è stata sviluppata di recente è stata denominata *network loadings*, o carichi di rete (λEGA; Christensen & Golino, 2021c), calcolata standardizzando la forza dei nodi (la somma delle edges a cui un nodo è connesso) suddivisa tra le dimensioni identificate dall'EGA. Similmente al caricamento dei fattori nell'analisi fattoriale, sono stati proposti dei valori soglia per l'interpretazione dell'entità dei carichi di rete: piccolo $\lambda EGA > 0,15$; moderato: $\lambda EGA > 0,25$; grande: $\lambda EGA > 0,35$. L'utilizzo dei carichi di rete è possibile quantificare il contributo dei singoli nodi nel far emergere le diverse dimensioni; inoltre tale indice ha permesso lo studio sulla valutazione dell'invarianza metrica (Jamison et al., 2022). Stabilire l'invarianza in un

modello psicometrico è necessario per garantire che l'adattamento del modello sia equivalente in tutti i sottogruppi contenuti nel campione.
L'utilizzo dell'approccio EGA nel framework dei network psicometrici sia diventato uno dei principali metodi innovativi utilizzati anche nella costruzione e revisione di strumenti per la valutazione dei costrutti psicologici e psicopatologici, tuttavia come ogni campo in evoluzione sono necessari ulteriori studi di simulazione, al pari delle tecniche di analisi tradizionali.

2.5 L'Accuratezza Diagnostica

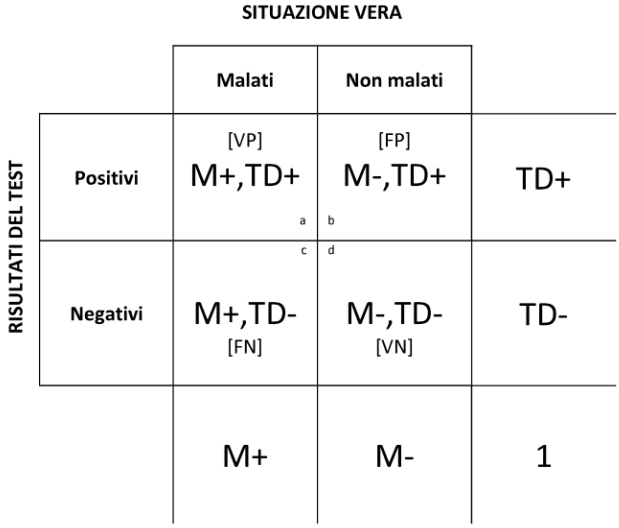
Fin qui sono state analizzate le proprietà fondamentali che uno strumento o una misura deve possedere e le metodiche analitiche con cui esaminarle. Questo con il fine di introdurre nella pratica clinica strumenti sempre più accessibili, affidabili e validi, soprattutto quando si parla di processo diagnostico. Alla base del processo diagnostico vi è l'acquisizione di informazioni, quali la misurazione di sintomi, risposte a stimoli o a domande che il clinico pone mediante l'utilizzo di test diagnostici. È proprio sulla base dei risultati ottenuti da uno o più test diagnostici che il clinico può decidere se sia necessario proseguire la raccolta di informazioni o se sia possibile attribuire un'eventuale categoria diagnostica alla persona. Tuttavia, il processo diagnostico non è privo di errori; molto spesso conduce ad una probabilità diagnostica piuttosto che ad una certezza di aver individuato o escluso una condizione clinica (Valenti et al., 2021). Valutare l'accuratezza diagnostica di procedure e test deve rispondere ad alcune delle domande principali, come ad esempio se il test verrà utilizzato in contesto clinico o di screening e se sia in grado di discriminare tra condizioni cliniche (Eusebi, 2013). Le misure di accuratezza diagnostica sono molto sensibili alle caratteristiche della popolazione in cui viene valutata l'accuratezza del test. Alcune di esse dipendono in larga misura dalla prevalenza della malattia, mentre altre sono molto sensibili allo spettro della malattia nella popolazione studiata (Simundic, 2009). Tra le misure utilizzate per valutare dell'accuratezza di uno strumento diagnostico, vi sono misure che valutano le proprietà discriminative e altre il valore predittivo. Quando si parla di proprietà discriminativa, si intende la capacità dello strumento di discriminare persone con e senza una condizione clinica; quindi il test avrà una validità clinica se è in grado di condurre il clinico ad una diagnosi positiva nelle persone affette da una determinata condizione e ad una diagnosi negativa nelle persone non affette (Valenti et al., 2021). In generale, quando si somministra un test, abbiamo un valore di cut-off che indica se un individuo può essere classificato come *positivo* (sopra/sotto il cutoff) *o negativo* (sotto/sopra il cutoff), e un gold standard (o metodo di riferimento) che ci dirà se lo stesso individuo è malato

o sano. I possibili risultati che si possono ottenere dividono la popolazione di soggetti esaminati in quattro sottogruppi:

- veri positivi (VP) - soggetti con una condizione clinica che risultano positivi al test;
- falsi positivi (FP) - soggetti che non hanno una condizione clinica ma che risultano positivi al test;
- vero negativo (TN) - soggetti che non hanno una condizione clinica e risultano negativi al test;
- falso negativo (FN) - soggetti con una condizione clinica ma che risultano negativi al test;

Questi sottogruppi possono essere visualizzati in una tabella 2x2 (Figura 1).

Figura 1. Tabella 2x2 che riporta la classificazione dei soggetti malati e non malati a cui è stato somministrato il test (Valenti et al., 2021).



La *sensibilità di un test*, misura la sua capacità di identificare correttamente i soggetti affetti da una determinata condizione clinica, ed è espressa in percentuale come la proporzione di veri positivi rispetto al totale dei soggetti con la malattia (VP/(VP+FN) x 100). La capacità del test, al contrario, di fornire risultati negativi in soggetti "non malati" è definita come *specificità del test* e definisce la proporzione dei veri negativi in un gruppo totale di soggetti senza una condizione clinica (VN/(VN+FP) x 100). Entrambi i parametri tendono al 100% quando la frequenza assoluta dei falsi negativi e dei falsi positivi si avvicina a 0; dunque, un test caratterizzato da un'elevata sensibilità, riduce al minimo i falsi negativi, garantendo un'alta

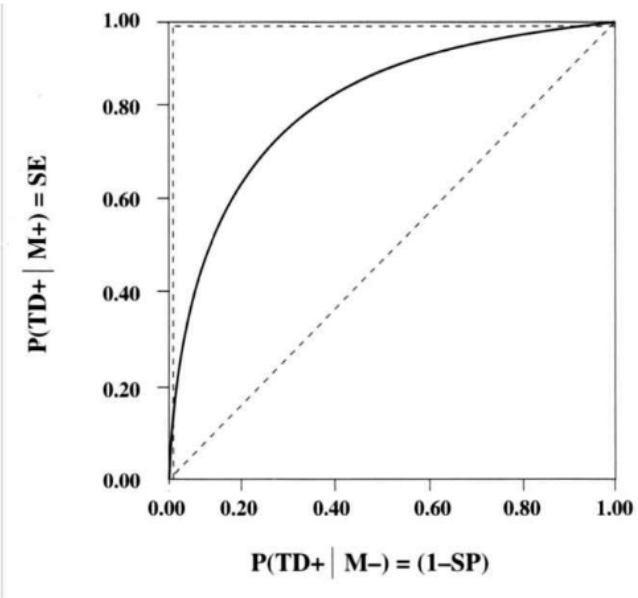
probabilità di rilevare correttamente i soggetti malati. Analogamente, un test con alta specificità minimizza i falsi positivi, riconoscendo accuratamente i soggetti sani e rilevando la condizione clinica solo in una minima parte di questi ultimi (Valenti et al., 2021). Sensibilità e specificità, in genere, vengono stimate su un campione costituito da un piccolo numero di soggetti e potrebbero risultare diverse dalla sensibilità e specificità "vere". La precisione delle stime può essere valutata, quindi, calcolando un intervallo di confidenza del 95% (CI), ossia l'intervallo di valori intorno al valore di sensibilità, o specificità, relativo alla popolazione (Valenti et al., 2021). Nella pratica clinica reale, tuttavia, non esiste uno strumento sensibile e specifico al 100%, di conseguenza sarà necessario valutare tali caratteristiche a seconda del contesto in cui viene utilizzato. Useremo un test sensibile nel caso si stia esaminando una condizione clinica grave, ma trattabile e il risultato di falsi positivi non comportano danni in termini psicologici ed economici al paziente. Al contrario, un test specifico verrà preferito in cui la condizione clinica è grave ma non trattabile ei risultati di falsi positivi comportano danni al paziente (Valenti et al., 2021).

La sensibilità e la specificità sono buoni indici della capacità discriminativa di un test; tuttavia, nella pratica clinica una comune linea di ragionamento è quella di sapere quanto un test sia predittivo di una condizione clinica (Eusebi, 2013). L'interesse dunque è rivolto a conoscere la probabilità che un soggetto "positivo" o "negativo" ad un test sia realmente affetto o meno dalla condizione clinica esaminata. Si parla di *valore predittivo positivo* (VPP) quando vi è la probabilità che un soggetto positivo al test sia realmente affetto (VP/(VP+FP) x 100); al contrario il *valore predittivo negativo* (VPN) definisce la probabilità che un soggetto negativo al test non sia malato, dunque, rappresenta la proporzione di soggetti senza malattia con un risultato negativo del test sul totale dei soggetti con risultati negativi (VN/(VN+FN) x 100). A differenza della sensibilità e specificità, il VPP e il VPN dipendono dalla prevalenza della malattia nella popolazione esaminata: il VPP aumenta mentre il VPN diminuisce con l'aumento della prevalenza della malattia in una popolazione (Eusebi, 2013; Simundic, 2009; Valenti et al., 2021).

Fino adesso l'accuratezza diagnostica è stata esaminata quando un test fornisce risultati dicotomici, positivo/negativo; tuttavia, nella pratica clinica, molti risultati di strumenti testistici vengono esaminati lungo scale continue, ad esempio un test psicologico fornisce un punteggio complessivo del costrutto esaminato. In questo caso la sensibilità e la specificità di un test dipendono dal valore soglia identificato per discriminare in "positivi" e "negativi". Considerando che non esiste un test sensibile e specifico al 100%, qualunque valore soglia (o cut-off) stabilito per discriminare tra positivi e negativi comporterà una certa porzione di

diagnosi errate, ovvero una certa quota di soggetti "malati" otterrà risultati negativi (FN) e una quota di soggetti "non malati" risulterà positiva (FP). Tali quote varieranno in funzione del cut-off considerato, in modo inversamente proporzionale l'una all'altra, modificando di conseguenza i valori di sensibilità e specificità (Valenti et al., 2021). La relazione tra sensibilità e specificità viene rappresentata attraverso la *curva ROC* (*Receiver Operating Characteristic curve*), consentendo di analizzare le caratteristiche di uno strumento nel caso non sia possibile, o non abbia un significato clinico, esprimere i risultati in forma dicotomica. Come mostrato nella Figura 2, nella curva ROC, sull'asse delle ascisse vengono riportate le probabilità che il test risulti positivo in un soggetto non malato (1- specificità)e, sull'asse delle ordinate, le probabilità che il test risulti positivo in un soggetto malato (la sensibilità).

Figura 2. Rappresentazione della curva ROC per un ipotetico strumento diagnostico



La diagonale tratteggiata (Figura 2) rappresenta la curva ROC di un test senza alcun valore clinico, in quanto non consente di discriminare tra soggetti "malati" e non; i soggetti "non malati" avranno la stessa probabilità di risultare positivi al test (FP) dei soggetti effettivamente "malati" (VP) per qualsiasi valore soglia. Al contrario, la linea tratteggiata ad angolo retto rappresenta la curva ROC di un test in grado di discriminare perfettamente tra "malati" e "non malati"; in questo caso il valore soglia corrisponde al vertice dell'angolo retto in alto a sinistra, corrispondente ad una sensibilità e specificità del 100% (Valenti et al., 2021). Nella maggior parte dei casi, la curva ROC di un test si colloca nel triangolo delimitato dalle rette tratteggiate; più la curva si si avvicinerà all'angolo superiore sinistro e più il test sarà valido. L'area che si viene a creare al di sotto della curva ROC (*area under curve, AUC)* rappresenta la misura

globale dell'accuratezza diagnostica e assume valori tra 0 e 1; un test diagnostico con un AUC di 1 avrà un'accuratezza diagnostica perfetta, mentre un test con un' AUC di 0,5 non sarà discriminativo, in quanto la probabilità di ottenere un risultato positivo o negativo sarà uguale al "lancio della moneta" (Eusebi, 2013).

Dunque, uno degli aspetti da considerare è che una metodologia diagnostica non potrà avere contemporaneamente requisiti di sensibilità e specificità; è fondamentale conoscere le caratteristiche di un test non solo per la corretta interpretazione dei risultati ma anche per la selezione di un test diagnostico che abbia requisiti di sensibilità e specificità adatti per la condizione clinica di interesse. Ad esempio, se si selezionano criteri più restrittivi, si otterrà un aumento della sensibilità a scapito della specificità; tale relazione può essere utile nella pratica clinica per adattare le caratteristiche di un testa a particolari condizioni cliniche (Valenti et al., 2021).

Parte II: Contributi Sperimentali[2]

[2] Il contenuto di questa sezione fa riferimento a studi attualmente pubblicati, sottomessi o in revisione:

- Le Donne, I., Valenti, M., Attanasio, M., Scurti, V., Masedu, F., Mazza, M. (2024). Early diagnosis in Autism Spectrum Disorder: evaluation of the predictive agreement between Toddler Module and Module 1 of the Autism Diagnostic Observation Schedule, Second Edition. (*Submitted in Psychological Assessment*)
- Le Donne, I. and Mazza, M., Attanasio, M., Covone, N., Greco, MP., Scurti, V., Valenti, M. (2024). Application of psychometric methods in dimensional analysis and integration of assessment tools in early diagnosis for Autism Spectrum Disorder (*Under review in Journal of Clinal Psychology).*
- Le Donne, I., Attanasio, M., Bologna, A., Vagnetti, R., Masedu, F., Valenti, M., & Mazza, M. (2023). Autism and intention attribution test: a non-verbal evaluation with comic strips. *Annals of General Psychiatry*, *22*(1), 29.
- Mazza, M., Le Donne, I., Vagnetti, R., Attanasio, M., Paola Greco, M., Chiara Pino, M., & Valenti, M. (2024). Normative values and diagnostic optimisation of three social cognition measures for autism and schizophrenia diagnosis in a healthy adolescent and adult sample. *Quarterly Journal of Experimental Psychology,* 77(3), 511-529.

Introduzione agli studi

La valutazione e la diagnosi di autismo pone particolari sfide per gli operatori sanitari, in quanto rappresenta una condizione eterogenea, con una variabilità nella manifestazione sintomatologica e nei livelli di gravità e, come per altre condizioni del neurosviluppo e disturbi psichiatrici, non esistono biomarcatori utilizzabili nella pratica clinica (Hayes et al., 2018). Attualmente la "best practice" nella valutazione dell'autismo è rappresentata da team multidisciplinari e il supporto di strumenti testistici nella formulazione e caratterizzazione della condizione. Negli ultimi anni la ricerca si sta incentrando sull'identificazione di metodi di assessment e di valutazione affidabili e validi con l'obiettivo di aumentare la conoscenza e la caratterizzazione sintomatologica dell'ASD. Sebbene vi sia stata un rapido progresso nel campo della valutazione con l'introduzione di sempre più misure, vi è stata anche l'introduzione di problemi metodologici, tra cui una limitata qualità psicometrica, della coerenza del metodo, della replicazione degli studi e dell'adattamento delle misure a livello transculturale (Harrison et al., 2017). Nello sviluppo di scale affidabili è necessario partire da una concettualizzazione chiara del costrutto e l'identificazione di item (o indicatori) rappresentativi dello stesso, la valutazione delle caratteristiche fondamentali di attendibilità e validità della scala, nonchè fornire dati normativi (Cicchetti, 1994; Clark e Watson, 1995; Gerbing & Anderson, 1988; Harrison et al., 2017). Dalla letteratura esistente si evidenzia come molti degli studi che esaminano ed indagano le caratteristiche dell'ASD mancherebbero di aderire agli standard psicometrici. Una revisione della letteratura (Harrison et al., 2017) ha sollevato alcune preoccupazioni riguardo la qualità psicometrica delle misure utilizzate negli studi esaminati. In particolare, è stato evidenziato come solo il 7% degli studi esaminati includeva informazioni sull'attendibilità e validità; il 45% degli studi non riportavano alcuna valutazione psicometrica delle misure. Inoltre, la lunghezza e complessità della somministrazione rappresentano fattori cruciali nella scelta degli strumenti da impiegare nei setting clinici (Pino et al., 2018). Questo aspetto è particolarmente rilevante nella valutazione diagnostica o di monitoraggio clinico dell'autismo, dove spesso si ricorre all'utilizzo di diversi test diagnostici, adattivi, funzionali e cognitivi che richiedono tempo ed energie all'individuo. Tali strumenti, tuttavia, possono risultare impegnativi sia in termini di tempo che di energie richieste all'individuo. Per questo motivo, nei contesti clinici emerge l'esigenza di adottare strumenti che siano non solo validi e affidabili, ma anche di facile somministrazione, accessibili e meno onerosi dal punto di vista temporale, per garantire un'esperienza più sostenibile sia per i soggetti valutati sia per i professionisti coinvolti. Lo sviluppo e la disponibilità di

strumenti validi e affidabili, hanno un impatto sull'individuo, sull'individuazione di punti di forza e di debolezza, in quanto permettono di garantire una valutazione corretta delle caratteristiche indagate, ma anche nella strutturazione e valutazione di interventi mirati. Uno degli obiettivi che sta diventando centrale tra i ricercatori nel framework dell'ASD è quello di evidenziare l'importanza dell'utilizzo degli approcci interdisciplinari che incorporano metodi psicometrici e analitici accanto alla pratica clinica e agli strumenti diagnostici; tale obiettivo permetterebbe sia la comprensione ma anche la caratterizzazione e generalizzazione della sintomatologia autistica, con implicazioni rilevanti nella strutturazione di percorsi diagnostici-riabilitativi individualizzati e focalizzati, con l'obiettivo del miglioramento della prognosi e della qualità della vita della persona con ASD. Tali approcci svolgono un ruolo fondamentale nella diagnosi e nell'identificazione precoce dei bambini a rischio.

La formulazione di una diagnosi precoce, accurata e stabile è fondamentale per migliorare le prospettive di intervento e di sostegno per i bambini con ASD consentendo l'attuazione tempestiva di strategie educative e terapeutiche individualizzate. Generalmente, l'ASD viene diagnosticato tra i 2 e i 3 anni di età, anche se i sintomi possono essere evidenti già tra i 12 e i 18 mesi (Barbaro & Dissanayake, 2017). Tuttavia, nonostante la maggiore consapevolezza dei segni precoci dell'autismo, solo il 42% dei bambini con ASD viene sottoposto a una valutazione iniziale prima dei 3 anni (Baio, 2018; Benedetto et al., 2021). Questa discrepanza può derivare dalla variabilità della manifestazione dei sintomi. In letteratura, l'insorgenza dei sintomi comportamentali dell'ASD è concettualizzata attraverso tre modelli: il modello a esordio precoce, in cui i bambini mostrano ritardi e anomalie nello sviluppo sociale e comunicativo già nelle prime fasi della vita; il modello regressivo, in cui i bambini inizialmente sviluppano le abilità come previsto ma poi subiscono un declino significativo; e il modello di stagnazione (o plateau) dello sviluppo (Ozonoff & Iosif, 2019; Shumway et al., 2011). Quest'ultimo modello è caratterizzato da abilità iniziali intatte, con il bambino che successivamente non riesce a compiere i progressi previsti (Ozonoff & Iosif, 2019). Oltre alla variabilità nell'insorgenza dei sintomi, i ritardi nella diagnosi possono anche essere correlati alle preoccupazioni relative alla stabilità diagnostica nei bambini identificati prima dei 3 anni, perché molti clinici sono riluttanti a stabilire una diagnosi a questa età precoce (Ozonoff et al., 2015). Tuttavia, diversi studi che indicano che la maggior parte dei bambini e dei ragazzi con diagnosi di ASD tra i 18 e i 48 mesi continua a presentare sintomi nelle valutazioni successive. Una percentuale di bambini con punteggi iniziali di gravità dei sintomi più elevati ha mostrato una riduzione della gravità dei sintomi, passando

a un intervallo di punteggio più lieve entro 1 anno (Benedetto et al., 2021). Di conseguenza, clinici e ricercatori sono interessati a identificare l'età più precoce in cui la diagnosi di ASD è possibile e stabile nel tempo. Attualmente, la diagnosi di ASD si basa su una combinazione di opinioni cliniche di esperti e sull'uso di strumenti diagnostici convalidati; in questo contesto, l'Autism Diagnostic Observation Schedule - Second Edition (ADOS-2) (Lord et al., 2012) è considerato il "gold standard" per la diagnosi di ASD e comprende diversi moduli adattati a diversi gruppi di età e livelli di sviluppo. Per la diagnosi precoce, i moduli ADOS-2 principalmente utilizzati sono il modulo Toddler e il Modulo 1, specificamente progettati per valutare i bambini molto piccoli (il modulo Toddler per l'età di 12-30 mesi e il modulo 1 per l'età ≥31 mesi). Entrambi i moduli valutano due costrutti principali: Affetto Sociale (AS) e i Comportamenti Ristretti e Ripetitivi (CRR). Questi costrutti sono centrali per la diagnosi di ASD e rappresentano i principali domini di deficit nei bambini con autismo. La distinzione tra i due moduli è che il Modulo Toddler stima un range di rischio per una diagnosi di ASD, mentre il Modulo 1 consente una diagnosi definitiva e un livello di gravità (Lord et al., 2012; Luyster et al., 2009). Una diagnosi accurata e certa consente di avviare interventi precoci che possono portare a un miglioramento dei risultati e della qualità della vita, nonché a una riduzione dei costi sociali a lungo termine (Benedetto et al., 2021; Chasson et al., 2007). Diversi studi hanno analizzato la sensibilità e la specificità degli algoritmi diagnostici e delle dimensioni dell'ADOS-2, in particolare per i Moduli 1, 2, 3 e 4 (de Bildt et al., 2011; Hong et al., 2021; Kamp-Becker et al., 2013; Medda et al., 2019). Per il Modulo Toddler, invece, gli studi sul suo utilizzo nella pratica clinica rimangono limitati (Hong et al., 2021; Lee et al., 2019). Inoltre, nel contesto Italiano non esistono studi che indagano l'accordo tra il Modulo Toddler e gli altri moduli.

In questo quadro, l'applicazione dei metodi psicometrici nella pratica clinica è essenziale, dove l'ottimizzazione della raccolta delle informazioni e la riduzione degli errori attraverso l'eliminazione degli item non informativi sono fondamentali per mantenere l'accuratezza (Donadello et al., 2017). Un tentativo nell'utilizzo di questo approccio è stato esplorato applicando la Valutazione Psicologica Formale (FPA) a tutti i moduli dell'ADOS-2 (Pino et al., 2018), studiando le relazioni tra gli item e le osservazioni cliniche per migliorare la qualità dei dati e consentire una valutazione adattiva per le persone con ASD. Ciò ha una particolare rilevanza nella pratica clinica, soprattutto quando si valutano bambini di circa 12 mesi che presentano un sospetto di ASD e sono sottoposti a una valutazione impegnativa che richiede una collaborazione prolungata. Questo approccio supporterebbe la pratica clinica facilitando la raccolta di informazioni chiave, consentendo ai clinici di concentrarsi

sulle caratteristiche essenziali per identificare i bambini a rischio e rendendo una valutazione più adattabile alle caratteristiche del bambino, come verrà esposto negli Studi 1 e 2.

Nella valutazione dell'autismo, oltre che parlare di strumenti diagnostici è altrettanto importante parlare dell'utilizzo di misure valide ed efficaci che possono contribuire all'identificazione dei deficit *core* della condizione, rappresentato, in particolare, dalla SC. Sebbene il costrutto della SC sia stato ampiamente studiato nell'ambito della ricerca sull'ASD, la sua valutazione in ambito clinico nelle persone con ASD è ancora limitata. Speso gli strumenti disponibili non possiedono proprietà psicometriche ben stabilite o, in alcuni casi, derivano da adattamenti di test utilizzati in altri contesti clinici. L'uso di strumenti standardizzati è cruciale per garantire la validità e la riproducibilità dei risultati, fornendo misurazioni oggettive dell'entità e della gravità dei deficit. Questi strumenti permettono anche di identificare le capacità residue o i punti di forza, utili per compensare le difficoltà individuali (Henry et al., 2015; Pino et al., 2018). Molti degli strumenti di SC utilizzati nella ricerca sull'autismo sono stati sviluppati per esplorare ipotesi teoriche sui deficit di SC, piuttosto che essere utilizzati come strumenti di valutazione formale comparabili ai test di intelligenza o di personalità (Brewer et al., 2017; Mazza & Valenti, 2019). Di conseguenza la qualità psicometrica e la disponibilità di dati normativi non rappresentano la priorità. Ricorrere alla valutazione della SC è fondamentale non solo in termini di accuratezza della diagnosi ma an- che per la realizzazione di trattamenti individualizzati che prendano in considerazione la specifica fase di sviluppo del soggetto e le abilità da potenziare (Mazza & Valenti, 2019). Se nei bambini è importante comprendere il livello di sviluppo e acquisizione delle abilità di SC per la pianificazione di interventi che mirino al potenziamento e recupero delle competenze sociali, negli adulti con ASD tale valutazione aiuta a comprendere i vari aspetti del funzionamento adattivo dell'individuo e a realizzare di trattamenti abilitativi e riabilitativi (Mazza & Valenti, 2019). Ciò è particolarmente importante poiché, anche in presenza di funzioni cognitive superiori alla norma, i soggetti con ASD possono manifestare gravi difficoltà nella SC e deficit nelle relazioni interpersonali. Questi problemi hanno spesso un impatto significativo sulla qualità della vita e sul benessere generale dell'individuo (Mazza & Valenti, 2019).

Nonostante la forte associazione tra ASD e deficit nelle abilità di SC, tra cui la ToM, la letteratura riporta risultati contrastanti (Baksh al., 2021; Brewer et al., 2017; Cole et al., 2018). Alcuni studi suggeriscono che gli adulti con ASD hanno notevoli difficoltà a dedurre gli stati mentali ed emotivi degli altri (Baron-Cohen et al., 2001; Castelli et al.,2002; Happè,

1994; Mazza et al., 2022). D'altra parte, altri studi riportano prestazioni simili sia negli individui con ASD sia nei controlli sani nei compiti di ToM (Kirkovski et al., 2016; Roeyers et al., 2001). Inoltre, vi è una scarsità di dati disponibili per dimostrare, a livello di gruppo, come gli individui con ASD, in particolare gli adolescenti e gli adulti, si confrontano con i controlli sani con QI corrispondente (Brewer et al., 2017). Negli ultimi anni, la ricerca ha sviluppato diverse misure e strumenti che indagano le diverse sfaccettature della SC, come la capacità di riconoscere le espressioni facciali e le emozioni attraverso i segnali vocali, il contagio emotivo, la ToM, l'identificazione degli inganni e il giudizio morale (Eddy, 2019) come il test Reading the Mind in the Eyes (Baron-Cohen et al., 2001; Olderbak et al., 2015), la Theory of Mind Assessment Scale (Bosco et al., 2016) e l'Edinburgh Social Cognition Test (Baksh et al., 2021), di cui sono state valutate le proprietà psicometriche. Tuttavia, l'uso delle misure della ToM rimane principalmente confinato al campo della ricerca e la loro applicazione nella pratica clinica rappresenta ancora una sfida. Inoltre, la maggior parte delle misure per la valutazione della ToM, come le Strange Stories (Happè, 1994), richiedono un linguaggio espressivo e ricettivo ben sviluppato (che include, per esempio, lunghe descrizioni verbali e istruzioni), o comportano un grande carico di memoria, rendendole difficili da somministrare a persone con ASD che hanno compromesso le capacità verbali e cognitive (Colle et al., 2007; Mazza et al., 2022; Pino et al., 2020). Sono state suggerite diverse strategie per semplificare i compiti ToM e facilitare la comprensione delle istruzioni: in alcuni casi, sono state presentate situazioni simili alla vita quotidiana dei soggetti; in altri, è stata offerta la possibilità di ripetere la storia quando il paziente sembrava non capire; in altri ancora, sono stati forniti ausili visivi come disegni o vignette (*comic-strip*) (Brunet et al., 2003; Drury et al., 1998; Frith & Corcoran, 1996; Pickup & Frith, 2001). Questo aspetto evidenzia una necessità, applicabile al setting clinico e alle popolazioni cliniche che evidenziano determinate compromissioni, ovvero quello di costruire test semplici, adattabili e di facile somministrazione, come verrà mostrato nello Studio 3.

Il ruolo fondamentale dell'interazione sociale per il benessere individuale e il funzionamento adattivo è ormai ben riconosciuto (Eddy, 2019), e ciò ha portato a una maggiore consapevolezza dell'importanza di valutare le abilità di SC non solo nell'ASD ma anche in condizioni psicopatologiche che possono presentare importanti aree di *overlapping* con l'autismo. E' il caso dei disturbi dello spettro schizofrenico (SSD; APA, 2013). Le due condizioni, ASD e SSD, condividono il termine "spettro", riferendosi alle diverse manifestazioni cliniche che caratterizzano entrambe le codizioni. Questa

definizione riflette la complessità della sintomatologia, la quale può variare a seconda del livello di gravità e manifestazione dei sintomi. Questa complessità può rappresentare una sfida per il clinico, in particolare, durante il processo di diagnosi differenziale. Sebbene ASD e SSD siano due condizioni indipendenti, con presentazioni funzionali e comportamentali diverse (DSM-5; APA, 2013), esistono prove a sostegno dell'idea che la compromissione delle componenti della SC e, in particolare, della ToM, sia il meccanismo alla base della disfunzione sociale in entrambi i disturbi (Baron-Cohen et al., 2015; Boada et al., 2020; Frith & Happé, 2005; Mazza et al., 2010, 2022). L'ASD è un disturbo del neurosviluppo in cui il funzionamento delle abilità sociali appare atipico fin dalla prima infanzia, manifestandosi come difficoltà a rappresentare gli stati mentali degli altri (Frith, 1992; Mazza et al., 2022; Pickup & Frith, 2001; Tin et al., 2018). L'SSD, tuttavia, rappresenta una condizione che si manifesta tipicamente nell'adolescenza o nella giovane età adulta ed è caratterizzata da deliri, allucinazioni, pensiero e linguaggio disorganizzati, comportamento motorio gravemente disorganizzato o anormale e sintomi negativi (APA, 2013). Le prospettive neuropsicologiche concettualizzano l'SSD come un disturbo della rappresentazione degli stati mentali la cui compromissione si verificherebbe dopo il primo episodio psicotico (Bliksted et al., 2017; Frith, 1992, 2004; Mazza et al., 2001, 2010, 2022). La compromissione si manifesterebbe principalmente con l'*ipermentalizzazione*, ovvero la tendenza a travisare gli stati mentali degli altri (Frith, 1992; Martinez et al., 2019; Mazza et al., 2022; Pickup & Frith, 2001; Tin et al., 2018). La letteratura fornisce prove contrastanti sull'ipotesi che la ToM sia compromessa in modo simile in entrambi i disturbi (Chung et al., 2014; Deste et al., 2020; Fernandes et al., 2018; Morrison et al., 2019; Pinkham et al., 2020; Sasson et al., 2011; Tin et al., 2018). Uno studio recente (Mazza et al., 2022) sottolinea l'importanza dell'utilizzo di strumenti di valutazione delle abilità di SC nella diagnosi differenziale dei due gruppi clinici, in particolare nell'adolescenza e nella giovane età adulta. Mazza e colleghi (2022) dimostrano che i pazienti con ASD e SSD ottengono risultati peggiori nei compiti ToM rispetto al gruppo a sviluppo tipico, ma nonostante le due condizioni cliniche abbiano punteggi uguali, mostrano un diverso stile di ragionamento per interpretare lo stato mentale degli altri. I dati sembrano fornire un supporto per l'identificazione di alcuni pattern di risposta come caratteristici degli individui SSD durante l'esecuzione dei compiti: essi non comprenderebbero le storie ToM, in particolare in scenari che includono costrutti complessi come il doppio bluff, le bugie bianche e la persuasione. Gli ASD, invece, comprendono la storia ma non identificano l'intenzione corretta del personaggio (Mazza et al., 2022).

Si evidenzia ancora una volta come vi sia la necessità di sviluppare e promuovere l'utilizzo di misure affidabili e valide che possano essere utilizzati in contesti clinici e di ricerca a supporto delle misure "gold-standard".

Sulla base dei presupposti appena presentati abbiamo condotto gli studi che saranno accuratamente descritti di seguito. Gli studi sono stati condotti seguendo un approccio interdisciplinare, che permette di evidenziare l'importanza di integrare la ricerca, l'utilizzo di metodiche statistiche-psicometriche (come ad esempio l'applicazione di network psicometrici e modelli IRT nella costruzione e revisione di misure) e la pratica clinica.

Studio 1

Diagnosi precoce nel disturbo dello spettro autistico: valutazione dell'accordo predittivo tra il Modulo Toddler e il Modulo 1 dell'Autism Diagnostic Observation Schedule, seconda edizione.

1.1 Scopo dello Studio

Sebbene l'intervento precoce sia fondamentale nel miglioramento della prognosi, la diagnosi di ASD viene generalmente stabilita intorno ai 2-3 anni di età, nonostante la comparsa dei sintomi già a 12-18 mesi. Fornire una diagnosi stabile e certa consente di avviare tempestivamente interventi riabilitativi precoci. In questo contesto, l'obiettivo dello studio è valutare l'accordo diagnostico tra due moduli dell'ADOS-2: il Modulo Toddler (per i bambini tra i 12 e i 30 mesi) e il Modulo 1 (per i bambini dai 31 mesi in su). A nostra conoscenza, nessuno studio nel contesto italiano ha esaminato l'accordo diagnostico tra il modulo Toddler e gli altri moduli. Inoltre, lo studio si propone di esaminare l'effetto predittivo dei due costrutti condivisi da entrambi i moduli, ovvero AS e CRR, per capire in che misura ciascun costrutto contribuisca alla diagnosi di ASD. Verificare l'accordo predittivo delle misure faciliterebbe la continuità diagnostica e l'identificazione precoce dei bambini a rischio di ASD. Inoltre, la comprensione di come i due costrutti chiave nella diagnosi di ASD (AS e CRR) influenzino la diagnosi, consente di strutturare interventi tempestivi e mirati alle caratteristiche del disturbo con l'obiettivo di migliorare la prognosi.

1.2 Metodo

Partecipanti

È stato condotto uno studio longitudinale, con 43 bambini reclutati presso il Centro di Riferimento Regionale per l'Autismo, Azienda Sanitaria Locale (ASL 1), Regione Abruzzo, Italia, valutati al loro primo accesso al servizio (T0) utilizzando il Modulo Toddler e dopo il compimento di 31 mesi (T1) utilizzando il Modulo 1, per la conferma diagnostica. La diagnosi è stata effettuata da clinici esperti secondo i criteri del DSM-5 (APA, 2013) e attraverso l'utilizzo dell'ADOS-2 (Lord et al., 2012). In particolare, il Modulo Toddler e il Modulo 1 sono stati utilizzati, rispettivamente, per le valutazioni alla prima consultazione (T0) e nel follow-up (T1). Nel campione, 34 bambini hanno ricevuto la conferma diagnostica di ASD, mentre 9 non l'hanno ricevuta. Le informazioni descrittive del campione sono riportate nella Tabella 1.

Tabella 1. Valori descrittivi del campione e risultati delle differenze tra i punteggi di T0 e T1 sull'ADOS-2

	T0	**T1**
N	43	43
Età media in mesi (SD)	25.27(3.6)	44.16(10.3)
Genere (M; F)	33;10	33;10

Autism Diagnostic Observation Schedule-Version 2

	Toddler	**Modulo 1**	**T-test**	**p-value**
AS	14.72(4.4)	10.3 (4.9)	$T_{(42)}$= 6.3	**< 0.001**
CRR	3.53(2.08)	3.67(2.02)	$T_{(42)}$= -0.39	0.693
Punteggio Totale	18.25(5.8)	14.07(6.3)	$T_{(42)}$= 4.96	**<0.001**
CS AS	7.55(2.2)	5.04(1.96)	$T_{(42)}$= 7.36	**<0.001**
CS CRR	6.58(2.08)	6.95(2.2)	$T_{(42)}$= -0.88	0.38
CS Totale	7.3(2.3)	5.51(2.2)	$T_{(42)}$= 5.608	**<0.001**

Note: AS= Affetto Sociale; CRR= Comportamenti Ristretti e Ripetitivi; CS AS= Comparative Score Affetto Sociale; CS CRR= Comparative Scor CRR; CS Totale= Comparative score Totale.

*Le differenze significative sono riportate in grassetto.

Misure

ADOS-2

L'ADOS-2 (Lord et al., 2012) è considerato il *gold standard* per la valutazione e la quantificazione della sintomatologia autistica. Si tratta di uno strumento osservativo semi-strutturato e standardizzato che comprende diversi compiti che elicitano comportamenti rilevanti per una diagnosi di ASD. L'ADOS-2 è composto da cinque moduli che vengono somministrati in base all'età cronologica e al livello di linguaggio espressivo del soggetto; ogni modulo richiede dai 40 ai 60 minuti per essere completato. Oltre al punteggio della sintomatologia totale, l'ADOS-2 fornisce punteggi per due domini: AS e CRR. In questo studio sono stati utilizzati il Modulo Toddler e il Modulo 1. Il Modulo Toddler consente di valutare i sintomi precoci legati all'ASD in bambini di età compresa tra i 12 e i 30 mesi, fornendo un intervallo di rischio da "nessuno a lieve", "da lieve a moderato" e "da moderato a grave". Il Modulo 1, invece, è adatto a bambini di età ≥31 mesi e fornisce punteggi di cut-off per la classificazione di "spettro autistico" e "autismo". Inoltre, la gravità dei sintomi può essere determinata utilizzando i punteggi di confronto disponibili nel Modulo 1.

1.3 Analisi

Per il primo livello di analisi, abbiamo effettuato un'analisi esplorativa dei dati costruendo una matrice di classificazione che incrocia le decisioni cliniche del Modulo Toddler con quelle del Modulo 1. I livelli di gravità forniti dai due moduli sono stati classificati come lieve (0), moderato (1) e grave (2). È stato applicato un test chi-quadro di Pearson alla matrice di classificazione per valutare qualsiasi associazione tra le classificazioni fornite dai due moduli. Da questa matrice, è stata stimata il livello di accordo utilizzando la somma dei valori diagonali e corretto per l'accordo casuale con la *kappa* di Cohen. Successivamente, data la natura ordinale delle variabili, abbiamo esaminato l'associazione statistica tra i rischi di gravità identificati dal Modulo Toddler e quelli identificati dal Modulo1. A tale scopo è stato eseguito un test di tendenza non parametrico basato sul rango per valutare la significatività di qualsiasi tendenza nei dati, in particolare se un aumento di una variabile corrisponde a un aumento o a una diminuzione di un'altra variabile (test Jonckheere-Terpstra) (Jonckheere, 1954; Terpstra, 1952). Infine, abbiamo calcolato due modelli di regressione logistica per esaminare il comportamento diagnostico relativo alla presenza o all'assenza di ASD in base ai due moduli somministrati e alla diagnosi finale disponibile. I modelli utilizzano i due domini, SA e CRR, sia del Modulo Toddler che del Modulo 1 come predittori.

1.4 Risultati

Matrice di classificazione

La matrice di classificazione che incrocia la decisione clinica del Modulo Toddler (livello di rischio) con quella del Modulo 1 (gravità dei sintomi ASD) è riportata nella Tabella 2. I risultati ottenuti dalla matrice di classificazione mostrano associazioni statistiche significative tra il Modulo Toddler e il Modulo 1 (χ^2= 13,69, df= 4, p= 0,008). Questa associazione è focalizzata sulla classificazione della condizione ASD in modo multimodale senza alcun riferimento alla natura ordinale della gravità della sintomatologia. Questo risultato è confermato dalla stima della V di Cramer (V= 0,399). Tuttavia, si sottolinea che l'accordo diagnostico di gravità (K= 0,06) è piuttosto basso. Come mostrato nella Tabella 2, il 67,7% dei bambini è stato classificato in maniera diversa, passando dal livello di gravità più alto a quello moderato.

È emerso che, tenendo conto delle informazioni ordinali fornite dalla variabile gravità, il test di Jonckheere-Terpstra ha mostrato una tendenza statisticamente significativa (T_{JT} = 286,00, z = 3,756, p < 0,0005).

Tabella 2. Matrice di classificazione per indagare l'associazione tra classificazione Toddler e Modulo 1

ASD Livello di Rischio	ASD Severità Sintomatologica			Totale
	0	1	2	
0	4(100%)	0(0%)	0(0%)	4(100%)
1	4(50%)	3(37.5%)	1(12.5%)	8(100%)
2	5(16.13%)	21/67.7%)	5(16.13%)	31(100%)
Totale	13(30%)	24(56%)	6(14%)	43(100%)

Modelli di Regressione Logistica

Per quanto riguarda la predittività dei costrutti AS e CRR dei due moduli esaminati, i risultati ottenuti dal modello logistico mostrano AS del Modulo Toddler come predittore significativo della diagnosi futura (β = 0,30, SE = 0,12, z = 2,45, p = 0,014) ma non CRR (β = 0,027, SE = 0,29, z = 0,09, p = 0,92). L'AS presenterebbe un aumento del 35% nel ricevere la diagnosi (OR = 1,35; 95% CI: 1,06-1,71).

Nel Modulo 1, i CRR sembra essere il predittore significativo per la diagnosi (β = 1,001, SE = 0,41, z = 2,41, p = 0,016) rispetto all'AS (β = - 0,09, SE = 0,13, z = -0,72, p = 0,47). In questo caso, la probabilità di ricevere la diagnosi aumenta in modo tale che l'OR di ASD è 2,72 (95% CI: 1,2- 6,14) per ogni unità di aumento del punteggio CRR. Si veda la Tabella 3.

Tabella 3. Risultati del modello di regressione logistica

			Diagnosi		
	β	**SE**	***p***	**OR**	**95%CI**
Predittori					
ADOS-2 TODDLER					
AS	0.30	0.12	**0.014**	1.35	1.06 – 1.71
CRR	0.027	0.29	0.92	1.01	0.57 – 1.82
ADOS-2 MODULO 1					
AS	-0.09	0.13	0.47	0.90	0.70 – 1.17
CRR	1.001	0.41	**0.016**	2.72	1.2 – 6.14

Note: AS= Affetto Sociale; CRR= Comportamenti Ristretti e Ripetitivi.
*Le differenze significative sono riportate in Grassetto.

1.5 Discussione

Una questione centrale riguardo l'ASD è l'identificazione dell'età minima in cui una diagnosi può essere fatta in modo affidabile e stabile nel tempo. La stabilità di una diagnosi di ASD

consente di strutturare e applicare percorsi di intervento precoci, finalizzati a migliorare i risultati sintomatici. Attualmente, l'ADOS-2 è utilizzato come strumento "gold standard" nella diagnosi di ASD ed è talvolta richiesto per l'accesso agli interventi precoci, anche se poche ricerche ne supportano l'uso clinico (Barbaresi et al., 2022). In particolare, pochi studi hanno esaminato il Modulo Toddler e l'accuratezza clinica (Hong et al., 2021; Lee et al., 2019) e, a nostra conoscenza, nessuno studio ha esplorato l'accordo diagnostico con il Modulo 1, in cui la diagnosi è tipicamente confermata o esclusa.

Nel presente studio abbiamo analizzato il livello di accordo tra la fascia di rischio (lieve, moderata o alta) stimata dal Modulo Toddler alla prima consultazione e la gravità della sintomatologia fornita dal Modulo 1 alla conferma diagnostica. Lo studio ha incluso 43 bambini dai 12 mesi che hanno partecipato al primo consulto clinico per sospetto ASD e sono stati sottoposti al Modulo Toddler per stimare il range di rischio. Successivamente, i bambini sono stati rivalutati con il Modulo 1 per la conferma diagnostica, con il risultato che 34 bambini sono stati diagnosticati con ASD, mentre 9 no. Dai risultati della matrice di classificazione, è stata riscontrata un'associazione significativa tra i due moduli nel classificare i bambini come affetti da ASD; tuttavia, l'accordo sulla gravità diagnostica era molto basso, come indicato dalla *kappa* di Cohen. Come mostrato nella Tabella 2, il 67,7% dei bambini viene riclassificato dal livello di gravità più alto nel Modulo Toddler a un livello di gravità moderata nel Modulo 1. Ciò suggerisce che, sebbene vi sia concordanza nella conferma diagnostica, il Modulo Toddler potrebbe sovrastimare la gravità dei sintomi, mentre il Modulo 1 riflette un profilo sintomatologico meno grave. La sovrastima della gravità non è necessariamente sfavorevole, perché permette di identificare i sintomi che possono essere a rischio di peggioramento, consentendo così un intervento più precoce e intensivo. Questo risultato sottolinea anche la variabilità dello sviluppo nei bambini piccoli, che può portare all'incertezza diagnostica, contribuendo alla riluttanza di molti clinici a stabilire una diagnosi prima dei 3 anni (Ozonoff et al., 2015). Tuttavia, i nostri risultati relativi all'accordo tra i due moduli sulla conferma diagnostica, piuttosto che sulla gravità dei sintomi, sono in linea con l'evidenza che molti bambini e ragazzi con diagnosi di ASD già all'età di 18 mesi continuano a presentare sintomi ai successivi follow-up, anche se spesso con una gravità ridotta dei sintomi a un anno di distanza (Benedetto et al., 2021). L'accordo osservato sulla conferma diagnostica precoce, a partire da circa 12 mesi di età, tra i due moduli potrebbe supportare la pratica clinica consentendo l'inclusione precoce di bambini a rischio in interventi riabilitativi per potenziare le aree di sviluppo e migliorare gli esiti sintomatici.

Come ulteriore analisi, il nostro studio ha esaminato in via preliminare l'influenza dei due costrutti dell'ADOS-2, AS e CRR, sulla conferma diagnostica. Questi costrutti rappresentano le componenti principali per la diagnosi di ASD secondo il DSM-5, con AS che comprende item sull'interazione sociale e la comunicazione reciproca e CRR che indaga la presenza di comportamenti e interessi ristretti e ripetitivi (Lord et al., 2012). I risultati dei due modelli di regressione logistica evidenziano che AS e CRR predicono la diagnosi in modo diverso e in fasi diverse. In particolare, sembra che nel Modulo Toddler la presenza di deficit nelle abilità di AS abbia un ruolo centrale nell'influenzare la conferma della diagnosi a un anno di distanza. Per un aumento unitario del punteggio AS aumenterebbe del 35% , la probabilità di conferma diagnostica. Al contrario, nel Modulo 1, quando è possibile formulare una diagnosi, il dominio CRR diventa più influente. Ogni aumento unitario del punteggio del dominio CRR aumenta la probabilità di una diagnosi di ASD. I Comportamenti stereotipati e motori compaiono anche in altre condizioni del neurosviluppo, in particolare nella disabilità intellettiva. Pertanto, nelle prime fasi dello sviluppo, sembrerebbe essere la compromissione della reciprocità e della reattività sociale, piuttosto che il dominio CRR, ad essere critica. Gli studi che analizzano le traiettorie di sviluppo indicano che i primi segni di comportamenti sociali legati all'ASD, come l'interesse sociale e l'affetto condiviso, sono presenti entro i primi 2 anni di vita (Bussu et al., 2018; Ozonoff & Iosif, 2019; Pino et al., 2024). I nostri risultati suggeriscono anche che i comportamenti ristretti e stereotipati diventano più significativi un anno dopo, forse riflettendo lo sviluppo di ulteriori abilità motorie, sensoriali e cognitive. Tuttavia, mentre i marcatori socio-comunicativi che emergono precocemente possono aiutare a prevedere una successiva diagnosi di ASD, si sa meno del valore predittivo dei comportamenti ristretti e ripetitivi precoci. Questi comportamenti, che possono includere elementi motori, sensoriali, vocali e cognitivi, potrebbero cambiare nel tempo sia per tipologia che per frequenza (Grossi et al., 2021). Sebbene gli aspetti dei comportamenti sociali possano aiutare a distinguere i bambini a rischio, recenti evidenze suggeriscono che è la frequenza, piuttosto che la varietà, dei comportamenti motori stereotipati a influenzare la gravità dell'ASD (Elison et al., 2014; Grossi et al., 2021).

1.6 Conclusioni

Sebbene questi risultati siano interessanti, lo studio presenta dei limiti che dovrebbero essere presi in considerazione per le ricerche future. Un limite primario è la dimensione del campione, in particolare la limitata rappresentanza di bambini classificati come non affetti da ASD. Un'altra limitazione riguarda la distribuzione di genere, che nel nostro studio presenta una prevalenza maschile; tuttavia, questo problema è comune negli studi sull'ASD, poiché si stima

che la condizione sia più frequente nei maschi che nelle femmine (APA, 2013; Valenti et al., 2019). Sono quindi necessari studi futuri per generalizzare i risultati su un campione più ampio. In conclusione, questo studio preliminare dimostra un accordo diagnostico tra il Modulo Toddler e il Modulo 1, supportando l'evidenza che una diagnosi stabile può essere possibile prima dei 2 anni di età. La conferma di una diagnosi stabile in questa fase precoce consentirebbe di avviare e strutturare interventi tempestivi e precoci. Inoltre, capire quali domini influenzano in modo più significativo la presenza di ASD e la gravità dei sintomi ha importanti implicazioni per la pianificazione degli interventi. Questa conoscenza consente di elaborare strategie di trattamento mirate a specifiche abilità compromesse, rendendo gli interventi più adattivi e tenendo conto della variabilità dello sviluppo in questo gruppo di età così delicato.

Studio 2

Applicazione dei metodi psicometrici nell'analisi dimensionale e nell'integrazione degli strumenti di valutazione nella diagnosi precoce del disturbo dello spettro autistico.

2.1 Scopo del lavoro

L'importanza della diagnosi precoce ha portato la ricerca nel campo dell'ASD allo sviluppo e alla revisione di strumenti di screening e diagnostici per l'autismo. Il presente studio si propone, come obiettivo preliminare, mediante l'utilizzo dell'Exploratory Graph Analysis (EGA), di replicare e indagare la struttura degli item del Modulo Toddler dell'ADOS-2 in una popolazione italiana di bambini di età compresa tra i 12 e i 30 mesi, con l'obiettivo di identificare fattori latenti ed eventuali pattern o cluster di item correlati. Analizzando i modelli di connessione tra gli item, possiamo valutare la consistenza interna del Modulo per bambini piccoli e identificare gli item che potrebbero non essere in linea con la struttura generale del test. Da questo presupposto abbiamo esaminato i compiti strutturati del Modulo Toddler che permettono la codifica degli item selezionati per l'algoritmo diagnostico, indagando le dimensioni primarie che spiegano la sintomatologia ASD nei bambini molto piccoli. Questo approccio ci permette di differenziare gli item essenziali da quelli ridondanti, facilitando un processo di valutazione adattivo ed efficiente in termini di tempo per i bambini molto piccoli, nella raccolta delle informazioni chiave nella pratica clinica.

2.2 Metodo

Partecipanti

Il campione dello studio è costituito da 91 bambini con un'età compresa tra i 12 e i 31 mesi (età media in mesi = 24,94 ± 4,25) alla prima consultazione clinica, arrivati per sospetto ASD al Centro Regionale di Riferimento per l'Autismo, ASL 1 Regione Abruzzo, Italia. I bambini sono stati sottoposti a una valutazione neuropsichiatrica con misure standardizzate e osservazione clinica secondo i criteri del DSM-5 (APA, 2013) eseguita da un'équipe multidisciplinare esperta nella diagnosi di ASD. Dopo l'esame neuropsichiatrico iniziale, i bambini sono stati sottoposti a una valutazione della sintomatologia ASD (fare riferimento alla sezione "misure") condotta da uno psicologo clinico esperto. La Tabella 1 mostra le caratteristiche demografiche e cliniche del campione.

Tabella 1. Caratteristiche demografiche e cliniche del campione

	Soggetti **N= 91**
Genere (n-%)	
Maschi	63(69,2%)
Femmine	28(30,8%)
Età media in mesi (SD)	24,79 (4,6)
ADOS-2, Modulo Toddler	
AS Totale	15,58 (4,3)
CRR Totale	3,59 (1,9)
Totale	19,21 (5,45)
Livello di rischio (n-%)	
Nessuno a lieve	3 (3,3%)
Lieve a Moderato	11 (12,1%)
Moderato a severo	77 (84,6%)

Misure

ADOS-2, Modulo Toddler

Nello studio è stato preso in considerazione solo il Modulo Toddler. Questo modulo consente di valutare i sintomi precoci legati all'autismo nei bambini di età compresa tra i 12 e i 30 mesi, fornendo una gamma di rischi. È costituito da 11 compiti, più 4 aggiuntivi, che permettono di codificare gli item selezionati per rappresentare i due algoritmi diagnostici: 1) algoritmo "*Tutti i bambini più piccoli/più grandi con poche o nessuna parola*" e 2) algoritmo "*Più grandi con alcune parole*". Per quanto riguarda il presente studio, il nostro campione iniziale era composto da 108 bambini; tuttavia, solo 17 bambini sono stati codificati con l'algoritmo "*Più grandi con alcune parole*". Pertanto, la nostra attenzione si è concentrata sulle prestazioni dei 14 item dell'algoritmo "*Tutti i bambini più piccoli/più grandi con poche o nessuna parola*".

2.3 Analisi

Le analisi sono state eseguite utilizzando IBM SPSS (versione 27.0) (Crop, 2020), il software R (Team, 2021) (R Core Team, 2023) e R Studio (RStudio Team, 2023).

EGA. Prima di eseguire l'EGA (H. F. Golino & Epskamp, 2017), è stata condotta un'analisi preliminare degli item utilizzando l'algoritmo "*Tutti i più piccoli/i più grandi con poche o nessuna parola*" per valutare la normalità degli item e identificare eventuali correlazioni

eccessive (rho > 0,85) tra di essi (Tabachnick & Fidell, 2007). L'approccio EGA è stato quindi applicato per valutare la struttura sottostante degli item del Modulo Toddler, con l'obiettivo di identificare i fattori latenti e gli eventuali cluster di item correlati tra loro. Lo scopo principale dell'EGA è quello di individuare cluster di nodi (o comunità) altamente connessi in base alla separazione funzionale (modularità) (Newman, 2006), che indica se ogni funzione o dimensione valuta una caratteristica distinta del costrutto valutato. Per stimare il numero di fattori o dimensioni è stato utilizzato il metodo GLASSO con correlazioni policoriche (Epskamp & Fried, 2018; H. F. Golino & Epskamp, 2017). Il risultato è una rete che mostra i nodi (item) e le loro relazioni come linee (correlazioni). Il numero esatto di dimensioni è stato ulteriormente esaminato utilizzando l'algoritmo Walktrap (H. F. Golino & Epskamp, 2017; Pons & Latapy, 2005). Per valutare la generalizzabilità dei risultati, abbiamo valutato la stabilità delle dimensioni e degli item utilizzando un approccio bootstrap (Christensen & Golino, 2021). È stato eseguito un bootstrap parametrico con 10.000 iterazioni, ottenendo statistiche descrittive come la mediana, gli intervalli di confidenza al 95% e la frequenza dei numeri dei fattori. La valutazione della stabilità consiste nel valutare la frequenza con cui le dimensioni vengono replicate e la coerenza con cui ogni item compare all'interno di una specifica dimensione (Christensen & Golino, 2021). Secondo Christensen e Golino (2021), la stabilità degli item varia da 0 a 1; valori ≥ 0,70 sono considerati accettabili e suggeriscono che un item è stabile e costantemente presente all'interno di una dimensione.
Inoltre, i carichi di rete (cioè la forza standardizzata dei nodi) sono stati esaminati per valutare il contributo di ciascun item (o nodo) alla coerenza dimensionale. Le dimensioni dei carichi sono state interpretate come segue: piccolo: λEGA > 0,15; moderato: λEGA > 0,25; grande: λEGA > 0,35 (Christensen & Golino, 2021; H. Golino et al., 2020). Infine, è stata applicata una CFA e la bontà dell'adattamento è stata valutata calcolando l'indice di adattamento comparativo (CFI), la radice quadratica media dell'errore di approssimazione (RMSEA), l'indice di bontà dell'adattamento (GFI) e il Chi-quadrato/gradi di libertà (Chisq/df), dove un buon adattamento è stato considerato da CFI > 0,90; RMSEA < 0,08, GFI > 0,90 e Chisq/df < 3.

Compiti e item del modulo Toddler. Successivamente, abbiamo identificato gli item valutati dai compiti strutturati del Modulo Toddler, facendo riferimento al manuale ADOS-2, e quali di questi rientrano nelle dimensioni sintomatologiche dell'ASD. Questo approccio consente alla valutazione di essere più adattabile, con compiti ridondanti che potrebbero essere utilizzati solo quando necessario. I risultati dell'analisi qualitativa sono riportati in appendice (Tabella A1).

2.4 Risultati

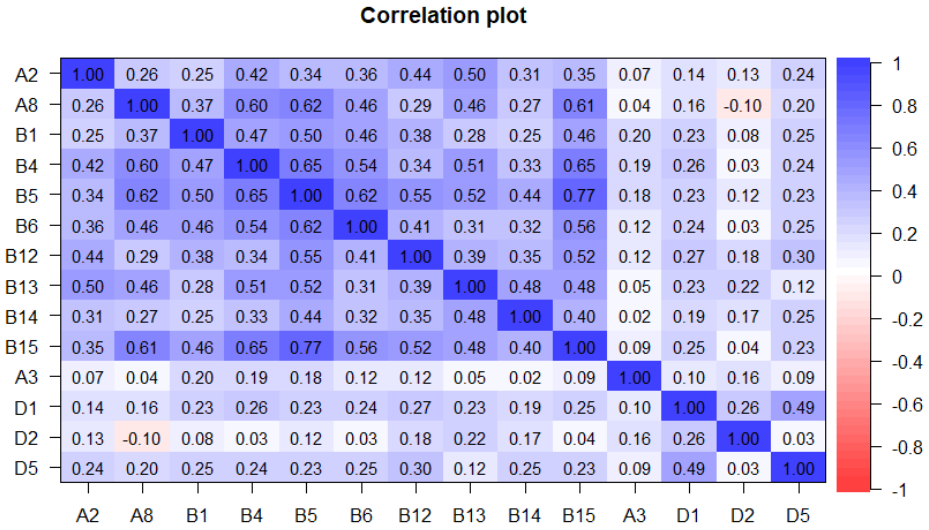
EGA

La Tabella 2 mostra le statistiche descrittive degli item del modulo Toddler. Gli item mostrano una distribuzione normale univariata con skewness e kurtosis entro le soglie desiderate. L'associazione tra gli item è stata valutata con i coefficienti di correlazione di Spearman (Figura 1); i risultati mostrano correlazioni da basse a moderate e, correlazioni negative tra gli item A8 e D2 (r = -0,10).

Tabella 2. Statistiche descrittive degli item del Modulo Toddler

Item		Mean (sd)	Skewness	Kurtosis	S.e
A2	Frequenza delle vocalizzazioni spontanee rivolte agli altri	2.36 (0.78)	-0.86	-0.41	0.08
A8	Gesti	1.69 (1.06)	-0.14	-1.27	0.11
B1	Contatto Oculare insolito	1.89 (0.67)	-0.95	1.65	0.07
B4	Espressioni facciali dirette agli altri	1.21(0.85)	0.03	-0.92	0.09
B5	Integrazione dello sguardo e di altri comportamenti durante le aperture sociali	1.85(0.93)	-0.11	-1.18	0.10
B6	Divertimento condiviso nell'interazione	1.65(1.06)	-0.22	-1.19	0.11
B12	Mostrare	2.51(0.91)	-1.67	1.50	0.10
B13	Inizio spontaneo dell'attenzione congiunta	2.12(0.94)	-0.87	-0.17	0.10
B14	Risposta all'attenzione congiunta	1.67(0.83)	0.09	-0.77	0.09
B15	Qualità delle aperture sociali	1.84(0.91)	0.15	-1.47	0.10
A3	Intonazione di vocalizzi e verbalizzazioni	0.65(0.92)	1.07	-0.19	0.10
D1	Interesse sensoriale insolito per materiali/persone di gioco	1.01(0.85)	0.52	-0.39	0.09
D2	Movimenti e posture di mani e dita	0.87(0.95)	0.57	-1.01	0.10
D5	Interessi insoliti e ripetitivi o comportamenti stereotipati	1.29(0.85)	0.08	-0.72	0.09

Figura 1. Heatmap dei coefficient di Spearman tra gli item del Toddler.



I risultati dell'EGA mostrano una rete caratterizzata da 3 dimensioni (Figura 2A); anche i risultati EGA bootstrappati mostrano una soluzione a 3 dimensioni (mediana= 3; 95% CI [1,41 - 4,58]). Inoltre, l'analisi indica che la soluzione a 3 dimensioni caratterizza la rete per il 46%, indicando una bassa stabilità strutturale. L'analisi della stabilità degli item all'interno delle dimensioni trovate mostra anche la presenza di item instabili (< 0,70) nelle dimensioni 1 e 2 (Figura 2B), in particolare gli item A3, B12 e D2.

Per trovare una soluzione strutturale migliore, abbiamo rianalizzato la rete eliminando i tre item instabili evidenziati dall'EGA. I risultati del secondo modello EGA (Figura 3A) mostrano nuovamente una soluzione tridimensionale, confermata dall'EGA bootstrappata (mediana = 3; 95%CI [1,83 - 4,12]); le dimensioni trovate sono state ottenute per il 69% della rete simulata. I risultati della coerenza strutturale mostrano una buona stabilità per la Dimensione 1 (74%), la Dimensione 2 (67%) e la Dimensione 3 (88%). La buona stabilità strutturale è confermata anche dall'analisi della stabilità degli item all'interno delle dimensioni. I risultati mostrano che gli item presentano una buona stabilità (>0,70) (Figura 3B).

Poiché l'EGA ha raggruppato gli item in modo diverso e ha identificato 3 dimensioni nella struttura del Toddler rispetto a precedenti teorizzazioni effettuate sull'ADOS-2 (Gotham et al., 2007), abbiamo eseguito un'analisi semantica del contenuto degli item raggruppati. L'analisi ha indicato che gli item della Dimensione 1 valutano i comportamenti legati all'"Intenzionalità sociale", gli item della Dimensione 2 valutano gli aspetti legati ai "Comportamenti comunicativi" e gli item della Dimensione 3 rappresentano l'area dei "Comportamenti sensoriali e stereotipati".

Figura 2. Rappresentazione grafica del network e stabilità degli item ottenuti dall'EGA

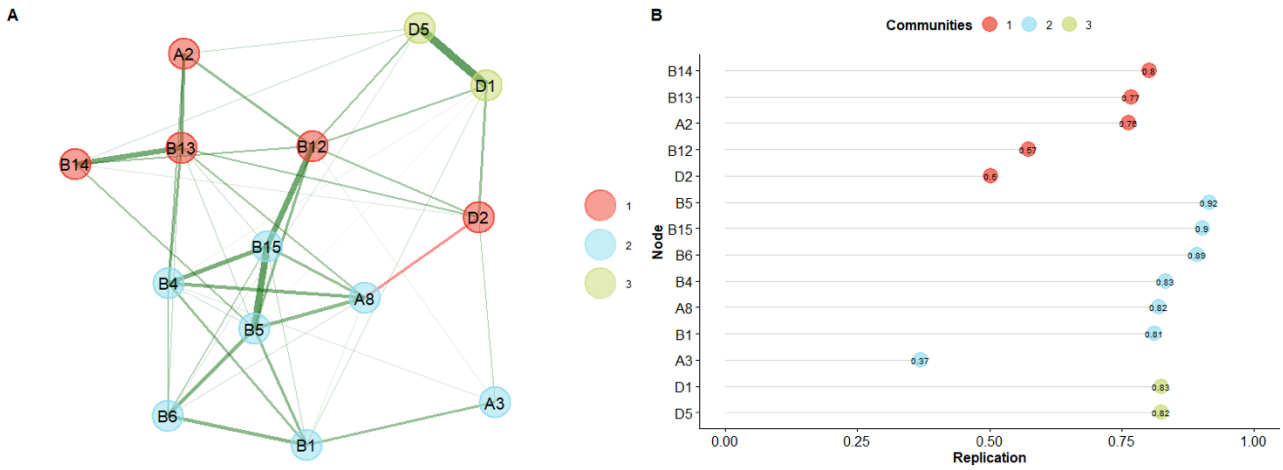
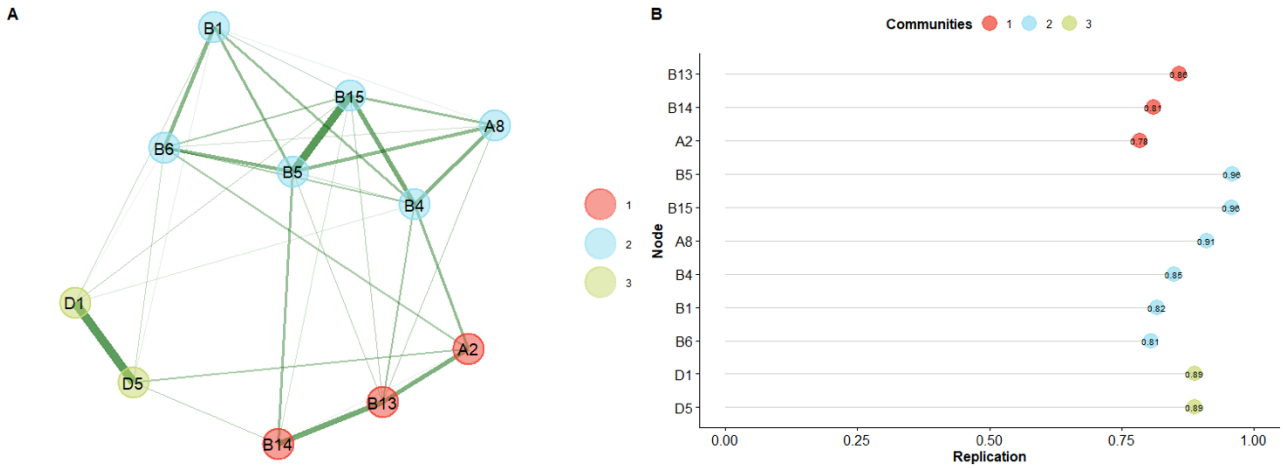


Figura 3. Secondo modello EGA dopo l'eliminazione degli item instabili



Inoltre, l'analisi dei carichi di rete (Tabella 3) ha mostrato che ogni item ha un'associazione ed è coerente con la dimensione. Alcuni item hanno mostrato carichi incrociati, ma questi valori hanno un intervallo da zero a piccolo, inferiori al carico sulla dimensione assegnata. Nella dimensione "Intenzionalità sociale", la λEGA variava da 0,39 (grande) a 0,17 (piccola); la dimensione "Comportamenti comunicativi" mostrava una λEGA da 0,45 (grande) a 0,24 (moderata); la dimensione "Comportamenti sensoriali e stereotipati" mostrava una λEGA di 0,39 (grande).

A conferma del modello strutturale, esso è stato ulteriormente analizzato mediante CFA. L'adattamento del modello ha dato come risultato CFI= 0,95, RMSEA= 0,07, GFI= 0,97, e Chisq/df= 1,48.

Tabella 3. EGA-based network loadings.

	Dimensioni		
Item	**Intenzionalità Sociale**	**Comportamenti Comunicativi**	**Comportamenti Sensoriali e Stereotipati**
B13	0.374	0.154	0
B14	0.205	0	0
A2	0.177	0.108	0
B5	0.144	0.45	0
B15	0	0.441	0
B4	0.205	0.318	0
B6	0	0.262	0
B1	0	0.243	0
A8	0	0.24	0
D1	0	0	0.399
D5	0.123	0	0.399

2.5 Discussione

L'importanza degli approcci multidisciplinari, che incorporano metodi psicometrici e analitici accanto alla pratica clinica e agli strumenti diagnostici, sta diventando centrale, in particolare per comprendere e cercare di standardizzare le manifestazioni sintomatiche dell'ASD. Tali approcci svolgono un ruolo fondamentale nella diagnosi e nell'identificazione precoce dei bambini a rischio. In letteratura ci sono diversi studi che suggeriscano che la manifestazione precoce dei sintomi dell'ASD è molto variabile tanto che viene concettualizzata attraverso tre modelli di insorgenza dei sintomi: il modello ad insorgenza precoce, il modello a regressione e il modello "plateau" (Ozonoff & Iosif, 2019; Shumway et al., 2011). Per migliorare l'efficienza e l'efficacia nella valutazione dell'ASD, l'ADOS-2 è stato esaminato in ricerche precedenti per esplorare le relazioni tra le evidenze cliniche e gli attributi in linea ai criteri del DSM-5 (Pino et al., 2018). In questo studio, è stata applicata un'EGA per replicare e indagare la struttura sottostante degli item del Modulo Toddler in un campione italiano di bambini di età compresa tra 12 e 30 mesi. Questa analisi ci ha permesso di confrontare i risultati dell'EGA con l'analisi dei compiti strutturati, esaminando se i cluster di item identificati si allineano con le categorie comportamentali valutate dai compiti del Modulo Toddler.

I risultati dell'EGA mostrano che i 14 item dell'algoritmo selezionato ("*Tutti i bambini più piccoli/più grandi con poche o nessuna parola*) generano 3 dimensioni come soluzione strutturale ottimale. Tuttavia, tre item - A3 "Intonazione dei vocalizzi e delle verbalizzazioni",

B12 “Mostrare” e D2 “Movimenti delle mani e delle dita/postura” – sono incoerenti all'interno delle dimensioni, compromettendo la stabilità complessiva della rete (Figura 2). In particolare, gli item A3 e D2, che teoricamente appartengono alla dimensione CRR (Lord et al., 2012), appaiono instabili nella nostra analisi, comparendo in dimensioni diverse da quelle ipotizzate. Questa incoerenza potrebbe essere legata alle caratteristiche di sviluppo dei bambini di questa fascia d'età. Ad esempio, l'item A3 può fornire informazioni limitate sulle caratteristiche indagate, soprattutto per i bambini più piccoli o per quelli con abilità verbali assenti. Il manuale dell'ADOS-2 suggerisce di codificare “...vocalizzi ripetitivi o non parole insolite”, ma questo potrebbe non essere applicabile se il bambino non vocalizza a sufficienza durante la valutazione. Inoltre, la codifica di A3 consente di attribuire un punteggio di “8” (non applicabile), che viene convertito in “0” nel punteggio finale, rappresentando una mancanza di informazioni piuttosto che una caratteristica osservata.

L'item D2, invece, codifica per “...movimenti o posture insolite e ripetitive di mani e dita”. L'instabilità all’interno della dimensione può derivare dalla specificità dell’item, in quanto si concentra esclusivamente sui movimenti delle mani, che potrebbero non essere chiaramente evidenti durante la valutazione. Movimenti più complessi, come “flapping hand”, sono invece codificati nell'item D3 (“Altri manierismi complessi”), che viene escluso dall'algoritmo di punteggio finale.

Per ottenere la migliore soluzione strutturale, abbiamo replicato l'analisi, escludendo gli item incoerenti e confermando una soluzione tridimensionale con una migliore stabilità (Figura 3). Ogni item ha mostrato indici di stabilità e replicabilità da buoni a eccellenti, oltre a carichi di rete buoni (Tabella 3). Pertanto, dall'analisi del contenuto semantico degli item raggruppati nelle tre dimensioni identificate dall'EGA, queste dimensioni sembrano rappresentare i domini relativi all'Intenzionalità Sociale (Dimensione 1), ai Comportamenti Comunicativi (Dimensione 2) e ai Comportamenti Sensoriali e Stereotipati (Dimensione 3). I risultati dell'EGA per il Modulo Toddler forniscono indicazioni su come gli item e le dimensioni latenti sono correlate e come sono posizionati graficamente, evidenziando le connessioni tra le dimensioni latenti (Christensen et al., 2019).

I nostri risultati non hanno replicato la soluzione a due fattori dell'ADOS-2 (Lord et al. 2012); tuttavia, la precedente versione dell'ADOS era stata costruita per esaminare tre dimensioni, coerentemente con il DSM-IV, che proponeva un modello triadico per la diagnosi di ASD. Questo modello triadico comprendeva compromissioni nell'interazione sociale, nella comunicazione e la presenza di comportamenti e interessi ristretti e ripetitivi. Con l'aggiornamento del DSM-5 (APA, 2013) e l'adozione di un modello diadico, nell'ADOS-2 i

due domini relativi all'interazione sociale e alla comunicazione sono stati combinati in un'unica dimensione (Affetto Sociale o AS). Una differenza rispetto agli studi precedenti è che nel nostro modello gli item di ciascuna dimensione sono raggruppati in modo diverso. Nella dimensione Intenzionalità Sociale (IS), gli item A2, B14 e B15 - che indagano i comportamenti che riflettono la "consapevolezza" degli altri e la reciprocità sociale, abilità che emergono con l'attenzione congiunta intorno ai 12 mesi di età (Happé & Frith, 2014) - sono raggruppati insieme. Secondo il manuale ADOS-2, questi item sono codificati in base ai tentativi del bambino di richiamare e rispondere all'attenzione del caregiver durante la valutazione. Per la dimensione Comportamenti Comunicativi (CC), la nostra analisi ha raggruppato gli item che descrivono le abilità di comunicazione del bambino nell'ambito delle interazioni sociali, comprendendo l'uso di gesti, il contatto visivo e le espressioni facciali, nonché i comportamenti che indicano il divertimento condiviso e l'avvio di interazioni sociali con il caregiver. Infine, come proposto nell'ADOS-2, gli item che valutano gli interessi sensoriali e i comportamenti ripetitivi (D1 e D5) rimangono nella dimensione ipotizzata. Sulla base dei risultati dell'EGA, la struttura dimensionale è stata testata anche con un modello CFA, che ha dimostrato buoni indici di adattamento.

Per passare dall'analisi del costrutto e della dimensionalità all'applicazione dei compiti strutturati dello strumento, abbiamo confrontato l'aggregazione degli item all'interno delle dimensioni derivate con i compiti del Modulo Toddler. Risulta che quattro compiti elicitano le caratteristiche principali rappresentate dalle tre dimensioni (IS, CC e CSS): "Gioco libero", "Gioco con le bolle", "Anticipazione di una routine con oggetti" e "Bagnetto" (vedi Materiali supplementari). Ciò suggerirebbe una struttura di compiti "principali" e compiti "opzionali/complementari" per la somministrazione di questi compiti strutturati. Il "Gioco Libero" rappresenterebbe il compito principale, in cui si crea una situazione in cui il bambino non viene sollecitato dall'esaminatore. Questo compito consente di valutare il coinvolgimento del bambino con il genitore (o l'esaminatore), i modelli di comunicazione e di affetto e la presenza di comportamenti stereotipati. In questo contesto, l'esaminatore può anche valutare aree specifiche, come la risposta al nome, il sorriso sociale e la risposta all'attenzione congiunta. Il compito "Gioco con le bolle" e il compito "Anticipazione di routine con gli oggetti", insieme ai loro subtest ("Scherzo" e "Non si gioca più"), consentono di codificare gli stessi item e le stesse caratteristiche, ovvero l'ingaggio, il divertimento condiviso, l'avvio spontaneo dell'attenzione, i comportamenti ripetitivi e le reazioni a situazioni sociali insolite create dall'esaminatore. Se l'esaminatore è in grado di codificare questi elementi attraverso, ad esempio, il "Gioco delle bolle", potrebbe essere superfluo somministrare un compito aggiuntivo

che susciti le stesse caratteristiche. Al contrario, questi compiti potrebbero integrarsi a vicenda quando è necessario indagare la stessa caratteristica (o item) in più contesti. Durante la valutazione, non è insolito osservare reazioni avverse nel compito "Anticipare una routine con oggetti" (ad esempio, il bambino prova paura per il rumore del palloncino gonfiato) o una mancanza di interesse, che può portare all'interruzione del compito. Reazioni simili possono verificarsi anche nel "Gioco delle bolle". Un approccio così strutturato ha implicazioni significative per la pratica clinica, perché consente di raccogliere informazioni utili per la diagnosi, riducendo i tempi di somministrazione e minimizzando la necessità di una cooperazione prolungata da parte del bambino. Ciò è particolarmente vantaggioso nelle valutazioni di bambini molto piccoli, che spesso presentano alcune caratteristiche specifiche di questa fase precoce dello sviluppo, come la reattività al test (ad esempio, si stancano facilmente) e la limitata capacità di attenzione. Risultati simili sono stati ottenuti utilizzando metodi analitici innovativi, come la FPA (Pino et al., 2018), che ha permesso di analizzare le relazioni tra gli item di tutti e cinque i moduli dell'ADOS-2 e gli attributi clinici per la diagnosi di ASD previsti dal DSM-5 (APA, 2013). Questa analisi ha identificato le relazioni di prerequisito tra i diversi compiti del Modulo Toddler, come il "Gioco delle bolle" che precede il compito "Anticipazione di una routine con oggetti" o l'"Imitazione funzionale e simbolica" che precede la somministrazione del "Bagnetto". Pertanto, rendere la valutazione più adattiva consentirebbe di ridurre i tempi di somministrazione e di concentrarsi su aspetti diagnostici specifici senza indagare ripetutamente la stessa caratteristica (Pino et al., 2018).

2.6 Conclusioni

Sebbene il presente studio riporti risultati preliminari inerenti alla strutturazione di prove del Modulo Toddler, è necessario riconoscere alcune limitazioni del presente studio. Un primo limite è data dalla grandezza del campione analizzato; un'altra limitazione è che abbiamo analizzato, anche se in via preliminare, solo un algoritmo diagnostico del Modulo Toddler ("*Tutti i bambini più piccoli/più grandi con poche o nessuna parola"*) perché il numero di bambini valutati con il secondo algoritmo del modulo *("Più grandi con alcune parole")* era molto ridotto.

Riteniamo che i nostri risultati siano rilevanti non solo per la ricerca sulla caratterizzazione della sintomatologia ASD, ma anche per la pratica clinica e riabilitativa. Rendere più efficiente ed efficace la valutazione precoce selezionando test "principali" e "opzionali/complementari" ha implicazioni non solo per ridurre i tempi di somministrazione, ma anche per ottimizzare la raccolta di informazioni.

Studio 3

Autismo e test di attribuzione dell'intenzione: una valutazione non verbale con fumetti

3.1 Scopo del lavoro

Lo studio ha avuto come obiettivo quello di proporre *l'Intentions Attribution - Comic Strip Test* (IA-CST), un test mirato a valutare la capacità di inferire le intenzioni dei personaggi e di comprenderne i comportamenti. Questo test può essere interpretato come una misura della capacità di base di ToM e come un precursore delle capacità di mentalizzazione di ordine superiore. Pertanto, lo scopo del nostro studio è duplice: 1) convalidare un nuovo test non verbale per la valutazione dell'attribuzione di intenzioni su un ampio campione, inclusi adolescenti e adulti (Studio 1); 2) confrontare la prestazione di individui con ASD con controlli abbinati per QI (Studio 2). In una prospettiva più ampia, il nostro studio mira a fornire uno strumento standardizzato per la valutazione di una capacità di ToM che sia pratica e priva di linguaggio, da integrare nell'ambiente clinico per supportare la valutazione diagnostica e la pianificazione dell'intervento nei disturbi dello spettro autistico.

3.2 Studio 1: Costruzione e validazione dell'IA-CST

3.2.1 Metodi

Procedura

La validazione della scala è stata eseguita seguendo tre fasi (Flake et al, 2017). Nella prima fase, gli item sono stati costruiti e rivisti da esperti (2 psicometrici, 1 neuropsicologo e 1 biostatistico) con una vasta conoscenza ed esperienza nel campo dell'autismo e della cognizione sociale. Successivamente, è stato utilizzato un campione di convenienza per rimuovere possibili elementi confondenti (vale a dire, elementi che potrebbero essere stati poco chiari o difficili da interpretare). Nella seconda fase, abbiamo eseguito una validazione strutturale della scala eseguendo un'Exploratory Graph Analysis (EGA) e un'Analisi Fattoriale Confermativa (CFA). Inoltre, è stata analizzata la relazione tra la prestazione sull'IA-CST e l'Advanced-Theory of Mind (Blair & Cipollotti, 2000; Prior et al., 2003), un test verbale di ToM cognitiva (validità concorrente). Infine, nella terza fase riportata nello Studio 2, valutiamo preliminarmente la validazione esterna della nuova misura che valuta la validità di gruppo nota. Il Comitato etico ha approvato il protocollo prima del reclutamento dei partecipanti, secondo i principi stabiliti dalla Dichiarazione di Helsinki. Il consenso informato scritto e le informazioni socio-demografiche sono stati forniti da tutti i partecipanti, così come dai loro genitori se minorenni,

prima della somministrazione del test. Ogni partecipante è stato valutato individualmente in una stanza tranquilla senza distrazioni. Le loro risposte sono state registrate utilizzando carta e matita. Uno psicologo era presente nella stanza durante la somministrazione per fornire ulteriori informazioni, se necessario.

Misure

Matrici Progressive Standard di Raven (SPM)

Le Matrici Progressive Standard di Raven (SPM) (Raven, 1983) sono state utilizzate per valutare l'intelligenza non verbale e il livello di QI è stato calcolato seguendo le tabelle di conversione. Le SPM sono costituite da 60 item, ognuno dei quali richiede il completamento di un set di figure con quella mancante, da uno schema presentato. Ogni item diventa progressivamente più difficile, richiedendo capacità di analisi, codifica e interpretazione. Abbiamo scelto di utilizzare le SPM perché la sua somministrazione è più breve, meno impegnativa e meno stressante dei tipici test del QI (ad esempio le scale Wechesler).

Advanced Theory of Mind (A-ToM)

A-ToM (Blair & Cipollotti, 2000; Prior et al., 2003) è un adattamento italiano di un compito di ToM cognitiva (ad es. Strange Stories) che Blair e Cipollotti (2000) hanno utilizzato e che è stato proposto per la prima volta da Happé (1994). Happé (1994) ha definito il compito Strange Stories come un compito di ToM "avanzato", ritenendolo utile per gli individui con forme di ASD ad alto funzionamento, che potrebbero superare i test di ToM (di primo ordine). Questo compito include un'indagine a due livelli degli stati mentali del protagonista della storia: le storie contengono una domanda di comprensione e una domanda chiave per spiegare la causa del suo comportamento (Pino & Mazza, 2016). Nel nostro studio, abbiamo utilizzato l'adattamento italiano (Prior et al., 2003), che consiste in una versione più breve di 13 storie che descrivono eventi reali; per un'interpretazione corretta, il compito richiede al soggetto di andare oltre il significato letterale del testo e di trarre un'inferenza sullo stato mentale del protagonista della storia. A ciascun item viene assegnato un punteggio pari a 1 se si risponde correttamente alle domande di comprensione e giustificazione, e 0 in caso contrario. Per maggiori dettagli, fare riferimento a Happé (1994) e Mazza e colleghi (2022).

Intention Attribution – Comic Strip Test (IA-CST)

IA-CST è un test a fumetti che valuta la capacità di dedurre le intenzioni dei personaggi e di comprenderne i comportamenti. Consiste in sei vignette simili a cartoni animati che illustrano

una sequenza di azioni intenzionali eseguite da un personaggio in uno scenario di vita quotidiana. Il test include anche una serie di item che valutano la causalità fisica. Per ogni item, ai partecipanti vengono mostrate tre vignette che descrivono l'azione o la relazione causale, dopodiché vengono presentate tre vignette, ciascuna contenente una conclusione alternativa allo scenario. Al partecipante viene chiesto di scegliere il finale corretto tra quelli proposti. Nello specifico, i tre possibili finali rappresentano: a) il finale corretto, che è intuibile se il partecipante riesce a comprendere l'intenzione del protagonista (o la relazione causale); b) un finale sbagliato molto simile all'ultima immagine della sequenza; c) un finale sbagliato che descrive un'azione quotidiana non associata alla sequenza. Le conclusioni corrette indicate dai partecipanti rappresentano un punteggio di 1 punto, altrimenti 0.

Costruzione del IA-CST

Un totale di 38 stimoli sono stati originariamente costruiti da un membro del team di ricerca. Gli stimoli sono stati progettati per suscitare la capacità del partecipante di dedurre l'intenzione del personaggio o l'inferenza causale fisica. In particolare, gli stimoli intenzionali sono stati progettati per cercare di suscitare intenzioni di primo o secondo ordine. Dopo aver costruito gli elementi, il team di ricerca si è riunito per rivedere ciascuno degli stimoli proposti seguendo un approccio di validità di contenuto qualitativo. Durante questo incontro, gli item considerati poco chiari o possibili fattori confondenti o non una misura adeguata delle dimensioni target sono stati rivisti o scartati. Un totale di 12 stimoli sono stati rimossi dal pool, mentre i rimanenti sono stati considerati adeguati dall'intero team di ricerca. Quindi, i 26 stimoli sono stati somministrati al campione sano (vedere la sezione Partecipanti). È stata calcolata la percentuale di risposte corrette per ciascun elemento. Questa analisi preliminare ha portato allo scarto di altri tre item perché presentavano una percentuale estremamente elevata di errori (> 70%); la successiva revisione di questi item ha indicato che le rispettive risposte corrette potevano essere difficili da interpretare e sono stati scartati. Alla fine di questa fase, il numero di item considerati adeguati per il passaggio successivo era 23, da qui in poi codificati da I1 a I23.

Partecipanti

Un totale di 261 individui sani (intervallo di età in anni 14-48, età cronologica media 20,09 ± 4,71, 122 maschi e 139 femmine, QI medio 94,57 ± 4,22), hanno partecipato allo studio durante la seconda fase dello studio. Erano madrelingua italiani e reclutati in modo occasionale da strutture e organizzazioni locali. Se un partecipante ha riferito una storia attuale o passata di

abuso di sostanze, disturbi neurologici e/o psichiatrici sono stati esclusi dallo studio. I dettagli delle caratteristiche dei partecipanti sono riportati nella Tabella 1.

Tabella 1. Dati sociodemografici e risultati delle misure di cognizione sociale del campione sano

	Gruppo Età		**Partecipanti**	**Statistics**	***p***
	14-18	**19-48**			
N	82	179	261		
Genere					
Maschi (N)	52	70	122	$x^2(2) = 13.35$	**<0.01**
Femmine (N)	30	109	139		
Età Media (SD)	15.18 (1.08)	22.34 (3.97)	20.09 (4.71)		
Media Anni Istruzione (SD)	10.16 (1.02)	14.97 (1.14)	13.46 (2.49)		
QI Medio (SD)	94.44 (4.46)	94.64 (4.11)	94.57 (4.22)	$F_{1,259}$= 0.123	0.73
Misure di Cognizione Sociale					
A-ToM	9.30 (1.49)	11.26 (1.92)	10.65 (2.01)	$F_{1,259}$= 62.93	**<.001**
Totale IA-CST	8.45 (0.71)	8.33 (0.68)	8.36 (0.69)	$F_{1,259}$= 1.85	0.17

*I Confronti significativi sono riportati in grassetto.

3.2.2 Analisi Statistiche

L'Exploratory Graph Analysis (EGA) è stata eseguita per indagare i fattori latenti della scala. Questo metodo è stato originariamente proposto da Golino e Epskamp (2017). In questo contesto, gli item della scala sono considerati nodi della rete e i legami tra i nodi coefficienti di correlazione parziale (Golino & Christensen, 2020). In questo approccio, i costrutti latenti sono caratterizzati in termini di sottoreti all'interno di una rete più ampia descritta da nodi (item) e bordi (correlazioni parziali). L'obiettivo principale dell'EGA è individuare cluster di nodi altamente connessi, in base a una funzione di separazione, nota come modularità (Newman, 2006), e a una corrispondente configurazione di separazione ottimale. Questo obiettivo viene raggiunto eseguendo esplorazioni separate di tipo random walk che dovrebbero convergere stocasticamente alla separazione ottimale dei cluster, se il modello esiste. Questo approccio dovrebbe consentirci di definire la struttura sottostante della sottorete. Alcuni punti di forza di questo approccio sono che i suoi risultati sono paragonabili, o addirittura migliori, di altre tecniche tradizionali utilizzate per rilevare le dimensioni latenti (Golino & Epskamp, 2017;

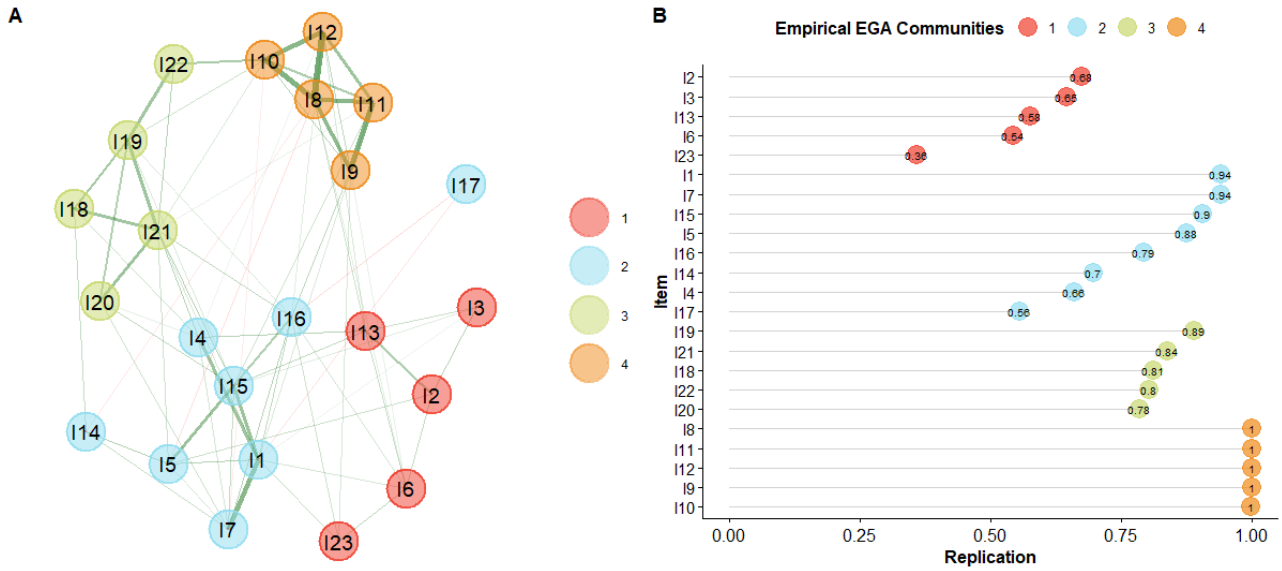
Golino & Christensen, 2020; Golino & Demetriou, 2017), può superare i problemi legati alla scelta di un metodo di rotazione (Browne, 2001; Christensen et al., 2020) e riduce il rischio di errori o bias legati alla ricerca (Browne, 2001; Golino & Christensen, 2020). Una Gaussian Graphical Model (GGM) (Lauritzen, 1996), è stata stimata attraverso una variante del least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996), ovvero il LASSO graphical (Friedman et al., 2008). Si tratta di una tecnica di regolarizzazione utilizzata per stimare il modello e i parametri della GGM (Epskamp et al., 2018). Un parametro di regolazione viene impostato minimizzando un Extended Bayesian Criterio (EBIC) (Chen & Chen, 2008), che è un indice per stimare l'adattamento ottimale del modello (Epskamp & Fried, 2018), che, a sua volta, è regolato da un parametro γ. EGA implementa un algoritmo che imposta γ in base alle connessioni risultanti dalla rete. Per una descrizione dettagliata si rimanda a Golino et al. (2020). LASSO riduce a zero le edges con valori minimi (Epskamp & Fried, 2018; McNeish, 2015) permettendo di limitare le edges falsi positivi e restituendo risultati conservativi e replicabili (Costantini et al., 2015). Il numero di dimensioni è stato individuato utilizzando l'algoritmo Walktrap (Pons & Latapy, 2006) proposto da Golino e Epskamp (2017). Successivamente, la stabilità delle dimensioni e degli item è stata valutata attraverso un approccio bootstrap per valutare la generalizzabilità dei risultati, come proposto da Christensen e Golino (2021). È stato eseguito un bootstrap non parametrico con 1000 iterazioni. Le statistiche descrittive come la mediana, l'IC 95% e la frequenza del numero di fattori sono state ottenute attraverso i bootstrap. Inoltre, riguardo la coerenza strutturale, ossia la frequenza con cui la dimensione empirica dell'EGA viene esattamente replicata, e la stabilità degli item, la frequenza con cui ogni item viene collocato in ogni dimensione, sono stati considerati accettabili valori di stabilità degli item ≥0,70, come indicato da Christensen e Golino (2021). Gli item instabili sono stati quindi rimossi e un altro modello è stato costruito senza quegli item. Il modello finale è stato ulteriormente valutato eseguendo l'analisi fattoriale confermativa (CFA) e la bontà dell'adattamento è stata valutata calcolando l'indice di adattamento comparativo (CFI), l'indice di approssimazione del quadrato medio dell'errore (RMSEA), l'indice di bontà dell'adattamento (GFI) e il Chi Quadrato/Degrees of Freedom (Chisq/df), dove un buon adattamento è stato considerato da CFI> 0,90; RMSEA < 0,08, GFI> 0,90 e Chisq/df< 3. La consistenza interna e l'affidabilità sono state valutate utilizzando l'α di Cronbach. Inoltre, abbiamo eseguito la correlazione di Pearson per indagare la relazione tra l'IA-CST e un test verbale di cognizione sociale (A-ToM).

3.2.3 Risultati

Il primo network ottenuto dall'EGA è riportata nella Figura 1A; il modello indica una soluzione a 4 dimensioni. Il numero mediano di dimensioni mostrato dalle reti bootstrappate era 4 (IC 95% [2,54, 5,45]). Questa analisi indica che il 48% delle reti è caratterizzato da 4 dimensioni. In termini di coerenza strutturale, ossia di frequenza di una dimensione, abbiamo ottenuto la seguente distribuzione: la dimensione 1 è stata osservata il 19% delle volte, la dimensione 2 il 22%, la dimensione 3 il 60% e la dimensione 4 il 99%, il che indica una bassa coerenza strutturale per tutte le dimensioni, tranne che per la dimensione numero 4.

Nel complesso, questi risultati sono stati interpretati come indice di una bassa stabilità strutturale delle reti. Questa conclusione è stata ulteriormente confermata dall'analisi della stabilità degli item (Figura 1B), dove tutti gli item della Dimensione 1 non sono risultati stabili (< 0,70). Questi risultati hanno portato alla rimozione di tutti gli item della Dimensione 1. Inoltre, la Dimensione 2 ha mostrato due item instabili (stabilità dell'item < 0,70) e un item con una stabilità vicina al cut off (I14, stabilità dell'item = 0,71); abbiamo scelto di includere questo item tra gli item instabili, di conseguenza i tre item sono stati rimossi. Gli item sono stati rimossi per trovare una soluzione strutturale migliore per la rete. Successivamente, gli item delle tre dimensioni (esclusa la Dimensione 1) sono stati rivisti per valutare il costrutto rappresentato da ciascuno di essi. La revisione ha indicato che gli item della Dimensione 2 valutavano *l'Attribuzione di Intenzione di Primo Ordine*, ad eccezione di un item che è stato considerato dal team di ricerca un item di effetto causale, quindi è stato rimosso da ulteriori analisi. La revisione degli item della Dimensione 3 ha indicato che essi misurano *l'Inferenza Causale*, mentre la revisione degli item della Dimensione 4 ha indicato che essi misurano *l'Attribuzione di intenzioni di Secondo Ordine*. Pertanto, gli item rimanenti sono stati utilizzati per condurre una seconda EGA.

Figure 1. Risultati dell'Exploratory Graph Analysis e della stabilità ottenuti dalla prima selezione di item.



I risultati del secondo modello EGA sono riportati nella Figura 2A. Il modello indica una soluzione a tre dimensioni, dove la dimensione 1 rappresenta *l'Inferenza Caus*ale, la dimensione 2 rappresenta *l'Attribuzione dell'intenzione di Secondo Ordine* e la dimensione 3 rappresenta *l'Attribuzione dell'intenzione di Primo Ordine*, in base alla revisione precedente. La descrizione di ciascuna dimensione è riportata nella Tabella 2. I risultati del bootstrap hanno indicato una mediana di 3 dimensioni con un CI relativo stretto del 95% [2,34, 3,65]; questo numero di dimensioni è stato ottenuto nell'88% delle reti simulate. I risultati della coerenza strutturale indicano una buona stabilità per la dimensione 1 (78%), la dimensione 2 (85%) e la dimensione 3 (99%). L'insieme di questi risultati è stato interpretato come indicativo di una buona stabilità strutturale delle reti. Questa conclusione è stata confermata anche dai risultati della stabilità degli item, dove hanno mostrato valori elevati di stabilità (tutti > 0,80, si veda la Figura 2B). Pertanto, una struttura a tre dimensioni per gli item considerati nell'analisi può essere considerata una soluzione stabile e replicabile.

Infine, come ulteriore conferma, il modello a tre dimensioni rappresentato nella Figura 2 è stato esaminato attraverso la CFA. L'adattamento del modello CFA ai dati ha dato come risultato CFI= 0,86; RMSEA= 0,07, GFI= 0,91 e Chisq/df= 2,55. Considerando che RMSEA, GFI e Chisq/df indicano un buon adattamento, mentre CFI ha raggiunto un valore prossimo a 0,90,

abbiamo interpretato questo risultato come un adattamento accettabile dei dati. Pertanto, sulla base dei risultati dell'EGA e della CFA, abbiamo stabilito come struttura finale dell'IA-CST il modello a tre dimensioni, mostrato nella Figura 2.

In sintesi, la network analysis ha effettuato due analisi di stabilità strutturale, una riferita al numero di dimensioni che rappresentano i nostri dati, l'altra alla stabilità degli item, permettendoci di capire quanto il caso abbia influenzato le strutture degli item latenti.

Figura 2. Exploratory Graph Analysis e risultati di stabilità degli item dopo la rimozione degli item instabili.

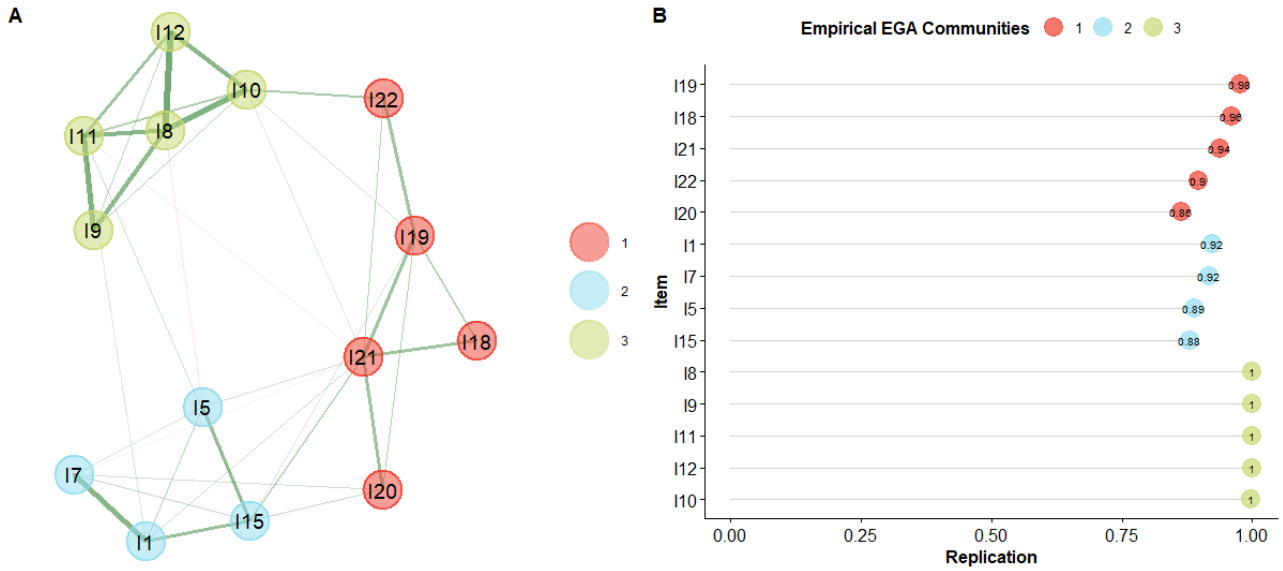


Tabella 2. Descrizione delle dimensioni del modello finale.

Dimensione	*Descrizione*	*Esempio Vignette*	*Item*
Inferenza Causale (Condizione di Controllo)	Una risposta corretta implica la comprensione della relazione causa-effetto degli oggetti o delle persone presenti nella scena. L'inferenza dell'intenzione non è richiesta per superare questi item.	La vignetta mostra un oggetto X che va contro un oggetto Y. La risposta corretta mostra Y che viene spostato dall'oggetto X.	I18, I19, I20, I21, I22
Attribuzione Intenzione Primo Ordine	Una risposta corretta implica la comprensione delle intenzioni del personaggio. La risposta corretta, che deve essere indicata dal partecipante,	La vignetta mostra una persona che vuole raggiungere un oggetto fuori dalla sua portata.	I8, I9, I10, I11, I12

	indica il momento stesso del raggiungimento dell'obiettivo da parte del protagonista della scena.	La risposta corretta raffigura l'uomo con l'oggetto in mano.	
Attribuzione Intenzione Secondo Ordine	Una risposta corretta implica la comprensione delle intenzioni del personaggio. La risposta corretta, che deve essere indicata dal partecipante, indica un'azione necessaria per raggiungere l'obiettivo, che non è ancora stato raggiunto dal protagonista.	La vignetta mostra una persona che vuole raggiungere un oggetto fuori dalla sua portata. La risposta corretta raffigura l'uomo mentre compie un'azione che gli permetterà di raggiungere l'obiettivo.	I1, I5, I7, I15

Scala finale: IA-CST

La scala finale è composta da 14 stimoli suddivisi in tre sottoscale: *Inferenza causale* (C IA-CST), *Intenzione di Primo Ordine* (1st IA-CST) e *Intenzione di Secondo Ordine* (2nd IA-CST). I punteggi totali ottenibili per ciascuna sottoscala vanno da 0 a 5 per il C IA-CST, da 0 a 5 per il 1st IA-CST e da 0 a 4 per il 2nd IA-CST. Per *l'Inferenza causale*, abbiamo calcolato il 5° percentile come soglia; i risultati indicano che per procedere alle altre due condizioni, il soggetto deve rispondere correttamente ad almeno 4 dei 5 item di controllo. Inoltre, è possibile ottenere un punteggio totale delle abilità di *Attribuzione di Intenzione,* sommando le dimensioni dei punteggi di Attribuzione di Intenzione di primo e secondo ordine (Totale IA-CST =1st IA-CST +2nd IA-CST, range 0-9).

L'affidabilità per la scala totale IA-CST è risultata buona (α = 0,7) e molto buona per la sottoscala 1st IA-CST (α = 0,84); un'affidabilità accettabile è stata riscontrata per le sottoscale 2nd IA-CST (α = 0,6) e della C IA-CST (α = 0,6). I risultati delle correlazioni mostrano che le sottoscale delle *Intenzioni di Primo e Secondo Ordine* sono significativamente correlate al punteggio totale dell'A-ToM (rispettivamente r= 0,324, p= 0,017 - r= 0,390, p= 0,004). Inoltre, il punteggio totale delle abilità di *Attribuzione delle Intenzioni* è correlato con il punteggio totale dell'A-ToM (r= 0,492, p<0,001).

3.3 STUDIO 2: Validità Clinica

3.3.1 Metodo

Partecipanti

Trentadue ASD di Livello 1 (età cronologica media 18,53 ± 2,53, QI medio 94,07 ± 10,58, anni di istruzione medi 10,84 ± 1,61) sono stati reclutati presso il Centro Regionale di Riferimento per l'Autismo (CRRA), ASL 1 Regione Abruzzo. La diagnosi è stata effettuata secondo l'ADOS-2 (Lord et al., 2012) e il DSM-3 (APA, 2013) da clinici esperti. I partecipanti con ASD sono stati esclusi dallo studio se presentavano condizioni psichiatriche o mediche concomitanti e un deterioramento cognitivo. Inoltre, trentadue partecipanti a sviluppo tipico (TD) (età cronologica media 19,19 ± 3,01, QI medio 95,70 ± 4,44, anni di istruzione medi 12,63 ± 2,83) sono stati abbinati agli ASD per QI, età e sesso.

3.3.2 Analisi

Per valutare in via preliminare il potenziale uso clinico dell'IA-CST e la validità "known group", il compito dell'IA-CST risultante dall'analisi precedente è stato valutato su un gruppo di 32 TD e 32 ASD appaiati per QI, sesso ed età cronologica. I punteggi dell'IA-CST sono stati confrontati con il test di Mann-Whitney e le differenze tra i gruppi sono state valutate con un test non parametrico poiché, date le dimensioni del campione, abbiamo eseguito un test di Shapiro-Wilk che ha indicato che le misure dell'IA-CST per entrambi i gruppi non seguivano una distribuzione normale. Abbiamo eseguito correlazioni di Spearman tra il QI, l'età e gli anni di istruzione con i punteggi IA-CST per entrambi i gruppi per esplorare le dimensioni associate alla misura. È stato implementato un modello logistico binomiale per esaminare se i punteggi dell'IA-CST potessero essere utilizzati per predire la diagnosi di ASD, quindi la performance diagnostica è stata valutata mediante la Receiver Operating Characteristic curve (ROC); il cut-off ottimale è stato stimato mediante l'indice di Youden, i valori di accuratezza AUC compresi tra 0,90-1,00 sono stati considerati eccellenti, 0,80-0,90 come buona accuratezza, 0,70-0,80 come discreta accuratezza e 0,70-0,60 come scarsa accuratezza (Metz, 1988). È stato poi eseguito un approccio EGA con il campione ASD per valutare le dimensioni dell'IA-CST all'interno del gruppo clinico; tuttavia, il modello più adatto è risultato una rete vuota, pertanto i risultati non sono riportati.

L'analisi è stata eseguita con R (R Core Team, 2020), il pacchetto EGAnet versione 0.9.9 (Golino & Christensen, 2021) e il pacchetto lavaan, versione 0.5 - 12 BETA (Rosseel, 2012). L'analisi della curva ROC è stata eseguita con SPSS 25.0 (IBM Corp, 2017).

3.3.3 Risultati

Trentadue giovani adulti con ASD (28 maschi e 4 femmine) sono stati confrontati con trentadue partecipanti TD (25 maschi e 7 femmine). I gruppi erano appaiati per quoziente intellettivo ($t(62)= 0,80$; $p= 0,42$) e non mostravano differenze per quanto riguarda l'età cronologica ($t(62) = 0,94$, $p = 0,34$) e il sesso (χ^2 $(1, N = 64) = 0,98$, $p = 0,32$). I due gruppi mostrano una differenza negli anni di istruzione ($t(62) = 3,09$, $p <0,01$). Le informazioni demografiche e cliniche sono riportate nella Tabella 3. In seguito, in base ai cluster di item riportati nella Figura 2 e nella Tabella 1, i gruppi sono stati confrontati sul punteggio totale di Inferenza causale (C IA-CST), Intenzione di primo ordine (1st IA-CST) e Intenzione di secondo ordine (2nd IA-CST) dell'IA-CST.

I punteggi dell'IA-CST per entrambi i gruppi sono presentati nella Tabella 4. I risultati hanno indicato una differenza significativa nel1st IA-CST ($U = 318,00$, $z= -2,95$, $p= 0,003$) in cui il gruppo ASD ha mostrato punteggi più bassi ($Mdn = 4,50$) rispetto al gruppo TD ($Mdn = 5,00$) e una differenza significativa nel 2nd IA-CST ($U = 292$, $z= -3,40$, $p<0,001$) in cui il gruppo ASD ha mostrato punteggi più bassi ($Mdn = 3,50$) rispetto al gruppo TD ($Mdn = 4,00$); una differenza significativa nell'IA-CST totale ($U = 251,50$, $z= -3,66$, $p< 0,001$) in cui il gruppo ASD ha mostrato punteggi più bassi ($Mdn = 8,00$) rispetto al gruppo TD ($Mdn = 9,00$). I risultati non hanno evidenziato una differenza per quanto riguarda il C IA-CST ($U = 386,00$, $z= -1,82$, $p= 0,068$) nel confronto tra ASD ($Mdn = 5,00$) e TD ($Mdn = 5,00$).

Le correlazioni di Spearman sono riportate nella Tabella 5; i risultati indicano che solo il QI del gruppo ASD è significativamente correlato con la C IA-CST ($r= 0,42$; $p= 0,02$).

Infine, è stata effettuata una regressione logistica per valutare l'effetto della IA-CST totale sulla probabilità di ricevere o meno una diagnosi di ASD. Il modello è risultato statisticamente significativo rispetto al modello nullo ($\chi^2(1) = 19.60$, $p<0,001$), inoltre ha spiegato il 36% della varianza nell'appartenenza al gruppo (secondo Nagelkerke R^2) e ha previsto correttamente il 68% del campione. L'IA-CST totale è risultato un predittore significativo ($\beta = -1.22$, $SE = 0.40$, χ^2 di Wald $(1) = 9.24$, $p = 0.002$), indicando che questo punteggio potrebbe aiutare a differenziare i gruppi ASD e TD. Sulla base di questo risultato, abbiamo costruito una curva ROC con il punteggio totale IA-CST per valutare la sua performance diagnostica. La curva ROC ha indicato una discreta accuratezza di classificazione dell'IA-CST Totalee (AUC= 0.75,

SE= 0.06, p=0.001, CI95% [0.63, 0.87]), in cui il cut-off migliore per una diagnosi di ASD era un punteggio < 9 (Sensibilità = 71%, Specificità = 66%, Classificazione corretta complessiva = 68%).

Tabella 3. Dati demografici dei campioni e informazioni cliniche relative all'ASD

	ASD (N=32)	**TD (N=32)**	**Test statistic**	***p***
Genere (M; F)	28; 4	25; 7	χ^2 (1, N = 64) = 0.98	0.32
Età Cronologica Media (SD)	18.53 (2.53)	19.19 (3.01)	t(62) = 0.94	0.34
QI medio (SD)	94.07 (10.58)	95.70 (4.44)	t(62) = 0.80	0.42
Media Anni Istruzione (SD)	10.84 (1.61)	12.63 (2.83)	t(62) = 3.09	**<0.01**
Informazioni Cliniche				
Media età prima diagnosi (SD)	10.68 (5.34)	-	-	-
ADOS-2 (Module 4)		-	-	-
Communicazione	4.33 (1.63)	-	-	-
Interazione Sociale	8.66 (3.35)	-	-	-
Comunicazione+Interazione Sociale	13.00 (4.62)	-	-	-
Comportamenti Ristretti e ripetitivi	0.40 (0.51)	-	-	-

I confronti significativi (p<0.05) sono riportati in grassetto.

Tabella 4. Confronto dei punteggi IA-CST tra i gruppi ASD e TD e statistiche descrittive in base al sesso

	Gruppo ASD		**Gruppo TD**			
	Campione totale (N= 32)		***Campione totale (N= 32)***			
	Media (SD)	**Mediana (1st-3rd quartile)**	**Media (SD)**	**Mediana (1st-3rd quartile)**	Z	*p*
C-IACST	4.54 (0.76)	5 (4-5)	4.83 (0.45)	5 (5-5)	-1.82	0.068
1st-IACST	4.28 (0.92)	4.50 (4-5)	4.83 (0.37)	5 (5-5)	-2.95	**0.003**
2nd-IACST	2.96 (1.25)	3.50 (2-4)	3.87 (0.34)	4 (4-4)	-3.40	**< 0.001**
Total-IACST	7.25 (1.84)	8 (6-9)	8.70 (0.46)	9 (8-9)	-3.66	**< 0.001**

IA-CST score per genere				
	ASD maschi (N= 28)		***ASD femmine (N=4)***	
	Media(SD)	**Mediana (1st-3rd quartile)**	**Media (SD)**	**Mediana (1st-3rd quartile)**
C-IACST	4.48 (0.80)	5 (4-5)	5.00 (0.00)	5 (5-5)
1st-IACST	4.25 (0.96)	4.5 (4-5)	4.50 (0.57)	4.5 (4-5)
2nd-IACST	2.96 (1.26)	3.5 (2-4)	3.00 (1.41)	3.5 (1.5-4)
Totale-IACST	7.21 (1.87)	8 (6-9)	7.50 (1.91)	8 (5.5-9)
	TD maschi (N=25)		***TD femmine (N=7)***	
	Media (SD)	**Mediana (1st-3rd quartile)**	**Media (SD)**	**Mediana (1st-3rd quartile)**
C-IACST	4.86 (0.35)	5 (5-5)	4.81 (0.54)	5 (5-5)
1st-IACST	4.86 (0.35)	5 (5-5)	4.81 (0.40)	5 (5-5)
2nd-IACST	3.86 (0.35)	4 (4-4)	3.87 (0.34)	4 (4-4)

Totale-IACST	8.73 (0.45)	9 (8-9)	8.68 (0.47)	9 (8-9)

Nota. Le statistiche Z sono ottenute dal test di Mann-Whitney; le differenze significative (p<0,05) sono riportate in grassetto.

Tabella 5. Correlazioni di Spearman per entrambi i gruppi tra IA ed età con i punteggi IA-CST

	C IA-CST	**1st IA-CST**	**2nd IA-CST**	**Totale IA-CST**
GruppoTD				
Età Cronologica	0.20	-0.01	0.26	0.17
QI	0.02	-0.10	-0.05	-0.12
Anni di Educazione	0.13	-0.02	0.25	0.17
Gruppo ASD				
Età Cronologica	-0.06	0.03	0.12	0.05
QI	**0.42**	0.23	0.23	0.32
Anni di Educazione	0.14	0.17	0.13	0.14

Nota: le correlazioni significative (p<0.05) sono riportate in grassetto.

3.4 Discussione

La Teoria della Mente è ampiamente studiata nella ricerca sull'ASD e sono stati sviluppati diversi strumenti per valutare questa capacità (Mazza et al., 2022; Morrison et al., 2019; Pino et al., 2020). Inoltre, la maggior parte dei test standardizzati di ToM richiede competenze linguistiche espressive e ricettive ben sviluppate che non possono essere utilizzati in individui con capacità verbali scarse o nulle (Colle et al., 2007; Mazza et al., 2022; Pino et al., 2020). Ciò porta a escludere un sottogruppo di individui con ASD (quelli con moderata o grave compromissione del linguaggio e/o individui con disabilità intellettiva) per la valutazione delle abilità di ToM, il cui deficit è considerato una delle caratteristiche principali di questa condizione clinica (Baron-Cohen, 1995; Baron-Cohen et al., 2001; Colle et al., 2007; Mazza et al., 2017). Inoltre, molti compiti di ToM sono lunghi o complessi, quindi potrebbero essere difficili da applicare in ambito clinico (Dodich et al., 2015). Il presente studio è stato condotto per contribuire a valutare l'affidabilità e la validità di un test non verbale, l'IA-CST, che potrebbe essere introdotto nella pratica clinica a supporto della valutazione diagnostica. L'IA-CST consiste in una serie di stimoli di tipo cartoon per valutare l'intenzionalità e l'inferenza causale (condizione di controllo). Alla nostra conoscenza, attualmente non esistono batterie standardizzate o test singoli che consentano di valutare l'attribuzione non verbale delle intenzioni. I risultati dell'analisi strutturale dell'IA-CST hanno evidenziato alcuni item instabili, che sono stati rimossi, e come struttura ottimale un modello a tre dimensioni; la CFA ha ulteriormente confermato questa soluzione, indicando un'adattabilità al nostro campione. Abbiamo quindi ottenuto tre sottopunteggi, ovvero *Attribuzione di Intenzioni Di Primo Ordine*,

Attribuzione Di Intenzioni Di Secondo Ordine e *Inferenza causale*, come punteggio di controllo; è stato possibile calcolare un punteggio totale di *Attribuzione di intenzioni* sommando i due punteggi di primo e second'ordine dell'attribuzione di intenzioni. Nello specifico, per gli item che indagano *l'Attribuzione di Intenzione di Primo Ordine*, la storia è costruita in modo che il finale corretto mostri il momento in cui il personaggio raggiunge l'obiettivo (ad esempio, il personaggio ha in mano l'oggetto desiderato); nell*'Attribuzione di Intenzione di Secondo Ordine*, il finale corretto mostra il personaggio che compie l'azione necessaria per raggiungere l'obiettivo. La condizione di *Inferenza Causale* non comporta un'inferenza di intenzioni, ma la comprensione della relazione causa-effetto di oggetti o persone nella scena (ad esempio, un oggetto che cade). Dal presente studio, suggeriamo di utilizzare il punteggio ottenuto nella condizione di *Inferenza Causale* come punteggio di controllo, che consente di accedere alla serie di attribuzione delle intenzioni se il soggetto è in grado di identificare almeno 4 dei 5 item. Inoltre, i nostri risultati mostrano che l'IA-CST ha una buona validità concorrente. Sebbene abbiamo riscontrato correlazioni significative ma deboli tra le due sottoscale di *Intenzioni di primo e secondo ordine* e l'A-ToM, il punteggio totale dell'*Attribuzione di intenzioni* ha mostrato una correlazione moderata con l'A-ToM. Questi risultati sono coerenti con l'ipotesi che la capacità di comprendere le intenzioni degli altri osservando le loro azioni sia un prerequisito per la capacità di produrre ragionamenti sullo stato mentale e spiegare i pensieri e i sentimenti degli altri, come misurato dall'A-ToM (Leslie, 1994; Premack & Woodruff, 1978). Infatti, secondo Happé e Frith (2014), il costrutto di cognizione sociale può essere inteso come una rete complessa che include componenti distinte, come l'identificazione dell'agente, l'auto-monitoraggio e l'attribuzione dello stato mentale. Tutte queste componenti sono interconnesse e influenzano lo sviluppo di comportamenti sociali appropriati (Happè & Frith, 2014; Happè et al., 2017; Pino, Mariano et al., 2020; Pino et al., 2020; Vagnetti et al., 2020). I nostri risultati dimostrano come le diverse componenti della cognizione sociale, misurate dai test utilizzati, rappresentino abilità diverse ma correlate e siano coinvolte nella comprensione degli agenti sociali e delle interazioni sociali. In particolare, sosteniamo l'ipotesi che una componente, come il riconoscimento dell'azione e l'attribuzione dell'intenzione valutati dall'IA-CST, possa essere una (sotto)componente di un'altra, come l'attribuzione dello stato mentale e l'empatia (Happè et al., 2017). È noto che nei soggetti con ASD vi è uno sviluppo atipico delle diverse (sotto)componenti della cognizione sociale (Pino, Mariano et al., 2020), per cui disporre di misure specifiche che valutano le singole abilità rappresenta un valore aggiunto sia in ambito clinico che di ricerca.

Nello Studio 2 abbiamo valutato in via preliminare la validità clinica dell'IA-CST e abbiamo confrontato le prestazioni del gruppo clinico, individui ASD, con quelle dei controlli con QI corrispondente. Il confronto tra gruppi ha indicato che gli individui ASD non mostrano alcuna difficoltà nella condizione di controllo rispetto al gruppo appaiato per QI. Al contrario, il gruppo ASD mostra una prestazione significativamente inferiore nelle due serie che indagano l'attribuzione di intenzione. Questo conferma che l'IA-CST è un test utile per identificare sottili compromissioni nell'attribuzione dell'intenzione, che è indipendente dalla comprensione degli eventi fisici ma è legata solo alle capacità di mentalizzazione. Infatti, i nostri risultati hanno mostrato che gli individui con ASD sono significativamente compromessi nelle condizioni di attribuzione di intenzione, evidenziando una compromissione dei meccanismi necessari per comprendere le intenzioni degli altri. Poiché è noto che i soggetti con ASD mostrano difficoltà nei processi di cognizione sociale (Baron-Cohen et al., 1986; Happè, 1994; Happè et al., 2017; Pino et al., 2020), ci si aspetterebbe che anche nell'IA-CST i soggetti con ASD incontrino difficoltà, per cui questo risultato indica la nota validità di gruppo del compito. L'analisi di regressione conferma questa evidenza, dimostrando che la capacità di attribuire le intenzioni è un predittore significativo nella differenziazione tra individui ASD e TD. Un'ulteriore conferma viene dall'analisi ROC, che sostiene l'efficacia dell'IA-CST nel discriminare, con un discreto livello di accuratezza, tra il gruppo ASD e quello TD. La migliore sensibilità e specificità dell'IA-CST, e quindi una maggiore probabilità di identificare individui con ASD, sono state ottenute con un punteggio < 9 (cut-off migliore). Nel complesso, questi risultati suggeriscono la possibilità di utilizzare il test nella pratica clinica.

L'IA-CST fornirebbe uno strumento aggiuntivo durante il processo diagnostico, incorporando e affiancando misure considerate gold-standard per la diagnosi e la valutazione dell'autismo. Le valutazioni neuropsicologiche standard e le procedure diagnostiche spesso mancano di informazioni sulle abilità di cognizione sociale e, di conseguenza, non forniscono indicazioni appropriate per il trattamento di questi deficit (Henry et al. 2016). I deficit nelle abilità di ToM sono precoci e rilevanti nell'ASD, quindi la valutazione di questi aspetti, insieme alle caratteristiche sintomatiche, cognitive, funzionali e adattive, è preziosa anche per strutturare gli interventi. Proponiamo uno strumento di facile somministrazione che consente di valutare un'abilità ToM di base necessaria per lo sviluppo di abilità di livello superiore. La valutazione delle diverse componenti della cognizione sociale nell'autismo è importante per delineare un profilo di funzionamento dell'individuo, evidenziando punti di forza e di debolezza. Attualmente, gli interventi per gli adolescenti e gli adulti con ASD più frequentemente riportati in letteratura (Lai et al., 2014) sono principalmente volti a migliorare le abilità comunicativo-

relazionali. Tuttavia, se il trattamento non considera la potenziale compromissione di abilità più basilari, il rischio è che gli interventi sulle abilità di livello superiore non siano efficaci. La costruzione degli interventi deve quindi essere individualizzata, basata sull'evidenza e mirata alla specifica abilità compromessa. L'uso di test specifici, come quello che abbiamo proposto, fornisce anche informazioni preziose per la strutturazione e il follow-up degli interventi per gli individui a basso e medio funzionamento. Siamo consapevoli che il nostro studio presenta alcune limitazioni. Uno di questi è la dimensione relativamente piccola del campione di ASD; studi futuri dovrebbero esplorare le differenze e le prestazioni diagnostiche dell'IA-CST con campioni più grandi. Inoltre, i nostri risultati preliminari hanno indicato un discreto livello di accuratezza in termini di classificazione, suggerendo il suo potenziale di utilizzo nelle valutazioni clinico-diagnostiche. Esistono strumenti gold-standard che chiaramente l'IA-CST non può sostituire, ma può essere utilizzato come strumento integrativo per valutare un'altra dimensione del funzionamento individuale. Da questo punto di vista, vorremmo sottolineare che un punto di forza dell'IA-CST è la sua breve lunghezza (14 item in totale) e la possibilità di somministrarlo a persone con scarse capacità verbali. Il nostro test potrebbe coinvolgere le abilità di elaborazione visuo-percettiva. La letteratura riporta spesso che gli individui con ASD hanno alti livelli di discriminazione visiva e di funzionamento percettivo (Baron-Cohen et al., 2009; Mottron et al., 2006), tuttavia questo aspetto rimane controverso (Funabiki &Shiwa, 2018; Trembath et al., 2015). Infatti, alcuni studi hanno dimostrato che i soggetti con ASD presentano un deficit nell'elaborazione visiva (Funabiki & Shiwa, 2018; Trembath et al., 2015; Williams et al., 2005). Ad esempio, è stato suggerito che i modelli di sguardo atipici negli individui ASD possono influenzare la capacità di comprendere le azioni osservate, principalmente a causa di anomalie nell'attenzione visiva (Masedu et al., 2021; Trembath et al., 2015).

3.5 Conclusioni

La popolazione con ASD è molto eterogenea, quindi è probabile che per alcuni individui con ASD l'uso di immagini possa facilitare le prestazioni, mentre per altri possa essere uno svantaggio (Trembath et al., 2015); quando si interpretano i risultati, è importante tenerlo presente. Uno dei compiti fondamentali del clinico è quello di scegliere i test più appropriati in base alle caratteristiche dell'individuo interessato. Durante il processo di valutazione, il clinico deve prendere in considerazione qualsiasi indagine aggiuntiva pertinente per comprendere appieno e interpretare correttamente i risultati.

Tra i limiti, inoltre, va segnalato che il processo di validazione potrebbe essere ulteriormente ampliato. Le correlazioni significative emerse tra l'IA-CST e l'A-ToM erano deboli/moderate,

per cui questo aspetto dovrebbe essere ulteriormente indagato. Inoltre, nel nostro framework, gli item sono stati rivisti da esperti e, successivamente, attraverso l'EGA, abbiamo valutato la struttura interna e la sua capacità di discriminare un campione che è noto per essere compromesso nel costrutto che lo strumento intende misurare. Inoltre, sebbene gli indici di fit assoluto (RMSEA , GFI) e parsimonioso (Chisq/df) raggiungano i criteri comunemente accettati, l'indice di fit incrementale (CFI) è appena al di sotto del cutoff proposto (>90) (Hopwood &Donnellan, 2010) e, da una prospettiva più ampia, miriamo a migliorarlo, sebbene questi indici non debbano essere interpretati come giudizi binari, ma piuttosto come un'approssimazione dei dati per un modello più realistico invece di giungere ad un fit perfetto (Knekta et al., 2019). Anche se i nostri risultati forniscono prove a favore della validità della misura proposta, ulteriori studi dovrebbero valutare ulteriori aspetti, ad esempio l'affidabilità test-retest, la validità predittiva e altri tipi di validità esterna. Infatti, il processo di validazione di una misura consiste nell'ottenere molte prove a favore della sua validità. Gli studi futuri dovrebbero superare queste limitazioni ed estendere l'applicazione del test anche a soggetti con ASD a basso funzionamento.

Studio 4

Valori normativi e ottimizzazione diagnostica di tre misure di cognizione sociale per la diagnosi di autismo e schizofrenia in un campione di adolescenti e adulti sani

4.1 Scopo del lavoro

Obiettivo dello studio è stato quello di esaminare le proprietà psicometriche mediante l'*Item Response Theory (IRT)* di tre misure di cognizione sociale, Compito Teoria della Mente Avanzato (*Advanced- Theory of Mind;* A-ToM), di Attribuzione dell'emozioni (*Emotion Attribution Task;* EAT) e delle Situazioni Sociali (*Social Situations Task*; SST), utili per indagare rispettivamente le abilità cognitive e affettive della ToM e la comprensione delle norme sociali. Un vantaggio del modello IRT è che permette di stimare i parametri specifici degli item tenendo conto delle abilità degli individui e della difficoltà degli item. Il presente studio estende gli studi precedenti concentrandosi su due aspetti: (a) rivedere, attraverso la metodica IRT, le tre misure di cognizione sociale menzionate e fornire valori normativi da un ampio campione di individui a sviluppo tipico; e (b) valutare la validità clinica delle tre misure riviste valutando le prestazioni di due popolazioni con un funzionamento atipico della cognizione sociale, ovvero ASD e SSD. I valori che forniamo per le tre misure riviste riguardano specificamente partecipanti dall'adolescenza all'età adulta (14-50 anni). Lo scopo dello studio è quello di fornire strumenti adatti a valutare gli aspetti della cognizione sociale che possano essere utilizzati in contesti clinici e di ricerca a supporto delle misure gold-standard.

4.2 Materiali e Metodi

Partecipanti

Campione con sviluppo tipico

Un totale di 580 controlli con sviluppo tipico ha partecipato allo studio. La loro età cronologica variava dai 14 ai 50 anni (M = 20,9, SD = 5,63), con 285 (49,1%) maschi e 295 (50,9%) femmine. Tutti erano di madrelingua italiana e sono stati reclutati per opportunità da scuole locali, facoltà universitarie, organizzazioni sociali e club sportivi. Tutti i partecipanti hanno partecipato volontariamente e tutti, così come i loro genitori se minorenni, sono stati informati sulle procedure e hanno dato il loro consenso informato. Sono stati esclusi i partecipanti che hanno riferito una storia attuale o passata di disturbi neurologici e/o psichiatrici o di abuso di

sostanze. L'intero campione a sviluppo tipico è stato suddiviso in due gruppi di età (adolescenti e adulti), come illustrato nella Tabella 1.

Tabella 1. Valori demografici del campione a sviluppo tipico diviso per due gruppi di età.

Gruppo	N	Maschi	Femmine	Età Media (SD)	Livello educazione Medio (SD)	Livello di Educazione N (%)		
		N (%)	N (%)			Diploma Scuole medie	Diploma Superiori	Università
14 - 19	260	65 (25)	195 (75)	17.79 (1.805)	12.8 (1.81)	75 (28.8)	185 (71.2)	-
20 - 50	320	220 (68.8)	100 (31.2)	23.43 (6.38)	16.4 (1.3)	-	127 (39.7)	193 (60.3)
Full sample	580	285 (49.1)	295 (50.9)	20.9 (5.63)	14.7 (2.36)	75 (12.9)	312 (53.8)	193 (33.3)

Campione clinico

- Un totale di 38 individui (35 maschi e 3 femmine) con ASD con un'età cronologica media di 17,63 anni (SD = 8,25), sono stati reclutati dal Centro Regionale di Riferimento per l'Autismo (CRRA; Valenti et al., 2019), ASL 1 sito in L'Aquila, Regione Abruzzo, Italia. La diagnosi è stata effettuata da clinici esperti, secondo i criteri del Manuale diagnostico e statistico dei disturbi mentali (5a ed.; DSM-5; APA, 2013) e utilizzando l'Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord et al., 2012). Tutti i partecipanti con ASD sono stati classificati come Livello 1 (alto funzionamento) in base ai criteri del DSM-5. I criteri di esclusione erano il deterioramento cognitivo e le condizioni mediche o psichiatriche concomitanti.

- Per lo studio sono stati reclutati 35 individui con diagnosi di SSD (16 maschi e 19 femmine), con un'età cronologica media di 24,43 anni (SD =7,24). Erano giunti per trattamento a seguito di un primo episodio di psicosi presso l'unità di Neuropsicologia Clinica del Dipartimento di Scienze Biotecnologiche e Cliniche Applicate dell'Università dell'Aquila, Italia. Le diagnosi per questo gruppo comprendevano schizofrenia, disturbo schizoaffettivo e disturbo schizofreniforme, effettuate da clinici esperti secondo i criteri del DSM-5 (APA, 2013). La valutazione clinica dei sintomi è stata effettuata utilizzando la versione italiana della Brief Psychiatric Rating Scale-24

(BPRS-24; Roncone et al., 1999). I criteri di esclusione erano il deterioramento cognitivo e le condizioni mediche o psichiatriche concomitanti.

Anche se ASD e SSD possono spesso avere condizioni co-occorrenti, eravamo interessati a esplorare le prestazioni legate a queste due condizioni di per sé piuttosto che all'interazione di effetti legati ad altre condizioni.

Procedure

Lo studio è stato approvato dal Comitato etico locale ed è stato condotto secondo i principi stabiliti dalla Dichiarazione di Helsinki.
Tutti i partecipanti hanno dato il consenso informato scritto e hanno fornito informazioni sociodemografiche prima della somministrazione del test. Non c'erano limiti di tempo; i partecipanti hanno completato le misure in un'unica sessione, impiegando circa 2 ore. Ogni partecipante è stato valutato individualmente in una stanza tranquilla e senza distrazioni; le risposte dei partecipanti sono state registrate con carta e matita. Uno psicologo era presente nella stanza durante la somministrazione per fornire tutte le informazioni necessarie allo svolgimento della batteria di test. L'ordine di somministrazione dei test è stato assegnato in modo casuale a ciascun partecipante, per controllare i bias legati all'ordine di somministrazione. Due membri del team di ricerca hanno effettuato la codifica e il punteggio dei test senza conoscere il gruppo di appartenenza dei partecipanti.

Misure

Matrici Progressive Standard di Raven

Le Matrici Progressive Standard di Raven (RSPM; Raven & Court, 1938) sono un compito a scelta multipla utilizzato per valutare l'abilità associata al ragionamento astratto, definita intelligenza fluida da Cattell (1963), e vengono utilizzate anche come indicatore dell'intelligenza generale (Raven et al., 2000). Il compito consiste in cinque serie di stimoli (A, B, C, D ed E), ciascuna composta da 12 item, per un totale di 60. Gli stimoli consistono in figure astratte in bianco e nero con una parte mancante. All'individuo viene chiesto di osservare gli stimoli, cercare le loro relazioni e indicare, tra le alternative proposte, quella che completa correttamente la figura. In ogni serie, la soluzione del primo item è quasi sempre ovvia, mentre le prove successive sono classificate in ordine di difficoltà crescente. Il test dipende poco dalle abilità linguistiche.

Advanced Theory of Mind (A-ToM)

L'A-ToM (Blair & Cipollotti, 2000; Prior et al., 2003) è un adattamento italiano di un compito ToM cognitivo (cioè Strange Stories) utilizzato da Blair e Cipollotti (2000) e proposto per la prima volta da Happé (1994). Happé (1994) definì il compito Strange Stories come un compito ToM "avanzato" e evidenziandone l'utilità per individui con forme di ASD ad alto funzionamento, che altrimenti avrebbero potuto avere successo nei test ToM di primo ordine. Questo compito prevede l'indagine a due livelli degli stati mentali del protagonista della storia, perché le storie contengono una domanda di comprensione e una domanda chiave per spiegare il suo comportamento (Pino & Mazza, 2016). Nel nostro studio abbiamo utilizzato l'adattamento italiano (Prior et al., 2003), che consiste in una versione più breve, 13 storie, che descrivono eventi reali; per una corretta interpretazione, il compito richiede al soggetto di andare oltre il significato letterale del testo e di trarre un'inferenza sullo stato mentale del protagonista della storia. Ogni storia rappresenta un diverso tipo di attribuzione di stato mentale, ovvero Fingere (due storie); Persuadere, Scherzare, Mentire (due storie); Bugia Bianca (due storie); Fraintendimento (due storie); Ironia; Doppio Bluff; Sarcasmo. Ogni storia è seguita da due domande: una di comprensione (ad esempio, "E' vero quello che ha detto X?") e una di giustificazione (ad esempio, "Perché X ha detto questo?"). Per ogni item viene assegnato un punteggio di 1 se si risponde correttamente alle domande di comprensione e di giustificazione; in caso contrario, viene assegnato un punteggio di 0. Una risposta alla domanda di giustificazione è stata considerata corretta se conteneva un'attribuzione fisica (cioè risposte che si riferivano a eventi non mentali, come l'aspetto fisico, l'azione di un oggetto, eventi ed esiti fisici) o mentale (cioè risposte che contenevano pensieri, sentimenti, desideri o figure retoriche, correttamente identificate). Le risposte non corrette comprendono quelle in cui il partecipante non ha compreso la situazione, ha fatto un'inferenza errata sull'intenzione del protagonista della storia o ha riportato le stesse parole usate nella storia. Per una descrizione più dettagliata di queste categorie, si rimanda a Happé (1994) e Mazza et al. (2022). Il punteggio totale può variare da 0 a 13, dove un punteggio più alto corrisponde a una migliore comprensione dello stato mentale degli altri. Il valore normativo proposto per la popolazione italiana è 12 (Prior et al., 2003).

Emotion Attribution Task (EAT)

L'EAT (Blair & Cipollotti, 2000) rappresenta un compito affettivo della ToM e valuta la capacità di attribuire stati emotivi agli altri. È composto da 58 brevi storie che descrivono situazioni emotive. Al partecipante viene chiesto di leggere le storie e di indicare l'emozione

che meglio descrive il sentimento provato dal protagonista della storia. La versione italiana di questo test è stata realizzata da Prior et al. (2003). Un esempio di storia è: "I quadri di Simone sono arrivati ultimi a un concorso"; poi il soggetto deve rispondere alla domanda: 'Come si sarebbe sentito Simone in questa situazione?' Per la codifica finale, l'esaminatore dispone di un elenco di risposte corrette (emozione target e termini che ne sono sinonimi) per ogni storia. La risposta corretta viene codificata come 1; altrimenti viene codificata come 0. Le storie della versione italiana sono state realizzate per suscitare attribuzioni di emozioni positive e negative, in particolare: 10 storie suscitano felicità, 10 tristezza, 10 paura, 12 imbarazzo, 3 disgusto, 10 rabbia e 3 invidia. Il punteggio totale per ogni categoria di emozione può essere ottenuto sommando il numero di attribuzioni emotive corrette per ogni categoria di emozione. Pertanto, il punteggio di Felicità è compreso tra 0 e 10; il punteggio di Tristezza è compreso tra 0 e 10; il punteggio di Paura è compreso tra 0 e 10; il punteggio di Imbarazzo è compreso tra 0 e 12; il punteggio di Disgusto è compreso tra 0 e 3; il punteggio di Rabbia è compreso tra 0 e 10; e il punteggio di Invidia è compreso tra 0 e 3. Sommando tutte le attribuzioni emotive corrette, si può ottenere un punteggio totale che va da 0 a 58. Un punteggio più alto equivale a una migliore comprensione dell'attribuzione delle emozioni. I dati normativi di Prior et al. (2003) sono stati realizzati su un campione di 120 individui (60 uomini e 60 donne; 37,79 ± 10,7 anni), dove hanno proposto un cut-off per ogni emozione.

Social Situation Test (SST)

Proposto da Dewey (1991), l'SST misura la capacità di giudicare l'adeguatezza del comportamento all'interno delle situazioni sociali e la capacità di distinguere i comportamenti normativi dalle violazioni. I comportamenti normativi sono le azioni riconosciute come generalmente adeguate, mentre le violazioni sono azioni che violano le norme sociali comuni e generano rabbia e disapprovazione (Blair & Cipollotti, 2000). Lo strumento è composto da 25 scene, ognuna delle quali può descrivere uno o più comportamenti sociali normativi, una o più violazioni o entrambi. Così, nel corso delle 25 scene, ai partecipanti viene chiesto di identificare 15 comportamenti normativi e 25 comportamenti di violazione. Il comportamento da identificare è riportato in corsivo. Quando viene descritto un comportamento, si chiede al partecipante di commentare quanto sia appropriato, dando un punteggio da A a D: A (comportamento normale), B (comportamento un po' strano), C (comportamento piuttosto strano) e D (comportamento estremamente strano).
Un esempio di scenario è il seguente: "Giovanna si reca con la figlia di tre anni all'aeroporto. C'erano file lunghissime di persone in attesa di imbarcare i bagagli. *Giovanna si è messa in fila*

con la figlia. (A, B, C, D). Dopo qualche minuto, Giovanna vide sua figlia correre attraverso le uscite del controllo passaporti. Giovanna era molto preoccupata, così chiese a una signora che la precedeva nella fila se poteva andare avanti perché aveva fretta. *La signora non rispose a Giovanna, che la spinse fuori dalla fila e la fece cadere a terra*(A, B, C, D)".

Vengono calcolati tre punteggi: SST Normativo (1 punto per ogni comportamento normativo correttamente identificato come A, per un totale di 15); SST Violazione (1 punto per ogni violazione correttamente indicata con B, C o D, per un totale di 25); e SST Gravità della violazione, calcolato assegnando 0 se il comportamento è giudicato con A, 1 se è giudicato con B, 2 se è giudicato con C e 3 se è giudicato con D, per un totale di 75. Il punteggio SST Gravità viene calcolato solo per gli item che descrivono una violazione delle norme sociali e permette di valutare in che misura i partecipanti giudicano le violazioni socialmente inappropriate. I cut-off per la popolazione italiana sono 13 per SST Normativo, 22 per SST Violazione e 45 per SST Gravità (Prior et al., 2003).

4.3 Analisi statistiche

Modello IRT

Come prima fase, per i tre compiti - A-ToM, EAT e SST - abbiamo condotto l'analisi del funzionamento differenziale degli item (DIF), che consente di indagare le caratteristiche differenziali di un item del test tra sottopopolazioni (o gruppi). L'analisi DIF è utile per identificare gli item che possono essere distorti verso una particolare popolazione (o gruppo) e poi, controllando per il costrutto latente, verificare se le risposte sono correlate all'appartenenza a un particolare gruppo (De Ayala et al., 2013; Kamata & Vaughn, 2004). Abbiamo condotto questa analisi per indagare le differenze di item nella misurazione dei due gruppi di età e di genere nel nostro campione normativo e abbiamo eliminato gli item che presentavano distorsioni nel modello di risposta.

Di conseguenza, gli item che mostravano DIF sono stati rimossi per migliorare l'accuratezza dei compiti nel misurare la loro dimensione latente, tenendo conto dei bias nelle risposte dovuti al sesso e all'età. In particolare, abbiamo utilizzato il metodo di Mantel-Haenszel. Sono stati eliminati anche gli item che presentavano un effetto tetto (risposte corrette ≥ 95%). Successivamente, è stato condotto un modello di Rasch per le risposte binarie degli item, stimato con il metodo della massima verosimiglianza condizionale (CML). Il CML considera la relazione tra le risposte di un individuo e il livello di abilità, fornendo importanti informazioni sull'item per affinare e calibrare le misure (Wright & Linacre, 1989). Inoltre, consente di generalizzare tra individui e item (Fox & Jones, 1998).

La bontà dell'adattamento è stata valutata con il test di Andersen Likelihood Ratio, utilizzando la media come criterio di divisione. Per valutare l'eterogeneità degli item è stato utilizzato anche un controllo grafico del risultato. L'adattamento degli item al test complessivo è stato valutato attraverso gli indici infit e outfit, considerando accettabili valori compresi tra 0,60 e 1,40 (Wright & Linacre, 1994).

Abbiamo quindi ricalcolato i nuovi totali per le tre scale riviste e generato dati normativi per i compiti, calcolando la media e la deviazione standard per i due gruppi di età e per il campione complessivo. Le analisi successive descritte sono state eseguite con le versioni riviste delle misure di cognizione sociale in base ai risultati ottenuti. Abbiamo eseguito la correlazione di Pearson tra l'età cronologica, l'istruzione e i punteggi di A-ToM, EAT e SST per esplorare se le misure potessero essere associate all'età cronologica o ai livelli di istruzione. Sono state eseguite anche correlazioni punto-biseriali per capire se anche il genere potesse presentare un'associazione. Infine, l'affidabilità di A-ToM, EAT e SST è stata valutata calcolando l'alfa di Cronbach, una misura di coerenza interna. L'analisi dei dati è stata eseguita utilizzando il pacchetto eRm (Mair et al., 2021) nel software R (R Core Team, 2021).

Validità clinica

La validità clinica è stata valutata confrontando le prestazioni dei controlli con sviluppo tipico (TD) con quelle di campioni clinici abbinati per età cronologica, sesso e quoziente di intelligenza (QI; misurato da RSPM): l'abbinamento ottimizzato del *propensity score* è stato eseguito utilizzando il pacchetto MatchIt (Stuart et al., 2011). I punteggi dei gruppi sono stati sottoposti ad un test di normalità, con il test di Shapiro-Wilk, e i risultati hanno indicato che non seguivano una distribuzione normale per entrambi i gruppi; di conseguenza, i gruppi sono stati confrontati con il test di Wilcoxon-Mann-Whitney e il calcolo delle dimensioni degli effetti. I confronti multipli sono stati corretti utilizzando il tasso di scoperta dei falsi (FDR).

Le prestazioni diagnostiche di A-ToM, EAT e SST sono state ulteriormente esplorate mediante analisi delle caratteristiche operative del ricevitore (ROC). Il cut-off ottimale è stato stabilito utilizzando l'indice di Youden, che ottimizza la capacità di differenziazione quando si attribuisce lo stesso peso alla sensibilità e alla specificità (Schisterman et al., 2005).

L'ottimizzazione delle prestazioni diagnostiche è stata valutata costruendo una curva ROC basata sui punteggi logit ottenuti da un modello logistico multiplo in cui le misure erano considerate variabili indipendenti e la diagnosi è stata considerata come variabile dipendente. I punteggi di Violazione e Gravità dell'SST sono stati utilizzati in due modelli separati perché sono altamente correlati (cioè rappresentano lo stesso costrutto). Pertanto, sono stati considerati

due possibili modelli. Un'area sotto la curva (AUC) compresa tra .90 e 1.00 è stata considerata come eccellente, tra .80 e .90 buona, tra .70 e .80 discreta, tra .60 e .70 scarsa e tra .50 e .60 nulla (Lüdemann et al., 2006; Metz, 1978). L'analisi ROC è stata eseguita con i pacchetti ROC it (Khan & Brandenburger, 2020) e ROC (Robin et al., 2011). Infine, forniamo i cut-off ottenuti da queste analisi preliminari in modo che queste misure possano essere considerate come potenziali strumenti nel contesto clinico italiano. Date le dimensioni limitate dei campioni dei gruppi clinici, i valori di cut-off dovrebbero essere considerati nell'ambito di un'analisi esplorativa che mira a sottolineare il potenziale uso clinico delle misure studiate.

4.4 Risultati

IRT

Advanced Theory of Mind (A-ToM)

I risultati hanno indicato una DIF dovuta all'età per gli item *2* ($x^2(1) = 10.22, p = 0.001$), *3* ($x^2(1) = 5.79, p = 0.016$) e *12* ($x^2\ (1) = 5.26, p = 0.021$), mentre non è stata riscontrata una DIF dovuta al genere. In particolare, gli item che differivano funzionalmente in base all'età riguardavano *Persuasione*, *Scherzo ironico* e *Sarcasmo*, dove gli odds ratio indicavano che per l'item 2, *Persuasione* (α_{MH} = 2,40), e per l'item 12, *Sarcasmo* (α_{MH} = 1,33), in media, gli adolescenti hanno ottenuto risultati migliori rispetto agli adulti, al contrario dell'item 3, *Scherzo ironico* (α_{MH} = 0,56), dove gli adolescenti hanno ottenuto risultati peggiori rispetto agli adulti. Il modello di Rasch per l'A-ToM ha mostrato un buon adattamento secondo il test di Andersen ($x^2(9) = 6.10, p = 0.73$), per cui il compito finale conteneva 10 item e un punteggio massimo di 10.

Emotion Attribution Task (EAT)

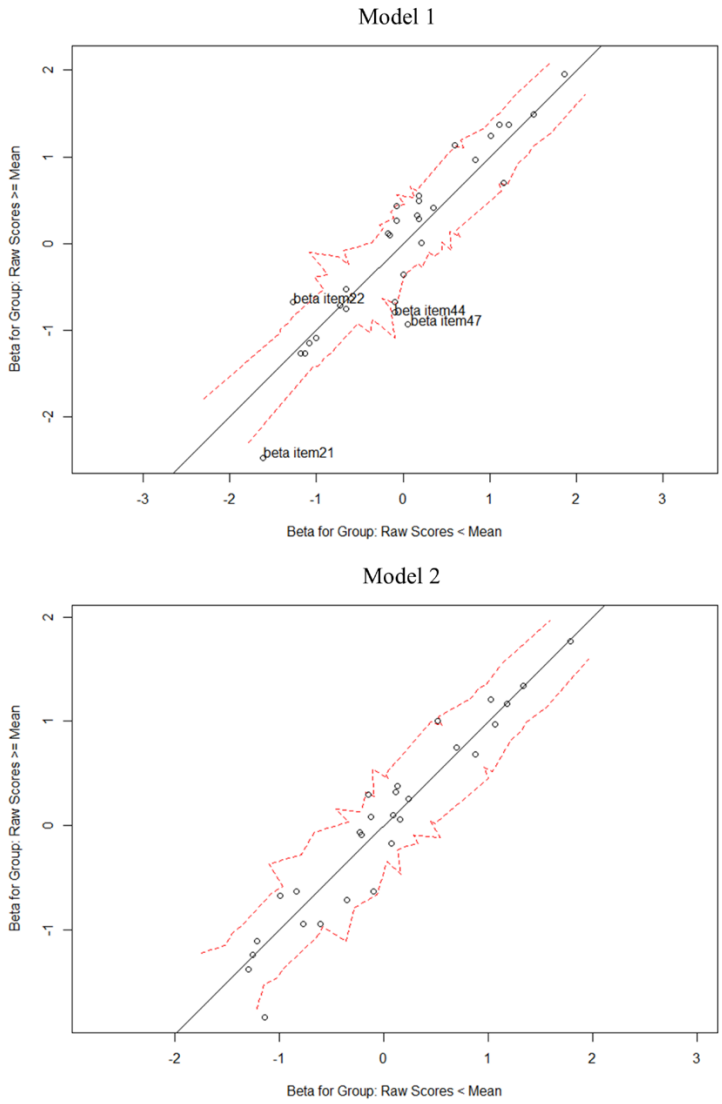
I risultati hanno indicato una DIF dovuta all'età per gli item *8 (*$x^2 = 9.46\ p = 0.002$), *41* ($x^2 = 29.40, p < 0.001$) e *57 (*$x^2 = 20.00, p < 0.001$), che valutano la *tristezza*; gli item *37 (*$x^2 = 16.03, p < 0.001$) e *51* ($x^2 = 7.89, p = 0.005$), che valutano la *paura*; gli item *17* ($x^2 = 6.23, p = 0.12$), *50* ($x^2 = 5.13, p = 0.02$), *52* ($x^2 = 27.03, p < 0.001$) e *54* ($x^2 = 5.30, p = 0.021$), che valutano l'*imbarazzo*; gli item *16* ($x^2 = 6.66, p = 0.009$), *18* ($x^2 = 10.29, p = 0.001$), *25* ($x^2 = 17.08, p < 0.001$), *30* ($x^2 = 15.56, p < 0.001$) e *38* ($x^2 = 7.81, p = 0.005$), che valutano la *felicità*; l'item *10* ($x^2 = 11.48, p < 0.001$), che valuta il *disgusto*; e gli item *12* ($x^2 = 17.66, p < 0.001$), *13* ($x^2 = 4.00, p = 0.045$), *35* ($x^2 = 5.85, p = 0.015$), *40* ($x^2 = 45.90, p < 0.001$) e *55* ($x^2 = 59.12, p < 0.001$) che valutano la *rabbia*. Abbiamo riscontrato risultati contrastanti tra adolescenti e adulti per quanto riguarda le probabilità di risposta corretta nelle varie categorie di emozioni, infatti gli adolescenti hanno presentato una

maggiore probabilità di risposta corretta all'item 57 (α_{MH} = 3,73) per *Tristezza*, all'item 37 (α_{MH} = 16,24) e all'item 51 (α_{MH} = 3,01) per *Paura*; all'item 17 (α_{MH} = 2,01), l'Item 50 (α_{MH} = 1,78) e l'Item 54 (α_{MH} = 2,14) per l'*Imbarazzo*; l'Item 16 (α_{MH} = 2,42), l'Item 30 (α_{MH} = 3,21) e l'Item 38 (α_{MH} = 2,40) per la *Felicità*; l'Item 10 (α_{MH} = 3,18) per il *Disgusto*; e l'Item 35 (α_{MH} = 1,70) per la *Rabbia*.

Gli adulti hanno mostrato maggiori probabilità di risposte corrette all'item 8 (α_{MH} = 0,51) e all'item 41 (α_{MH} = 0,29) per la *tristezza*; all'item 52 (α_{MH} = 0,11) per l'*imbarazzo*; all'item 18 (α_{MH} = 0,50) e all'item 25 (α_{MH} = 0,37) per la *felicità*; e all'item 12 (αMH = 0,38) e all'item 13 (αMH = 0,64) e l'item 25 (α_{MH} = 0,37) per la *Felicità*; all'item 12 (α_{MH} = 0,38), l'item 13 (α_{MH} = 0,64), l'item 40 (α_{MH} = 0,16) e l'item 55 (α_{MH} = 0,08) per la *Rabbia*. I risultati hanno anche indicato una DIF dovuta al genere per l'item 9 ($x^2 = 4.18, p = 0.041$), che valuta la *tristezza*; l'item 46 ($x^2 = 4.41, p = 0.035$), che valuta la *paura*; l'item 48 ($x^2 = 5.68, p = 0.017$), che valuta l'*imbarazzo*; l'item 15 ($x^2 = 14.08, p < 0.001$) e l'item 23 ($x^2 = 11.18, p < 0.001$), che valutano l'*invidia*; e l'item 33 ($x^2 = 14.22, p < 0.001$), che valuta la *rabbia*. Più specificamente, le donne, in media, hanno mostrato maggiori probabilità di rispondere correttamente rispetto agli uomini all'item 46 per la *Paura* (α_{MH} = 1,73); all'item 48 per l'*Imbarazzo* (α_{MH} = 1,80); all'item 15 (α_{MH} = 2,30) e all'item 23 (α_{MH} = 2,00) per l'*Invidia*; e all'item 33 per la *Rabbia* (α_{MH} = 2,73), mentre hanno presentato minori probabilità all'item 9 (α_{MH} = 0,65) per la *Tristezza*.

L'EAT non ha mostrato un buon adattamento nel modello di Rasch ($x^2(31) = 71.71, p < 0.001$); tuttavia, un controllo grafico del modello ha indicato che l'item *21* per la *Paura*, gli item *44* e *47* per l'*Imbarazzo* e l'item *22* per la *Felicità* erano eterogenei tra i gruppi creati secondo il criterio della divisione. Sono stati quindi rimossi. Questa soluzione ha migliorato l'adattamento del modello e il modello si è adattato ai dati ($x^2(27) = 33.19, p = 0.19$). Il compito finale era così composto da 28 item, con un punteggio massimo di 28.

Figura 1. Risultati dei due modelli del test di verosimiglianza di Andersen per l'EAT, solo gli elementi rimossi sono etichettati.



Social Situation Test (SST)

Per il subtest *Violazione*, i risultati hanno indicato una DIF dovuta all'età sugli item 22 ($x^2(1) = 4.76, p = 0.02$) e 24 ($x^2(1) = 4.51, p = 0.03$), e una DIF dovuta al genere sugli item 3b ($x^2(1) = 4.20, p = 0.04$) e 7b ($x^2(1) = 4.87, p = 0.02$). Per quanto riguarda l'SST *Violazione*, gli adolescenti hanno mostrato minori probabilità di risposte corrette rispetto agli adulti sull'item 22 (α_{MH} = 0,22) e maggiori probabilità sull'item 24 (α_{MH} = 2,75), mentre gli odds ratio hanno indicato maggiori probabilità di risposte corrette per gli uomini rispetto alle donne sull'item 3b (α_{MH} = 0,39) e sull'item 7b (α_{MH} = 0,41). Tuttavia, gli item di questo subtest dell'SST - ovvero 1b, 8b, 9, 10, 11, 12b, 16, 17c, 18, 20, 21b, 23b e 25b, nonché gli item del

Subtest *Comportamento normativo* (ovvero 1a, 2a, 5a, 8a, 12a, 14a, 14b, 17b, 21a e 25a) - hanno mostrato un effetto tetto (risposte corrette ≥95%) e sono stati quindi rimossi. Il modello di Rasch ha mostrato un buon adattamento secondo il test di Andersen ($x^2(11) = 11.14, p = 0.43$), quindi nella versione finale il subtest *Normativo* era composto da cinque item e il subtest *Violazione* era composto da otto item; il subtest *Gravità* aveva quindi un punteggio massimo di 24. Il compito finale era quindi composto da 11 scenari.

Le statistiche degli item ottenute dai modelli di Rasch sono riportate in Appendice (Tabella A3) per ogni misura.

Dati normativi dalla popolazione

La correlazione di Pearson non ha mostrato un'associazione significativa tra l'età cronologica e l'A-ToM, l'EAT o l'SST (tutti p>.10), mentre le correlazioni punto-biseriali hanno rivelato un'associazione significativa tra il sesso e il punteggio totale dell'A-ToM (r = 0.12, p = 0.001), il punteggio *Normativo* (r = 0.17, p = 0.001), il punteggio di *Violazione* (r = 0.12, p = 0.001) e il punteggio di *Gravità* (r = 0,13, p = 0,001); tuttavia, i risultati hanno evidenziato che l'entità dell'associazione era estremamente bassa. Inoltre, non è stata trovata alcuna associazione tra sesso e punteggio EAT. Non abbiamo considerato il QI nei valori normativi perché, quando significativo, ha mostrato una piccola associazione con i punteggi totali nel campione normativo: A-ToM (r = 0,12, p = 0,003), EAT (r = 0,10, p = 0,01), punteggio *Normativo* (r = 0,14, p = 0,001), punteggio *Violazione* (r = 0,01, p = 0,81) e punteggio *Gravità* (r = -0,04, p = 0,34). Le correlazioni hanno indicato che solo l'A-ToM era significativamente associato agli anni di istruzione (r = 0,21, p= 0,001), mentre non lo erano l'EAT (r = -0,04, p= 0,33), il punteggio *Normativo* (r = -0,07, p= 0,09), il punteggio di *Violazione* (r= 0,01, p= 0,81) e il punteggio di *Gravità* (r= 0,04, p= 0,33). La correlazione significativa spiegava solo il 4% della varianza del punteggio (R^2= 0,04), quindi non abbiamo corretto i punteggi normativi per questa variabile demografica. L'A-ToM ha mostrato una buona coerenza interna (α di Cronbach = .61), così come l'SST (α di Cronbach = .68), mentre l'EAT ha mostrato un'elevata coerenza interna (α di Cronbach = .83).Una descrizione dettagliata dei punteggi normativi è riportata nella Tabella 2 per il campione complessivo. I punteggi per gruppi di età sono riportati in Appendice (Tabella A3).

Tabella 2. Dati normativi della versione rivista di A-ToM, EAT e SST ottenuti attraverso l'analisi di Rasch per l'intero campione.

	N Item	Media (SD)	95% Intervallo di Confidenza	
			Lower Bound	Upper Bound
A-ToM	10	8.64 (1.36)	8.53	8.75
Emotion Attribution Task	28	19.03 (5.22)	18.58	19.49
Tristezza	6	4.22 (1.55)	4.09	4.36
Paura	6	4.52 (1.40)	4.39	4.64
Imbarazzo	5	2.63 (1.34)	2.51	2.74
Felicità	4	3.11 (1,12)	3.01	3.21
Disgusto	2	1.51 (0,68)	1.45	1.57
Invidia	1	0.66 (0.47)	0.62	0.70
Rabbia	4	2.36 (1.16)	2.26	2.46
Social Situation Task				
Normativo	5	4.05 (0.98)	3.95	4.16
Violazione	8	7.07 (1.10)	6.95	7.19
Gravità	8	14.01 (3.60)	13.62	14.41

Validità clinica

Confronto ASD vs TD

Utilizzando il propensity score matching, abbiamo identificato 38 controlli TD dai campioni normativi che erano comparabili con il campione clinico di 38 individui con ASD in base all'età cronologica, al sesso e al QI. Le statistiche descrittive sono riportate nella Tabella 3.

Tabella 3. Statistiche descrittive dei campioni abbinati per età cronologica, sesso e QI

ASD e TD appaiati				
	ASD (N=38)	TD (N=38)	Test statistic	p
Età cronologica media (SD)	17.63 (8.25)	20.50 (4.69)	t(74)= 1.86	.07
Genere (M; F)	35; 3	32; 7	$Chi^2(1)$= 1.84	.17
QI medio (SD)	94.71 (12.08)	100.69 (13.19)	t(74)= 0.38	.71

SSD e TD appaiati

	SSD (N=35)	TD (N=35)	Test statistic	p
Età cronologica media (SD)	24.43 (7.24)	24.23 (6.02)	t(68) = 0.12	.90
Genere (M; F)	19; 19	17; 21	$Chi^2(1)$= 0.21	.64
QI medio (SD)	90.60 (10.71)	92.14 (18.56)	t(68) = 0.42	.67

Il confronto dei punteggi tra i nuovi totali per l'A-ToM, l'EAT e le tre subcales dell'SST (Normativo, Violazione e Gravità) ha mostrato punteggi significativamente più bassi nell'ASD rispetto al gruppo TD per l'A-ToM (U = 906, p = 0,03, r = 0,36), l'EAT (U = 843. 5, p = 0,02, r = 0,30), il punteggio Normativo dell'SST (U = 937,5, p< 0,001, r = 0,48) e il punteggio di Gravità dell'SST (U = 895, p< 0,001, r = 0,40), mentre i punteggi di Violazione non erano significativamente diversi tra i gruppi (U = 743, p = 0,13, r = 0,19). Per ulteriori dettagli si rimanda alla Tabella 4.

Tabella 4. Confronto delle mediane (1°-3° quartile) tra gruppi clinici e TD sui punteggi A-TOM, EAT e SST.

			ASD vs TD			
	Score range	ASD (N=38)	TD (N=38)	Test statistic	Effect size (r)	p
A-ToM	0-10	8.00 (6.25-9.75)	10 (8.25-10.00)	U = 906	0.36	**0.03**
EAT	0-28	18.50 (14.50-21.00)	21.00 (19.00-22.00)	U = 843.5	0.30	**0.02**
SST						
Normativo	0-5	3.00 (2.00-4.00)	4.00 (4.00-5.00)	U = 937.5	0.48	**0.00**
Violazione	0-8	6.00 (6.00-7.25)	8.00 (6.00-8.00)	U = 743	0.19	0.13
Gravità	0-24	11.00 (9.00-13.25)	14.00 (12.25-16.00)	U = 895	0.40	**0.00**

			SSD vs TD			
	Score range	SSD (N=35)	TD (N=35)	Test statistic	Effect size (r)	p
A-ToM	0-10	8.00 (6.00-9.00)	8.00 (8.00-9.50)	U = 847	0.34	**0.00**
EAT	0-28	19.00 (11.00-24.00)	19.00 (16.00-21.00)	U = 547	0.01	0.95
SST						
Normativo	0-5	3.00 (2.00-5.00)	4.00 (3.00-4.00)	U = 468	0.14	0.33
Violazione	0-8	4.00 (3.00-5.00)	7.00 (6.25-8.00)	U = 795	0.67	**0.00**
Gravità	0-24	4.50 (3.00-9.50)	13.00 (10.25-15.75)	U = 753	0.58	**0.00**

nota: i valori di p sono corretti per FDR; i confronti significativi sono riportati in grassetto

Il modello logistico multiplo ha indicato i seguenti predittori come significativi: il punteggio A-ToM (β = -0,74, *SE* = 0,30, *z* = -2,44, *p* = 0,014), il punteggio *Normativo* dell'SST (β = -1,02, *SE* = 0,37, *z* = -2,76, *p* = 0,005) e il punteggio *Gravità* dell'SST (β = -0,27, *SE* = 0,10, *z* = -2,68, *p* = 0,007). L'EAT non era un predittore significativo (β = -0,19, *SE* = 0,10, *z* = -1,83, *p* = 0,66). Solo i predittori significativi di questo modello sono stati utilizzati per costruire un punteggio combinato. È stato quindi eseguito un modello logistico multiplo senza la variabile indipendente EAT; i coefficienti beta aggiornati sono riportati nella formula sottostante e la loro significatività è stata mantenuta. Di conseguenza, i punteggi sono stati calcolati con la formula:

$$\pi_i = \frac{e^{(12.68-(0.65\ A-ToM)-(1.10\ NORMATIVE)-(0.23\ GRAVITY))}}{1+e^{(12.68-(0.65\ A-ToM)-(1.10\ NORMATIVE)-(0.23\ GRAVITY))}}$$

Il modello logistico con il punteggio di violazione come variabile indipendente non è stato generato, perché non ha fornito risultati significativi in base al confronto delle medie. Le curve ROC ottenute dai punteggi dei compiti e dai compiti uniti secondo il modello sono mostrate nella Figura 2a.

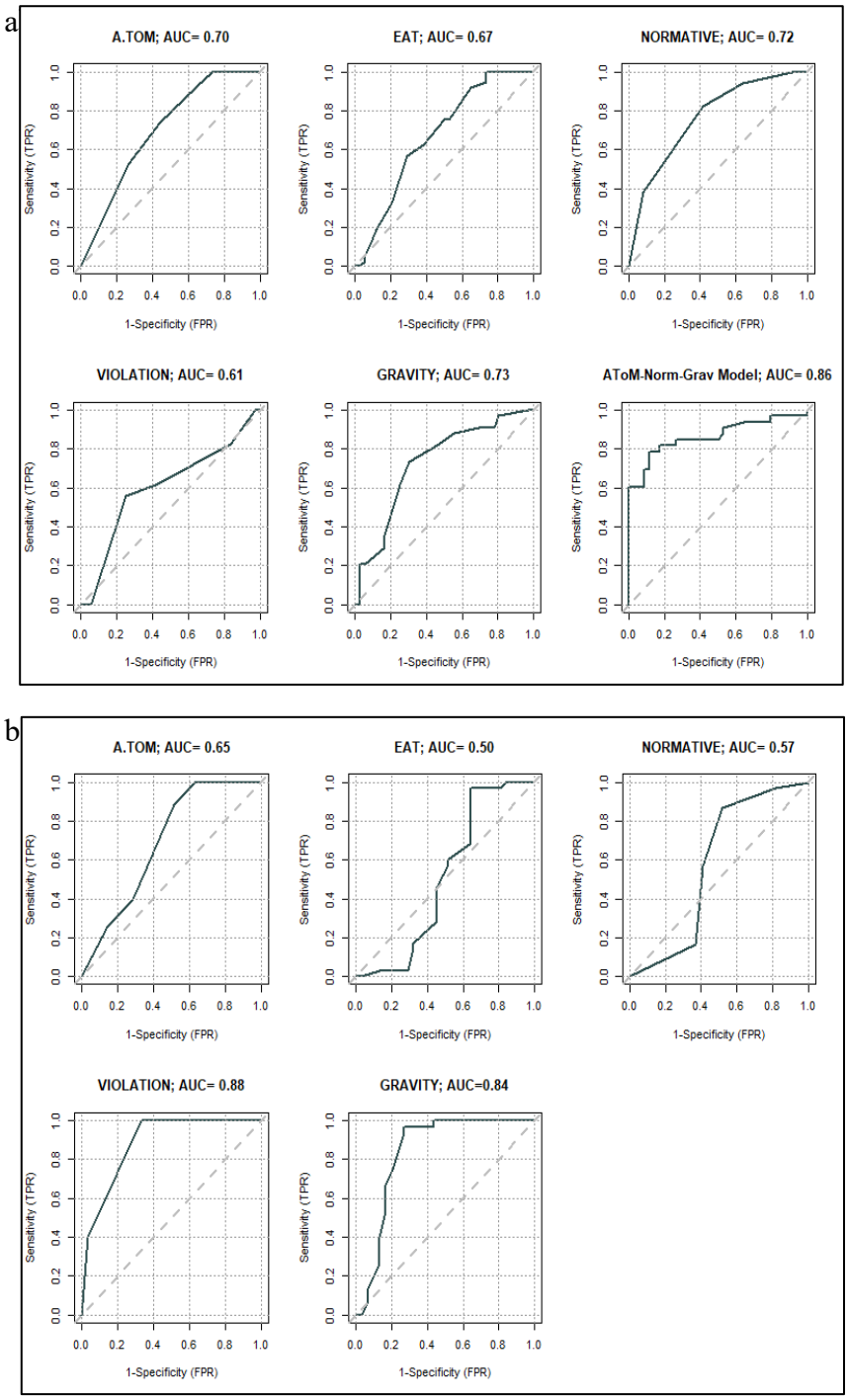
L'AUC ottenuta dalla combinazione dell'A-ToM e dei subtest *Normativo* e *Gravità* ha raggiunto una buona accuratezza ($AUC_{AToM\text{-}Normative\text{-}Gravity}$ = 0,86, Se =85%, Sp = 74%); le accuratezze per le sottoscale *Normativo* e *Gravità* sono state $AUC_{Normative}$ = 0,72 (Se = 82%, Sp = 50%) e $AUC_{Gravity}$ = 0,73 (Se = 73%, Sp = 50%) rispettivamente, mentre A-TOM era $AUC_{A\text{-}ToM}$ = 0,70 (Se = 74%, Sp = 56%); EAT (AUC_{EAT} = 0,67, Se = 76%, Sp = 50%) e *Violazione* ($AUC_{Violation}$ = 0,61, Se = 56%, Sp = 75%) hanno raggiunto una scarsa accuratezza nel distinguere i due gruppi.

Confronto SSD vs TD

Abbiamo identificato 35 controlli TD che corrispondevano ai 35 individui del nostro campione di SSD per età cronologica, sesso e QI (si veda la Tabella 3). I risultati (si veda la Tabella 4) indicano una prestazione ridotta del campione SSD su A-ToM (U = 847, p< 0,001, r = 0,34), sul punteggio di *Violazione* dell'SST (U = 795, p < 0,001, r = 0,67) e sul punteggio di *Gravità* dell'SST (U = 753, p< 0,001, r = 0,58). I gruppi hanno mostrato prestazioni simili sull'EAT (U = 547, p = 0,95, r = 0,01) e sul punteggio *Normativo* dell'SST (U = 468, p = 0,33, r = 0,14). Sono state eseguite due regressioni logistiche multiple, tenendo separati i punteggi di *Violazione*

e *Gravità* in quanto rappresentano lo stesso costrutto. Nel primo modello, il punteggio di *Violazione* è risultato un predittore significativo (β = -0,93, *SE* = 0,24, *z* = 3,80, *p* <0,001) mentre A-ToM non è risultato significativo (β = -0,18, *SE* = 0,21, *z* = -0,83, *p* = 0,40). Nel secondo modello, il punteggio di *Gravità* dell'SST (β =-0,26, *SE* = 0,07, *z* = -3,39, *p* < 0,001) predice significativamente la diagnosi, mentre l'A-ToM (β = -0,31, *SE* = 0,21, *z* = -1,50, *p* = 0,13) no. Quindi, le curve ROC sono state costruite separatamente per ciascun compito.

Figura 2. Prestazioni diagnostiche dei compiti nell'ambito della diagnosi clinica per: a) ASD, il modello si basa sulla combinazione dei punteggi A-ToM, Normativo e Gravità; b) SSD.



Le curve ROC sono riportate nella Figura 2b. I risultati hanno indicato una scarsa performance diagnostica per l'A-ToM ($AUC_{A\text{-}ToM}$ = 0,65, Se = 100%, Sp = 38%), per l'EAT (AUC_{EAT} = 0,50, Se = 97%, Sp = 35%), per il punteggio *Normativo* dell'SST ($AUC_{Normativo}$ = 0,57, Se = 85%, Sp = 50%). Una buona accuratezza è stata invece riscontrata per il punteggio di *Violazione* dell'SST ($AUC_{Violation}$ = 0,88, Se = 80%, Sp = 77%) e per il punteggio di *Gravità* dell'SST ($AUC_{Gravity}$ = 0,84, Se = 97%, Sp = 74%). Infine, nella Tabella 5 sono riportati i migliori cut-off ottenuti da tutte le analisi.

Tabella 5. Miglior cut-off, AUC di sensibilità (Se) e specificità (Sp) per ciascun gruppo di confronto sui punteggi A-TOM, EAT, Normativo, Violazione e Gravità rivisti.

ASD cut-offs				
	Best cut-off	AUC	Se	Sp
A-TOM	9	0.70	74%	56%
EAT	19	0.67	76%	50%
Normativo	4	0.72	82%	50%
Violazione	8	0.61	56%	75%
Gravità	13	0.73	73%	70%
A-ToM/Normativa/Gravità Model	0.36°	0.86	85%	74%
SSD cut-offs				
	Best cut-off	AUC	Se	Sp
A-TOM	7	0.65	100%	38%
EAT	13	0.50	97%	35%
Normativo	3	0.57	85%	50%
Violazione	6	0.88	80%	77%
Gravità	9	0.84	97%	74%

°il cut-off é stato calcolato secondo la formula: $\pi_i = \frac{e^{(12.68-(0.65\ A-ToM)-(1.10\ NORMATIVE)-(0.23\ GRAVITY))}}{1+e^{(12.68-(0.65\ A-ToM)-(1.10\ NORMATIVE)-(0.23\ GRAVITY))}}$

4.5 Discussione

Nel corso degli anni, la complessità del costrutto di cognizione sociale ha portato all'eterogeneità dei modelli teorici e allo sviluppo di numerose misure e strumenti che ne indagano le diverse sfaccettature. Tuttavia, le misure di cognizione sociale sono state utilizzate principalmente nel campo della ricerca per descrivere il funzionamento atipico di queste abilità in popolazioni cliniche come ASD e SSD (Bliksted et al. 2017; Frith, 1992, 2004; Mazza et al., 2010; Mazza et al., 2022; Pino et al., 2017, 2018) e la mancanza di valori normativi per le popolazioni a sviluppo tipico e clinico rimane una sfida importante (Beaudoin et al., 2020; Henry et al., 2016).

Il presente studio ha quindi utilizzato l'IRT per valutare le proprietà psicometriche di tre misure di cognizione sociale (A-ToM, EAT e SST) per fornire una versione rivista e i relativi valori normativi di un campione di partecipanti a sviluppo tipico. Inoltre, è stata studiata la validità clinica per valutare le prestazioni del nostro campione clinico di ASD e SSD.

Scale riviste nel quadro IRT

L'A-ToM è uno dei test più utilizzati per valutare la comprensione degli stati mentali fin dalla prima versione nota in letteratura come Strange Stories (Happè, 1994). Nel corso degli anni, diversi studi hanno cercato di valutare le proprietà psicometriche del test, di produrre versioni riviste e di estendere l'applicabilità e la generalizzazione a diversi gruppi di età (Osterhaus & Bosacki, 2022). Recentemente è stato suggerito (Nawaz et al., 2023) che le Strange Stories presentano incongruenze su alcuni item che dovrebbero valutare lo stesso costrutto di base. I nostri risultati hanno mostrato che tre item dell'A-ToM presentavano un bias nel modello di risposta dovuto all'appartenenza al gruppo, in particolare negli item che rappresentano l'attribuzione dello stato mentale di *Persuasione*, *Scherzo ironico* e *Sarcasmo*. Per quanto riguarda la *Persuasione*, abbiamo ipotizzato che il diverso funzionamento fosse legato a stili di risposta diversi, dovuti alla flessibilità del pensiero nelle situazioni di persuasione (Nippold et al., 2005). Tuttavia, è interessante notare che, sebbene le capacità persuasive maturino durante lo sviluppo (Clark & Delia, 1976), i nostri risultati indicano che il gruppo degli adolescenti aveva maggiori probabilità di dare risposte corrette rispetto agli adulti. Per quanto riguarda il costrutto di *Scherzo ironico*, poiché il DIF era presente nell'item 3 ma non nell'item 10, che rappresenta comunque lo stesso stato mentale, abbiamo ipotizzato che fosse legato alla struttura dell'item piuttosto che al costrutto valutato. Nell'item 12 di *Sarcasmo*, si è ipotizzato che la DIF fosse legata ai personaggi coinvolti nella storia, in quanto possono rappresentare diametralmente i due gruppi di età: una madre (adulta) ha speso molto tempo per cucinare il piatto preferito della figlia (adolescente); poiché la figlia non ha ringraziato, la madre dice: “Bene, questo è un bel comportamento, non trovi?”.

Per quanto riguarda l'EAT, sono stati rimossi 30 item, che hanno mostrato risultati contrastanti tra adolescenti e adulti per quanto riguarda la probabilità di risposte corrette nelle varie categorie di emozioni (tristezza, paura, imbarazzo, felicità, disgusto e rabbia). Inoltre, i risultati hanno indicato anche un DIF legata al genere, ma anche in questo caso con risultati contrastanti per le diverse categorie di emozioni. I nostri risultati sembrano essere in linea con quanto suggerito da un precedente studio italiano (Gatta et al., 2015) che ha analizzato preliminarmente l'affidabilità delle dimensioni emotive dell'EAT. Gatta e colleghi (2015) hanno riportato che

per alcuni item c'era una sovrapposizione di diverse dimensioni emotive (ad esempio, alcuni item di Tristezza interpretati con l'emozione di Rabbia), suggerendo la necessità di indagare ulteriormente l'affidabilità del test in successivi studi di validazione.

Per quanto riguarda l'SST, abbiamo rimosso un totale di 27 item, alcuni dei quali mostravano un bias di risposta in relazione all'età e al genere. A nostra conoscenza, non ci sono studi che abbiano condotto un'analisi degli item di questo strumento, che è principalmente utilizzato in letteratura a scopo di ricerca per indagare le capacità di riconoscimento di comportamenti sociali appropriati in popolazioni cliniche (Blair & Cipollotti, 2000; Prior et al., 2013). La validazione di questa misura è di particolare importanza per la valutazione delle abilità di cognizione sociale, in quanto la capacità di giudicare un comportamento come socialmente appropriato o inappropriato richiede l'acquisizione di norme sociali che sono alla base di un adeguato funzionamento sociale e promuovono il mantenimento delle relazioni interpersonali (Bas-Hoogendam et al., 2017; Cialdini & Goldstein, 2004; Rosenberger et al., 2019). Dunque, la versione rivista delle misure comprende 10 item per l'A-ToM, 28 item per l'EAT e 13 item totali tra i subtest *Normativo e Violazione* (Tabella 2), che hanno mostrato una buona coerenza interna. Un vantaggio dell'eliminazione degli item con informazioni limitate sulle reali capacità dei partecipanti è quello di rendere lo strumento più preciso e generalizzabile. Inoltre, l'uso di strumenti con un gran numero di item richiede molto tempo e può limitarne l'applicazione in ambito clinico e di ricerca. L'uso di misure più brevi è vantaggioso in termini di riduzione del tempo di somministrazione e del carico cognitivo nei pazienti che non sono in grado di partecipare a procedure lunghe (Prieto et al., 2003).

Sebbene nel nostro campione non sia stata riscontrata una relazione significativa tra i test di cognizione sociale e l'età cronologica, così come non sono state riscontrate differenze nelle prestazioni tra adolescenti e adulti, forniamo valori normativi per i due gruppi di età e per il campione totale (Tabella 2 e Tabella A3 in Appendice). I nostri risultati sembrano allinearsi con quanto riportato in studi precedenti sulla capacità di mentalizzazione in età adulta. Nel complesso, i ricercatori non hanno riportato differenze significative tra giovani e adulti, suggerendo cambiamenti nelle dimensioni cognitive e affettive della ToM solo nella tarda età adulta (cioè dopo i 60 anni; Baksh et al., 2018; Henry et al., 2013; Raimo et al., 2022). La cognizione sociale rappresenta un'abilità evolutiva, che si sviluppa lungo un continuum e segue fasi definite (Happé & Frith, 2014; Peterson et al., 2005; Pino et al., 2017; Wellman & Liu, 2004), ed è opinione comune che il comportamento sociale e la capacità di rispondere in modo appropriato al comportamento degli altri si sviluppino intorno ai 3-5 anni di età (Happé & Frith, 2014; Mazza et al., 2017). L'adolescenza, tuttavia, è un periodo cruciale per la formazione

dell'identità e si verificano cambiamenti nelle modalità di interazione sociale (Albiero et al., 2010; Burnett & Blakemore, 2009). Il nostro campione comprende adolescenti, in quanto si tratta di una fascia d'età in cui gli aspetti della cognizione sociale non sono indagati quanto quelli dei bambini e degli adulti, per cui spesso mancano strumenti adeguati per comprendere lo sviluppo di queste abilità in questo periodo cruciale. Pertanto, disporre di valori normativi applicabili nella pratica clinica consente di confrontare le prestazioni di un soggetto con quelle di altri sulla base delle prestazioni medie di un ampio campione (Mazza & Valenti, 2019; Vogel et al., 2020).

I risultati hanno anche mostrato una differenza di genere nelle prestazioni nell'A-ToM, nel subtest Normativo, nel subtest Violazione e nel subtest Gravità, ma di entità trascurabile. Suggeriamo cautela nell'interpretazione di questi risultati, poiché le correlazioni deboli potrebbero essere significative quando la dimensione del campione è sufficientemente grande, come nel caso in questione. In effetti, ogni correlazione significativa è di debole entità e, considerando le dimensioni degli effetti, calcolando il coefficiente di determinazione R^2, la correlazione più alta con il genere ($r = 0,17$) ha dato come risultato $R^2= 0,02$, indicando che il genere rappresenta solo il 2% della varianza. In pratica, suggeriamo di considerare l'associazione ottenuta come trascurabile. In letteratura esistono risultati contrastanti sulle differenze di genere nelle abilità di cognizione sociale (Di Tella et al., 2020): alcuni studi non mostrano differenze (Barrett et al., 1998; Derntl et al., 2010), mentre altri riportano che le femmine ottengono risultati migliori dei maschi in compiti che indagano specificamente le abilità di ToM e di empatia (Baron-Cohen et al., 1997, 1999, 2001). I nostri risultati supportano la letteratura che non indica differenze di genere; tuttavia, vale la pena notare che l'analisi degli item ha indicato vantaggi misti per uomini e donne nei compiti di cognizione sociale.

Sebbene il nostro studio non possa confermare le ipotesi relative al genere, vale la pena considerare che, a nostra conoscenza, altri studi non hanno considerato le differenze di genere in relazione al funzionamento degli item, il che rappresenta un aspetto trascurato e potrebbe spiegare i risultati contrastanti presenti in letteratura. Possiamo solo ipotizzare che il DIF possa contribuire a questo aspetto, almeno in una certa misura. Tuttavia, potremmo considerare questo punto come un punto di forza delle nostre misure riviste, perché consentirebbe un confronto indipendentemente dal sesso dei partecipanti, fornendo a clinici e ricercatori la rimozione di un possibile fattore di confondimento e facilitando lo studio dei fattori legati alla cognizione sociale. Gli studi futuri dovrebbero indagare ulteriormente su come le caratteristiche del compito possano influire sulle prestazioni nelle abilità di cognizione sociale. Per quanto riguarda i livelli di istruzione, riteniamo che l'analisi DIF in base alle classi di età abbia ridotto

la possibile associazione tra istruzione e punteggi; tuttavia, consideriamo questo risultato come un punto di forza delle misure proposte, perché hanno il vantaggio di rendere le prestazioni comparabili tra individui con livelli di istruzione diversi.

Validità clinica delle scale riviste

Per quanto riguarda il nostro secondo obiettivo, ovvero indagare la validità clinica dei tre compiti rivisti, abbiamo confrontato le prestazioni dei due campioni clinici, ASD e SSD, con un gruppo di controllo del campione normativo, appaiato per età, sesso e quoziente intellettivo, prevedendo una prestazione ridotta nei gruppi ASD e SSD. Abbiamo anche valutato la performance diagnostica di queste misure nel discriminare le due condizioni cliniche dal campione non clinico utilizzando le curve ROC.

In generale, i nostri risultati mostrano che i gruppi ASD e SSD presentano prestazioni inferiori rispetto a TD nelle misure di cognizione sociale. Gli individui con ASD e SSD presentano una compromissione delle abilità di ToM (Barneveld et al., 2014; Baron-Cohen et al., 2015; Bishop-Fitzpatrick et al., 2017; Maat et al., 2012; Martinez et al., 2019; Mazza et al., 2022) e la valutazione delle diverse manifestazioni contribuirebbe alla comprensione di come questo dominio possa essere considerato un marcatore di questi disturbi.

Per quanto riguarda i nostri risultati, entrambi i disturbi hanno mostrato punteggi più bassi in A-ToM e SST, ma solo l'ASD ha mostrato punteggi più bassi in EAT rispetto al TD.

Nel SST, è interessante notare che il gruppo ASD, ma non quello SSD, ha ottenuto punteggi simili a quelli dei TD nel sottotest Violazioni, identificando correttamente le azioni che violano le norme sociali comuni; tuttavia, giudicano le violazioni meno gravi rispetto ai TD. Pertanto, gli individui con ASD mostrerebbero un grado inferiore di disapprovazione delle violazioni delle norme sociali. Inoltre, il gruppo ASD ha mostrato punteggi più bassi nel subtest Normative rispetto ai TD. Studi precedenti sulla percezione delle situazioni sociali hanno riportato che gli individui con ASD mostrano una mancanza di consapevolezza delle regole sociali non scritte e una difficoltà nell'identificare le norme sociali (Dewey, 1991; Loveland et al., 2001; Nah & Pooh, 2011). Gli individui con ASD identificano correttamente le violazioni delle regole, in quanto rappresentano un comportamento legato alle regole; tuttavia, nell'identificare la norma sociale fanno riferimento alle proprie esperienze (Dewey, 1991; Heavey et al., 2000; Loveland et al., 2001; Nah & Pooh, 2011).

È interessante anche considerare che, in alcune situazioni, una minore disapprovazione delle violazioni non rappresenta necessariamente un limite. In effetti, potrebbe anche indicare che questi individui potrebbero essere maggiormente in grado di accettare altri che sono

socialmente diversi, con conseguenti potenziali benefici. Riteniamo che questa considerazione possa spingere gli studi futuri verso altri aspetti preziosi legati a questi individui.

Al contrario, il gruppo SSD ha ottenuto punteggi comparabili ai TD nel subtest Normativo, ma ha mostrato punteggi inferiori nei subtest Violazione e Gravità. Ciò suggerisce che, a differenza di ASD, i soggetti SSD sono in grado di identificare la norma sociale, ma non riconoscono il comportamento inappropriato, giudicandolo meno grave. La capacità di giudicare il comportamento come appropriato o meno è una conoscenza sociale e sembra essere dovuta a un processo di apprendimento che avviene in età precoce, intorno ai 3 anni (Blair & Cipollotti, 2000; Grafman, 1994; Grafman et al., 1996; Pegado et al., 2018; Schmidt et al., 2016). Gli individui con SSD, infatti, mostrano un deterioramento delle abilità cognitive sociali a seguito del primo episodio psicotico (Bliksted et al. 2017; Frith, 1992, 2004; Mazza et al., 2010; Mazza et al., 2022; Pino et al., 2017, 2018), e la caduta di questa dimensione è stato spiegato con difficoltà nel ricordare comportamenti tipici negli script sociali (Matsui et al., 2006, 2009).

L'analisi ROC sottolinea le differenze tra campioni clinici e controlli nell'esecuzione dei tre compiti di cognizione sociale in studio. In particolare, confrontando ASD e TD, i punteggi A-ToM e SST sono il miglior predittore diagnostico. Come delineato dal modello logistico, la combinazione dei punteggi dei subtest A-ToM, Normativo e Gravità delle SST ha predetto in modo significativo la diagnosi di ASD. Studi precedenti (Pino et al., 2018, 2020) hanno già dimostrato che i bambini con ASD hanno prestazioni peggiori rispetto ai bambini TD in un compito di elaborazione di situazioni sociali (Social Information Processing Interview; SIPI; Ziv & Sorongon, 2011; Ziv et al., 2014) e lo strumento ha mostrato una valida accuratezza nel distinguere tra bambini ASD e TD.

Per quanto riguarda le misure che differenziano i TD dagli SSD, l'analisi ROC indica un’accuratezza per i punteggi dei subtest Violazione e Gravità delle SST, mostrando la più alta performance diagnostica. Ciò indica che la capacità di riconoscere quando una norma sociale viene violata rappresenta la misura migliore per differenziare TD e SSD durante la valutazione clinica. Infatti, l'SST richiede la capacità di contestualizzare un comportamento, di distinguere tra comportamenti regolatori e violazioni e di giudicare la normalità o la stranezza di un comportamento sulla base di regole sociali (Prior et al., 2003). I nostri risultati sono in accordo con molti studi che suggeriscono che gli individui ad alto funzionamento con ASD sono in grado di rilevare violazioni delle regole sociali (Baez et al., 2012; Zalla et al., 2011), probabilmente grazie a un apprendimento esplicito delle regole sociali che funge da strategia compensatoria per i deficit di cognizione sociale. Al contrario, gli individui con SSD mostrerebbero una compromissione nell'individuazione delle violazioni a causa di una

difficoltà nell'elaborazione del contesto sociale e nell'identificazione delle informazioni sociali (Champagne-Lavau & Charest, 2015; Corrigan & Penn, 2001; Penn et al., 2002).

Nonostante i risultati promettenti di questo studio, ci aspettavamo una migliore performance diagnostica per la misura ToM. Anche se i risultati indicano che entrambi i gruppi presentano performance atipiche sull' A-ToM, mostrando punteggi inferiori per gli aspetti cognitivi della ToM rispetto agli individui con sviluppo tipico, i risultati ROC indicavano il potenziale utilizzo del compito A-ToM nella pratica clinica insieme a misure gold standard per supportare il medico nell'identificazione di ASD ma non SSD, data la sua bassa specificità (38%). Tuttavia, potremmo supporre che considerare aspetti più qualitativi della misura potrebbe anche fornire informazioni significative per migliorare la sua capacità discriminatoria; ad esempio, uno studio precedente ha dimostrato che queste condizioni sono caratterizzate da stili di risposta specifici nell'attribuzione di stati mentali (Mazza et al., 2022). Nello specifico, mentre gli individui con ASD non identificano l'intenzione dei protagonisti, quelli con SSD mostrano grandi difficoltà nel comprendere lo scenario sociale (Mazza et al., 2022). Ciò consente la raccolta di informazioni preziose durante il processo diagnostico e nella differenziazione delle condizioni cliniche.

Per quanto riguarda l'EAT, gli individui con ASD hanno mostrato prestazioni atipiche, mentre quelli con SSD hanno avuto prestazioni simili ai controlli, contrariamente alle nostre aspettative. Di conseguenza, considereremmo l'EAT come uno strumento adatto per coloro che sono sospettati di ASD, ma non per SSD, perché questo risultato ha indicato che, per il gruppo SSD, c'è ancora bisogno di fornire una misura più dettagliata della ToM affettiva nel contesto clinico. Le prestazioni diagnostiche dell'EAT confermano questa considerazione; in effetti, non ha mostrato un'accuratezza accettabile nel differenziare tra SSD e controlli e suggeriamo cautela nell'uso dell'EAT per l'identificazione di SSD. Come prospettive future uno degli obiettivi è di superare tale limitazione attuale. Quindi, i dati forniti dal nostro lavoro suggeriscono valori di cut-off per i tre compiti di cognizione sociale considerati nello studio e consentono agli esaminatori di utilizzare questi test con valori normativi.

Considerando la diagnosi di ASD, suggeriremmo rigorosamente di combinare il punteggio A-ToM con il punteggio Normativo e il punteggio Violazione dell'SST, come indicato dalla formula fornita, perché questa soluzione raggiunge la massima accuratezza nel differenziare tra ASD e controlli. Suggeriamo inoltre di somministrare l'EAT per ottenere altre informazioni che potrebbero aiutare il clinico nel processo diagnostico, perché differenzia tra ASD e controlli ma con un'accuratezza inferiore. Tuttavia, sottolineiamo che il punteggio combinato dovrebbe essere il principale strumento diagnostico per una possibile diagnosi di ASD a supporto delle

misure gold standard. Infatti, anche se il nostro campione consente solo una valutazione preliminare, i risultati sono incoraggianti rispetto ad altre misure testate su adulti senza disabilità intellettiva, come l'algoritmo rivisto e l'Autism Diagnostic Interview-Revised (ADI-R) (Fusar-Poli et al., 2017). Come possibile prospettiva futura, vorremmo confrontare direttamente il punteggio combinato con l'ADI-R per confermare questo possibile risultato. Tuttavia, dobbiamo notare che, rispetto ai gold standard, i nostri compiti revisionati presentano alcuni vantaggi: ad esempio, non richiedono a un individuo di ricordare condizioni passate e non richiedono il coinvolgimento attivo del clinico nelle attività e nell'osservazione dei comportamenti. Nella ricerca futura, sarebbe utile integrare l'EAT nel punteggio complessivo e valutare i vantaggi della procedura. Vorremmo anche notare che, in base agli obiettivi della valutazione, il clinico o il ricercatore potrebbero somministrare tutti i compiti proposti o solo alcuni specifici.

Per quanto riguarda la valutazione clinica SSD, suggeriremmo di considerare il punteggio Violazione e Gravità di SST come i criteri principali per la diagnosi di SSD, perché hanno raggiunto la migliore performance diagnostica. Inoltre, ci si aspetterebbe una convergenza di queste misure poiché la loro performance diagnostica è strettamente simile. Suggeriamo di utilizzare l'A-ToM e il punteggio Normativo con questa popolazione, insieme ai punteggi Violazione e Gravità, perché, anche se limitati, potrebbero differenziare tra controlli e pazienti e possono fornire al medico informazioni utili per la procedura clinica e gli interventi.

Nonostante queste considerazioni, dovremmo notare che le misure proposte dovrebbero essere utilizzate insieme ai gold standard. Sebbene il nostro studio indichi il potenziale utilizzo di queste misure nel contesto clinico, riteniamo che potrebbero ancora essere migliorate in termini di accuratezza. Questo è un possibile obiettivo per studi futuri. Nonostante gli interessanti risultati, siamo consapevoli che il presente studio ha dei limiti. Le relazioni tra le dimensioni della cognizione sociale devono ancora essere studiate e pienamente comprese; ciò è particolarmente rilevante per l'ottimizzazione diagnostica. La dimensione del campione delle popolazioni cliniche era limitata e studi futuri dovrebbero considerare le prestazioni di questi test in campioni clinici più ampi. I valori di cut-off riportati dovrebbero essere considerati con cautela, poiché suggeriscono il potenziale utilizzo nella pratica clinica in un quadro esplorativo, a causa delle dimensioni limitate del campione dei gruppi clinici. Studi futuri dovrebbero fornire valori di cut-off considerando campioni più ampi. Da una prospettiva futura, sarebbe interessante confrontare le prestazioni tra gruppi clinici senza e con condizioni cliniche concomitanti per valutare possibili effetti dovuti ad altre condizioni o difficoltà. ASD e SSD rappresentano infatti due condizioni cliniche con probabili comorbilità associate, alcune delle

quali potrebbero ulteriormente influenzare la prestazione in questi test. Tuttavia, riteniamo che i cut-off proposti potrebbero essere utili anche in presenza di comorbilità, in quanto è molto probabile che i punteggi ottenuti da questi pazienti rientrino in un intervallo di interesse clinico. Inoltre, è importante sottolineare che il medico dovrebbe essere in grado di selezionare gli strumenti appropriati alle caratteristiche dell'individuo, tenendo conto di eventuali condizioni cliniche e mediche concomitanti che, in alcuni casi, potrebbero rendere difficile, se non improbabile, la somministrazione di tali test.

Abbiamo deciso di utilizzare le SPM, poiché è un test facile da somministrare e ci consente di non affaticare i partecipanti e di escludere i partecipanti con significativo deterioramento cognitivo, poiché, negli individui con schizofrenia, gli errori nella cognizione sociale sono correlati ai loro livelli di QI (Sahl et al., 2022). Inoltre, per quanto riguarda il gruppo ASD, i partecipanti sono stati classificati come Livello 1 (alto funzionamento) ed è stato somministrato il modulo 4 ADOS-2, che richiede buone capacità linguistiche espressive e recettive. Tuttavia, non rappresenta una misura del QI verbale, ma ha alcune associazioni con esso (Kluever et al., 1995; McLaurin & Farrar, 1973), quindi ciò rappresenta comunque una limitazione da tenere in considerazione.

Per quanto riguarda il nostro campione di individui sani, non abbiamo incluso partecipanti di mezza età più anziani (> 50 anni), il che rappresenta un'altra limitazione del presente studio. In letteratura, i ricercatori hanno studiato le capacità di mentalizzazione nel corso dell'età adulta e hanno riportato risultati contrastanti, tra cui cambiamenti selettivi nelle dimensioni della ToM negli anziani (vale a dire circa 60 anni) (Baksh et al., 2018; Henry et al., 2013; Raimo et al., 2022). Fornire dati normativi per la popolazione anziana dovrebbe richiedere un'attenta considerazione; infatti, sarebbero necessari studi aggiuntivi che includano campioni di anziani e che prendano in considerazione la possibile influenza dei profili neuropsicologici sul funzionamento della ToM nell'invecchiamento (Raimo et al., 2022). Inoltre, nel campione 20-50 c'erano pochi partecipanti di età pari o superiore a 30 anni, il che potrebbe spiegare l'assenza di una correlazione significativa tra compiti di cognizione sociale ed età; tuttavia, considerando gli studi sui cambiamenti nelle prestazioni di cognizione sociale durante l'età adulta (Cho et al., 2019; Love et al., 2015; Raimo et al., 2022), non dovremmo aspettarci cambiamenti legati all'età nel campione di adulti considerato. Anche se il nostro campione di individui con sviluppo tipico è ragionevolmente ampio e proveniente da contesti diversi, i nostri partecipanti non sono stati selezionati in modo casuale, il che rappresenta una limitazione dello studio attuale; i lavori futuri dovrebbero superare questa limitazione considerando anche altri fattori demografici, come il background occupazionale e l'etnia, tra gli altri.

4.6 Conclusioni

In conclusione, abbiamo dimostrato che le misure di cognizione sociale proposte possono migliorare la pratica clinica e i nostri risultati incoraggiano il loro utilizzo come supporto per altri strumenti ampiamente utilizzati nel processo diagnostico. Ciò aumenterebbe l'accuratezza nel distinguere condizioni che condividono caratteristiche cliniche, l'identificazione di interventi mirati e la valutazione della loro efficacia.

Conclusioni generali

L'ASD rappresenta una condizione del neurosviluppo (APA,2013) con un impatto *longlife* sull'individuo, la cui ampia eterogeneità nella manifestazione clinica rappresenta una sfida per la valutazione diagnostica e nella strutturazione di interventi. Nel corso degli anni sono stati proposti diversi strumenti di valutazione con l'obiettivo di caratterizzare la sintomatologia ASD, tuttavia molti degli strumenti non sono applicabili alla pratica clinica, sia in termini di complessità nella somministrazione ma anche di mancanza di una buona qualità psicometrica. Lo sviluppo e la disponibilità di strumenti validi e affidabili, hanno un impatto sull'individuo, sia per quanto riguarda nel fornire un quadro clinico accurato ma anche nella strutturazione e valutazione di interventi mirati.

La pratica clinica dell'ASD ne fornisce un esempio rappresentativo, in particolare quando si parla di diagnosi e identificazione precoce di bambini con o a rischio ASD, mediante l'utilizzo dei cosiddetti "gold standard" per la diagnosi. Allo stato attuale, l'ADOS-2 è il test più utilizzato nell'identificare i comportamenti relati alla condizione ASD, ma anche per permettere l'avvio degli interventi riabilitativi. Tuttavia, la diagnosi di ASD non può essere ricondotta ad un singolo test; per la sua natura multifattoriale, per arrivare ad una diagnosi precisa, soprattutto quando giungono ai servizi adolescenti e adulti, è necessario affiancare ai test "gold-standard" misure valide e affidabili che contribuiscono a identificare i deficit *core*, come le abilità di SC. Tuttavia, il contesto clinico italiano appare ancora poco operativo nell'integrare gli strumenti di SC nell'iter diagnostico e valutativo per l'autismo, dovuto anche a una limitata disponibilità di proprietà psicometriche.

Il presente lavoro di tesi ha avuto, dunque, l'obiettivo di evidenziare l'importanza degli approcci interdisciplinari, come l'utilizzo di metodologie statistiche e psicometriche solide nell'ambito della psicologia clinica, in particolare per l'ASD. Con tale approccio abbiamo indagato l'affidabilità e la stabilità della diagnosi di ASD in età molto precoce. Una diagnosi accurata e certa consente di avviare interventi tempestivi che possono portare a un miglioramento dei risultati e della qualità della vita, nonché a una riduzione dei costi sociali a lungo termine (Benedetto et al., 2021; Chasson et al., 2007). È il caso del Modulo Toddler, utilizzato per stimare il livello di rischio di ricevere una diagnosi di autismo in bambini dai 12 mesi. Allo stato attuale, non ci sono studi riguardo l'accordo diagnostico tra il Modulo Toddler e i restanti Moduli, con cui, invece, la diagnosi di ASD viene posta; e rimangono limitati gli studi che ne indagano il suo utilizzo nella pratica clinica. I risultati dello Studio 1 hanno evidenziato l'esistenza di una concordanza nella conferma diagnostica tra il Modulo Toddler e il Modulo 1, ma non vi è accordo sulla gravità diagnostica, evidenziando come il Modulo

Toddler, “sovrastimi” la severità sintomatologica. Il risultato ottenuto evidenzia la variabilità dello sviluppo in bambini molto piccoli, che può portare all’incertezza diagnostica e di conseguenza alla riluttanza di alcuni clinici nel porre una diagnosi certa prima dei 3 anni di età. Tuttavia, la presenza di un accordo tra i due moduli sulla conferma diagnostica, piuttosto che sulla gravità dei sintomi, a partire da circa 12 mesi di età, potrebbe supportare la pratica clinica consentendo l'inclusione precoce di bambini a rischio in interventi riabilitativi per potenziare le aree di sviluppo e migliorare gli esiti sintomatici. Inoltre, è stato osservato come i due costrutti diagnostici AS e CRR, predicano in maniera diversa e in tempi diversi, la diagnosi di autismo, permettendo la strutturazione degli interventi che siano mirati e adattabili alle fasi di sviluppo. Sempre sulla stessa scia, abbiamo tentato di replicare la struttura sottostante del Modulo Toddler con l’applicazione di nuove metodiche, l’EGA. Analizzando i modelli di connessione tra gli item, possiamo valutare la consistenza interna del Modulo Toddler e identificare gli item che potrebbero non essere in linea con la struttura generale del test. I nostri risultati non hanno replicato la soluzione a due fattori dell'ADOS-2 (Lord et al. 2012), ma suggeriscono una soluzione a 3 dimensioni (SI, CC e CSS), in linea con il modello triadico presente nella prima versione dell’ADOS. Inoltre, anche gli item all’interno delle dimensioni vengono aggregati in maniera diversa. Per passare dall'analisi del costrutto all'applicazione dei compiti strutturati dello strumento, abbiamo confrontato l'aggregazione degli item all'interno delle dimensioni con i compiti del Modulo Toddler. Risulta che quattro compiti elicitano le caratteristiche principali rappresentate dalle tre dimensioni (IS, CC e CSS): “Gioco libero”, “Gioco con le bolle”, “Anticipazione di una routine con oggetti” e “Bagnetto”. Ciò suggerirebbe una struttura di compiti “principali” e compiti “opzionali/complementari” per la somministrazione dei compiti, rendendo la valutazione più adattiva in termini di tempo di somministrazione e nell’ottimizzazione della raccolta delle informazioni, minimizzando la necessità di una cooperazione prolungata da parte di un bambino.

Abbiamo, inoltre, dimostrato come sia possibile l’introduzione di misure di SC nel percorso valutativo per l’ASD, sia in persone con un limitato repertorio linguistico e di funzionamento cognitivo, ma anche nel differenziare condizioni cliniche che presentano una sovrapposizione sintomatologica, come ASD e SSD. Con l’utilizzo del metodo EGA, è stato costruito un nuovo strumento non verbale, l’IAC-CST per la valutazione della capacità di attribuire le intenzioni (pre-requisito della ToM), che ha dimostrato il suo potenziale utilizzo nella pratica clinica. Il test proposto ha il vantaggio di essere di facile somministrazione, in termini di tempo data la sua breve lunghezza (14 item in totale) e la possibilità di somministrarlo a persone con scarse capacità verbali. Ne è stata valutata anche la validità clinica confrontando le prestazioni del

gruppo ASD con quelle dei controlli con QI corrispondente. Il confronto tra gruppi ha indicato che gli individui ASD non mostrano alcuna difficoltà nella condizione di controllo rispetto al gruppo appaiato per QI. Al contrario, il gruppo ASD mostra una prestazione significativamente inferiore nelle due serie che indagano l'attribuzione di intenzione. Un'ulteriore conferma viene dall'analisi ROC, che sostiene l'efficacia dell'IA-CST nel discriminare, con un discreto livello di accuratezza, tra il gruppo ASD e quello TD.

Inoltre, come ulteriore obiettivo, abbiamo revisionato, indagato le proprietà psicometriche e fornito valori normativi per tre misure di SC utilizzati ampiamente in letteratura, A-ToM, EAT e SST, utilizzando il modello IRT. Delle tre misure revisionate ne abbiamo indagato anche la validità clinica e la performance diagnostica nelle due popolazioni ASD e SSD, utilizzando la curva ROC, che ne evidenzia le differenze tra campioni clinici e controlli nell'esecuzione dei tre compiti di cognizione sociale in studio. In particolare, confrontando ASD e TD, i punteggi A-ToM e SST sono il miglior predittore diagnostico, mentre tra TD e SSD la performance diagnostica migliore è data dal compito SST.

I risultati ottenuti hanno implicazioni importanti nella pratica clinica. Non solo permettono di contribuire alla conoscenza e alla caratterizzazione delle manifestazioni cliniche dell'ASD, ma evidenziano come il processo valutativo nell'ASD diviene sempre più accurato nell'utilizzo di strumenti che siano adattabili alla persona, ma anche con l'integrazione di misure che contribuiscano ad indagare i deficit principali. A tal proposito i risultati ottenuti contribuirebbero all'applicazione nei setting clinici di misure di SC con buone caratteristiche psicometriche e valori normativi. Inoltre, l'integrazione di misure forniscono supporto nella strutturazione trattamenti individualizzati, volti al potenziamento delle abilità che risultano compromesse e, attraverso valutazioni di follow-up, monitorarne l'andamento e gli esiti.

In conclusione obiettivo della tesi è stato quello di evidenziare la necessità dell'utilizzo di metodologie statistiche e psicometriche nell'ambito della psicologia clinica, ma soprattutto di sottolineare l'importanza del legame che c'è tra ricerca e pratica clinica, e viceversa.

Appendice

Studio 1

Tabella A1. Corrispondenza delle prove strutturate del modulo Toddler con i rispettivi item e dimensioni latenti emerse dal modello EGA.

Task	ITEM	DIMENSION
1. GIOCO LIBERO **1a. Gioco libero- palla**	A2. Frequenza delle vocalizzazioni spontanee rivolte agli altri A3. Intonazione di vocalizzi e verbalizzazioni B1. Contatto oculare insolito B4. Espressioni facciali dirette agli altri B6. Divertimento condiviso nell'interazione B13. Inizio spontaneo dell'attenzione congiunta B12. Mostrare B15. Qualità delle aperture sociali D1. Interesse sensoriale insolito per materiali/persone D2. Movimenti e postura di mani e dita D5. Interessi insoliti e ripetitivi o comportamenti stereotipati	IS CC CSS
2. Bloccare l'accesso al gioco	A2. Frequenza delle vocalizzazioni spontanee rivolte agli altri B1. Contatto oculare insolito B4. Espressioni facciali dirette agli altri	IS CC
3. Risposta al nome	A2. Frequenza delle vocalizzazioni spontanee rivolte agli altri B1. Contatto oculare insolito B4. Espressioni facciali dirette agli altri	IS CC
4. Bolle di sapone **4a. Fare uno scherzo**	A2. Frequenza delle vocalizzazioni spontanee rivolte agli altri A8. Gesti B1. Contatto oculare insolito B4. Espressioni facciali dirette agli altri B5 Integrazione dello sguardo e di altri comportamenti durante le aperture sociali B6. Divertimento condiviso nell'interazione B13. Inizio spontaneo dell'attenzione congiunta B15. Qualità delle aperture sociali D1. Interesse sensoriale insolito per materiali/persone D2. Movimenti e postura di mani e dita D5. Interessi insoliti e ripetitivi o comportamenti stereotipati	IS CC CSS
5. Anticipazione di routine con oggetti **5a. Non si gioca più**	A2. Frequenza delle vocalizzazioni spontanee rivolte agli altri A8. Gesti B1. Contatto oculare insolito B4. Espressioni facciali dirette agli altri B5 Integrazione dello sguardo e di altri comportamenti durante le aperture sociali B6. Divertimento condiviso nell'interazione B13. Inizio spontaneo dell'attenzione congiunta	IS CC CSS

	B14 Risposta all'attenzione congiunta B15. Qualità delle aperture sociali D1. Interesse sensoriale insolito per materiali/persone D2. Movimenti e postura di mani e dita D5. Interessi insoliti e ripetitivi o comportamenti stereotipati	
6. Anticipazione di routine sociali	A2. Frequenza delle vocalizzazioni spontanee rivolte agli altri A8. Gesti B1. Contatto oculare insolito B4. Espressioni facciali dirette agli altri B5 Integrazione dello sguardo e di altri comportamenti durante le aperture sociali B6. Divertimento condiviso nell'interazione B13. Inizio spontaneo dell'attenzione congiunta B15. Qualità delle aperture sociali	IS CC
7. Risposta all'attenzione congiunta	B1. Contatto oculare insolito B14 Risposta all'attenzione congiunta	IS CC
8. Eisposta al sorriso sociale	A2. Frequenza delle vocalizzazioni spontanee rivolte agli altri B1. Contatto oculare insolito B4. Espressioni facciali dirette agli altri B5 Integrazione dello sguardo e di altri comportamenti durante le aperture sociali B6. Divertimento condiviso nell'interazione	IS CC
9. Bagnetto **9a. Ignorare**	A2. Frequenza delle vocalizzazioni spontanee rivolte agli altri B1. Contatto oculare insolito B4. Espressioni facciali dirette agli altri B6. Divertimento condiviso nell'interazione B13. Inizio spontaneo dell'attenzione congiunta B15. Qualità delle aperture sociali D2. Movimenti e postura di mani e dita D5. Interessi insoliti e ripetitivi o comportamenti stereotipati	IS CC CSS
10. Imitazione Funzionale e simbolica	B6. Divertimento condiviso nell'interazione	CC
11. Snack	A2. Frequenza delle vocalizzazioni spontanee rivolte agli altri A8. Gesti B1. Contatto oculare insolito B4. Espressioni facciali dirette agli altri B5. Integrazione dello sguardo e di altri comportamenti durante le aperture sociali B15. Qualità delle aperture sociali	CC

Nota. Gli item che sono risultati instabili attraverso l'analisi EGA e sono stati eliminati dalle dimensioni sono indicati in rosso; IS: Intenzionalità sociale; CC: Comportamenti comunicativi; CSS: Comportamenti sensoriali e stereotipati.

Studio 4

Tabella A2. Indici degli item ottenuti dal modello di Rasch per le versioni riviste di A-ToM, EAT e SST.

Item	Infit	Outfit	Discrimination
A-ToM			
ToM 1	0.89	0.87	0.33
ToM 4	0.98	0.91	0.07
ToM 5	0.94	0.88	0.24
ToM 6	0.80	0.60	0.54
ToM 7	0.93	0.87	0.21
ToM 8	1.00	1.04	-0.02
ToM 9	0.98	0.98	0.03
ToM 10	1.05	1.04	-0.22
ToM 11	1.03	1.01	-0.12
ToM13	0.85	0.82	0.43
EAT			
SADNESS			
Item 1	0.97	0.92	0.21
Item 4	0.96	0.96	0.33
Item 6	1.00	0.99	0.18
Item 19	0.97	1.01	0.25
Item 31	0.97	0.91	0.23
Item 36	1.03	1.07	0.15
FEAR			
Item 2	1.00	0.99	0.18
Item 7	0.93	0.76	0.26
Item 11	0.98	0.96	0.16
Item 26	0.94	0.92	0.20
Item 29	0.98	0.98	0.23
Item 43	0.94	1.02	0.18
EMBARRASSMENT			
item 3	1.09	1.11	0.02
item 14	1.02	1.05	0.12
item 24	0.95	0.84	0.27
item 28	0.97	0.96	0.25
item 58	0.95	0.93	0.26
HAPPINESS			
Item 5	0.95	0.86	0.18
Item 20	1.00	0.97	0.22

Item 32	0.88	0.94	0.30
Item 42	0.98	0.94	0.17
DISGUST			
Item 27	1.06	1.10	0.06
Item 34	0.96	1.09	0.17
ENVY			
Item 39	0.99	1.03	0.19
ANGER			
Item 45	1.00	1.03	0.20
Item 49	0.99	0.99	0.22
Item 53	0.96	0.90	0.26
Item 56	1.01	1.01	0.16
SST			
NORMATIVE			
Item 4	1.05	1.05	- 0.79
Item 7a	0.98	1.06	-0.32
Item 15	0.96	1.01	0.12
Item 17a	1.02	1.15	- 0.65
Item 23a	1.09	1.13	- 0.80
VIOLATION			
Item 2b	0.93	0.86	0.28
Item 3a	1.03	1.06	- 0.05
Item 3c	0.93	0.85	0.65
Item 5b	0.95	0.90	0.37
Item 5c	0.93	0.78	0.53
Item 6	0.94	0.87	0.23
Item 13	0.89	0.79	0.67
Item 19	1.03	1.19	- 0.41

Tabella A3. Dati normativi della versione rivista dell'A-ToM, dell'EAT e dell'SST ottenuti attraverso l'analisi di Rasch per i due gruppi di età e per il campione complessivo.

	N Item	Mean (SD)	95 % Confidence interval	
			Lower Bound	Upper Bound
A-ToM	10			
14 - 19		8.46 (1.42)	8.29	9.64
20 - 50		8.79 (1.29)	8.64	8.93
Full sample		8.64 (1.36)	8.53	8.75
Emotion Attribution Task	28			
14 - 19		19.36(2.99)	18.92	19.79
20 - 50		18.85(6.16)	18.17	19.53
Full Sample		19.03(5.22)	18.58	19.49
Sadness	6			
14 - 19		4.22(1.19)	4.04	4.39
20 - 50		4.23(1.73)	4.04	4.42
Full Sample		4.22(1.55)	4.09	4.36
Fear	6			
14 - 19		4.67(1.02)	4.52	4.81
20 - 50		4.43(1.58)	4.26	4.60
Full Sample		4.52(1.40)	4.39	4.64
Embarrassment	5			
14 - 19		2.68(1.22)	2.50	2.86
20 - 50		2.59(1.40)	2.44	2.75
Full Sample		2.63(1.34)	2.51	2.74
Happiness	4			
14 -19		3.23(0,91)	3.10	3.37
20 - 50		3.04(1,22)	2.90	3.17
Full Sample		3.11(1,12)	3.01	3.21
Disgust	2			
14 - 19		1.62(0,60)	1.53	1.70
20 - 50		1.45(0,72)	1.37	1.53
Full Sample		1.51(0,68)	1.45	1.57
Envy	1			
14 -19		0.67(0.46)	0.60	0.74
20 - 50		0.65(0.47)	0.60	0.70
Full Sample		0.66(0.47)	0.62	0.70
Anger	4			
14 -19		2.24(1.03)	2.09	2.39
20 - 50		2.42(1.23)	2.29	2.56
Full Sample		2.36(1.16)	2.26	2.46
Social Situation Task				

Normative	5			
14 -19		4.09(0.93)	3.95	4.23
20 - 50		4.01(1.04)	3.84	4.18
Full Sample		4.05(0.98)	3.95	4.16
Violation	8			
14 -19		7.13(1.16)	6.96	7.30
20 - 50		7.00(1.03)	6.83	7.17
Full Sample		7.07(1.10)	6.95	7.19
Gravity	8			
14 -19		14.13(3.71)	13.57	14.68
20 - 50		13.87(3.48)	13.31	14.44
Full Sample		14.01(3.60)	13.62	14.41

Bibliografia

Albiero, P., Matricardi, G., & Toso, D. (2010). La basic empathy scale, uno strumento per la misura della responsività empatica negli adolescenti: Un contributo alla validazione Italiana. Psicologia clinica dello sviluppo, 14(1), 205–218.

Alcalá-López, D., Vogeley, K., Binkofski, F., & Bzdok, D. (2019). Building blocks of social cognition: Mirror, mentalize, share?. *Cortex*, 118, 4-18.

Allen, D. A. (1988). Austistic spectrum disorders: clinical presentation in preschool children. *Journal of child neurology*, 3(1_suppl), S48-S56.

Altschuler, E. L., Vankov, A., Hubbard, E. M., Roberts, E., Ramachandran, V. S., & Pineda, J. A. (2000). Mu wave blocking by observation of movement and its possible use as a tool to study theory of other minds. In *Soc. Neurosci* (Vol. 68, No. 1).

American Psychiatric Association (1980): Diagnostic and Statistical Manual of Mental Disorders, Washington, DC, American Psychiatric Assossation (3rd ed.). Trad. it. DSM-III: manuale diagnostico e statistico dei disturbi mentali, Milano, Masson. 1983.

American Psychiatric Association (1987): Diagnostic and Statistical Manual of Mental Disorders, Washington, DC, American Psychiatric Assossation (3rd ed., rev.). Trad. it. DSM-III-R: manuale diagnostico e statistico dei disturbi mentali, Milano, Masson. 1991.

American Psychiatric Association (1994): Diagnostic and Statistical Manual of Mental Disorders, Washington, DC, American Psychiatric Assossation (4th ed.). Trad. it. DSM-IV: manuale diagnostico e statistico dei disturbi mentali, Milano, Masson. 1995.

American Psychiatric Association (2000): Diagnostic and Statistical Manual of Mental Disorders, Washington, DC, American Psychiatric Assossation (4th ed., text rev.). Trad. it. DSM-IV-TR: manuale diagnostico e statistico dei disturbi mentali, text revision, Milano, Masson. 2005.

American Psychiatric Association (2013): Diagnostic and Statistical Manual of Mental Disorders, Washington, DC, American Psychiatric Assossation (5th ed.). Trad. it. DSM5: manualeiagnostico e statistico dei disturbi mentali, Milano, Raffaello Cortina Editore, 2014.

American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.), Upper Saddle River, Prentice Hall.

Anunciacao, L. (2018). An overview of the history and methodological aspects of psychometrics: History and methodological aspects of psychometrics. *Journal for ReAttach Therapy and Developmental Diversities*, *1*(1), 44-58.

Apperly, I. A. (2008). Beyond simulation–theory and theory–theory: why social cognitive neuroscience should use its own concepts to study "Theory of Mind". *Cognition*, 107(1), 266-283.

Asperger, H. (1944). Die "autistichen Psychopathen" im Kindersalter. *Archive fur psychiatrie und Nervenkrankheiten*, 117, 76–136.

Baddeley, A., & Wilson, B. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition*, 7(2), 212–230.

Baez, S., Rattazzi, A., Gonzalez-Gadea, M. L., Torralva, T., Vigliecca, N., Decety, J., . . .Ibanez, A. (2012). Integrating intention and context: Assessing social cognition in adults with Asperger syndrome. *Frontiers in Human Neuroscience*, 6, Article 302.

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., & Rutter, M. (1995). Autism as a strongly genetic disorder: evidence from a British twin study. *Psychological medicine*, 25(1), 63-77.

Baio, J. (2018). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2014. MMWR. Surveillance Summaries, 67.

Baksh, R. A., Abrahams, S., Bertlich, M., Cameron, R., Jany, S., Dorrian, T., ... & Auyeung, B. (2021). Social cognition in adults with autism spectrum disorders: Validation of the Edinburgh Social Cognition Test (ESCoT). The Clinical Neuropsychologist, 35(7), 1275-1293.

Barbaranelli, C., & Natali, E. (2005). *I test Psicologici: teorie e modelli psicometrici*. Roma: Carocci Editore

Barbaresi, W., Cacia, J., Friedman, S., Fussell, J., Hansen, R., Hofer, J., Roizen, N., Stein, R. E. K., Vanderbilt, D., & Sideridis, G. (2022). Clinician Diagnostic Certainty and the Role of the Autism Diagnostic Observation Schedule in Autism Spectrum Disorder Diagnosis in Young Children. *JAMA Pediatrics*, 176(12), 1233.https://doi.org/10.1001/jamapediatrics.2022.3605

Barbaro, J., & Dissanayake, C. (2017). Diagnostic stability of autism spectrum disorder in toddlers prospectively identified in a community-based setting: Behavioural characteristics and predictors of change over time. *Autism*, 21(7), 830-840.

Barneveld, P. S., Swaab, H., Fagel, S., van Engeland, H., & de Sonneville, L. M. (2014). Quality of life: A case-controlled long-term follow-up study, comparing young high-functioning adults with autism spectrum disorders with adults with other psychiatric disorders diagnosed in childhood. *Comprehensive Psychiatry*, 55(2), 302–310.

Baron-Cohen, S. (1991). Precursors to a theory of mind: Understanding attention in others. *Natural theories of mind: Evolution, development and simulation of everyday mindreading*, 1, 233-251

Baron-Cohen, S. (2004). The cognitive neuroscience of autism. *Journal of Neurology, Neurosurgery & Psychiatry*, 75(7), 945-948.

Baron-Cohen, S., Bowen, D. C., Holt, R. J., Allison, C., Auyeung, B., Lombardo, M. V., Smith, P., & Lai, M. C. (2015). The "reading the mind in the eyes" test: Complete absence of typical sex difference in~ 400 men and women with autism. *PLOS ONE*, 10(8), Article e0136521.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry*, 38(7), 813–822.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition* 21, 37–46.

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux-pas by normally develop- ing children and children with Asperger syndrome or high functioning autism. *Journal of Autism and Developmental Disorders*, 29(5), 407–418.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001b). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of autism and developmental disorders*, *31*(1), 5-17.

Baron-Cohen, S. (1989). Perceptual role taking and protodeclarative pointing in autism. *British Journal of developmental psychology*, 7(2), 113-127.

Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J., & Walker, J. (1995). Are children with autism blind to the mentalistic significance of the eyes?. *British Journal of Developmental Psychology*, 13(4), 379-398.

Barrett, L. F., Robin, L., Pietromonaco, P. R., & Eyssell, K. M. (1998). Are women the "more emotional" sex? Evidence from emotional experiences in social context. Cognition & Emotion, 12(4), 555–578.

Barrett, P. (2003). Beyond psychometrics: Measurement, non-quantitative structure, and applied numerics. *Journal of Managerial Psychology*, *18*(5), 421-439.

Bas-Hoogendam, J. M., van Steenbergen, H., Kreuk, T., Van der Wee, N. J., & Westenberg, P. M. (2017). How embarrassing! The behavioral and neural correlates of processing social norm violations. PLOS ONE, 12(4), Article e0176326.

Bastiaansen, J. A., Thioux, M., Nanetti, L., van der Gaag, C., Ketelaars, C., Minderaa, R., & Keysers, C. (2011). Age-related increase in inferior frontal gyrus activity and social functioning in autism spectrum disorder. *Biological psychiatry*, *69*(9), 832-838.

Baylor, C., Hula, W., Donovan, N. J., Doyle, P. J., Kendall, D., & Yorkston, K. (2011). An introduction to item response theory and Rasch models for speech-language pathologists.

Beaudoin, C., & Beauchamp, M. H. (2020). Social cognition. In Handbook of clinical neurology (Vol. 173, pp. 255-264). Elsevier.

Bechi, M., Agostoni, G., Buonocore, M., Gritti, D., Mascia, M., Spangaro, M., Bianchi, L., Cocchi, F., Guglielmino, C., Bosia, M., & Cavallaro, R. (2019). The association of autistic traits with Theory of Mind and its training efficacy in patients with schizophrenia. *Schizophrenia research. Cognition*, *19*, 100164.

Benedetto, L., Cucinotta, F., Maggio, R., Germanò, E., De Raco, R., Alquino, A., ... & Gagliano, A. (2021). One-year follow-up diagnostic stability of autism spectrum disorder diagnosis in a clinical sample of children and toddlers. *Brain sciences*, 11(1), 37.

Bennetto, L., Pennington, B. F., & Rogers, S. J. (1996). Intact and impaired memory functions in autism. *Child Development*, 67(4), 1816–1835.

Bermúdez, J.L. (Ed.), (2006). *Philosophy of Psychology: Contemporary Readings*. Routledge

Bertelli, M. O., Merli, M. P., Bradley, E., Keller, R., Varrucciu, N., Del Furia, C., & Panocchia, N. (2015). The diagnostic boundary between autism spectrum disorder, intellectual developmental disorder and schizophrenia spectrum disorders. *Advances in mental health and intellectual disabilities*, *9*(5), 243-264.

Bettelheim, B. (1967). The Empty Fortress. Infantile Autism and the Birth of the Self. New York: The Free Press.

Binet, A. & Simon, T. (1905). *Mèthodes nouvelles pour le diagnostic du niveau intellectual des anormax.* Année Psycologique, 11, 191-244.

Birnbaum, A. (1968). *Some latent trait models*. Statistical theories of mental test scores.

Bishop-Fitzpatrick, L., Mazefsky, C. A., Eack, S. M., & Minshew, N. J. (2017). Correlates of social functioning in autism spectrum disorder: The role of social cognition. *Research in Autism Spectrum Disorders*, 35, 25-34.

Blair, R. J., & Cipolotti, L. (2000). Impaired social response reversal: A case ofacquired sociopathy'. *Brain*, 123(6), 1122-1141.

Bleuler, E. (1911). *Dementia praecox oder Gruppe der Schizophrenien*. New York: International Universities Press.

Bliksted, V., Videbech, P., Fagerlund, B., & Frith, C. (2017). The effect of positive symptoms on social cognition in first-episode schizophrenia is modified by the presence of negative symptoms. Neuropsychology, 31(2), 209.

Boada, L., Lahera, G., Pina-Camacho, L., Merchán-Naranjo, J., Díaz-Caneja, C. M., Bellón, J. M., Ruiz-Vargas, J. M., & Parellada, M. (2020). Social Cognition in Autism and Schizophrenia Spectrum Disorders: The Same but Different?. *Journal of autism and developmental disorders*, 50(8), 3046–3059.

Bock, D. (1997). A brief history of item theory. *Educational measurement: issues and practice*, 16(4), 21-33.

Boncori, L. (1993). *Teoria e tecnica dei test*, Torino, Bollati Boringhieri.

Bosco, F. M., Gabbatore, I., Tirassa, M., & Testa, S. (2016). Psychometric properties of the Theory of Mind Assessment Scale in a sample of adolescents and adults. *Frontiers in psychology*, 7, 566.

Bottema-Beutel, K., Park, H., & Kim, S. Y. (2018). Commentary on social skills training curricula for individuals with ASD: Social interaction, authenticity, and stigma. *Journal of autism and developmental disorders*, 48, 953-964.

Boucher, J. (2012). Putting theory of mind in its place: psychological explanations of the socio-emotional-communicative impairments in autistic spectrum disorder. *Autism*, 16(3), 226-246.

Bowler, D. M. (1992). ‘‘Theory of mind’’ in Asperger’s syndrome. *The Journal of Child Psychology and Psychiatry*, 33(5), 877–893.

Brewer, N., Young, R. L., & Barnett, E. (2017). Measuring theory of mind in adults with autism spectrum disorder. *Journal of autism and developmental disorders*, 47(7), 1927-1941.

Broekhof, E., Ketelaar, L., Stockmann, L., van Zijp, A., Bos, M. G., & Rieffe, C. (2015). The understanding of intentions, desires and beliefs in young children with autism spectrum disorder. *Journal of autism and developmental disorders*, 45(7), 2035-2045

Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. Handbook of structural equation modeling, 361, 379.

Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate behavioral research*, 36(1), 111-150.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological methods & research*, 21(2), 230-258.

Brugha, T. S., McManus, S., Bankart, J., Scott, F., Purdon, S., Smith, J., Bebbington, P., Jenkins, R., & Meltzer, H. (2011). Epidemiology of autism spectrum disorders in adults in the community in England. *Archives of general psychiatry*, 68(5), 459–465.

Brunas-Wagstaff, J. (1999). *La personalità*, Bologna, Il Mulino.

Brunet, E., Sarfati, Y., Hardy-Baylé, M. C., & Decety, J. (2003). Abnormalities of brain function during a nonverbal theory of mind task in schizophrenia. *Neuropsychologia*, 41(12), 1574-1582

Burnett, S., & Blakemore, S. J. (2009). The development of adolescent social cognition. *Annals of the New York Academy of Sciences,* 1167(1), 51–56.

Bussu, G., Jones, E. J. H., Charman, T., Johnson, M. H., Buitelaar, J. K., Baron-Cohen, S., Bedford, R., Bolton, P., Blasi, A., Chandler, S., Charman, T., Cheung, C., Davies, K., Elsabbagh, M., Fernandes, J., Gammer, I., Garwood, H., Gliga, T., Guiraud, J., … the BASIS Team. (2018). Prediction of Autism at 3 Years from Behavioural and Developmental Measures in High-Risk Infants: A Longitudinal Cross-Domain Classifier Analysis. Journal of Autism and Developmental Disorders, 48(7), 2418–2433. https://doi.org/10.1007/s10803-018-3509-x

Butcher, J. N. (2006). Assessment in Clinical Psychology: A Perspective on the Past, Present Challenges, and Future Prospects. *Clinical Psychology: Science and Practice*, *13*(3), 205.

Cadman, T., Spain, D., Johnston, P., Russell, A., Mataix-Cols, D., Craig, M., Deeley, Q., Robertson, D., Murphy, C., Gillan, N., Wilson, C. E., Mendez, M., Ecker, C., Daly, E., Findon, J., Glaser, K., MRC AIMS Consortium, Happé, F., & Murphy, D. (2015). Obsessive-Compulsive Disorder in Adults with High-Functioning Autism Spectrum Disorder: What Does Self-Report with the OCI- R Tell Us?. *Autism research: official journal of the International Society for Autism Research*, *8*(5), 477–485.

Carpenter, M., Pennington, B. F., & Rogers, S. J. (2001). Understanding of others' intentions in children with autism. *Journal of autism and developmental disorders*, *31*(6), 589-599.

Carpenter, S. K. (2005). Some neglected contributions of Wilhelm Wundt to the psychology of memory. *Psychological reports*, *97*(1), 63-73.

Carruthers, P. (1966). Simulation and self-knowledge. *Theories of Theories of Mind (Cambridge University Press, 1996)*, 22-38.

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, 125(8), 1839-1849.

Cattell, J. M. (1890). *Mental tests and measurements*. Mind, 15, 373-380.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. Journal of Educational Psychology, 54(1), 1–22.

Chaidez, V., Hansen, R. L., & Hertz-Picciotto, I. (2014). Gastrointestinal problems in children with autism, developmental delays or typical development. *Journal of autism and developmental disorders*, *44*(5), 1117- 1127

Chambon, V., Farrer, C., Pacherie, E., Jacquet, P. O., Leboyer, M., & Zalla, T. (2017). Reduced sensitivity to social priors during action prediction in adults with autism spectrum disorders. *Cognition*, 160, 17-26.

Champagne-Lavau, M., & Charest, A. (2015). Theory of mind and context processing in schizophrenia: The role of social knowledge. *Frontiers in Psychiatry*, 6, Article 98.

Chasson, G. S., Harris, G. E., & Neely, W. J. (2007). Cost comparison of early intensive behavioral intervention and special education for children with autism. Journal of Child and Family Studies, 16, 401-413.

Chen, J., & Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3), 759–771.

Chen, M. H., Su, T. P., Chen, Y. S., Hsu, J. W., Huang, K. L., Chang, W. H., Chen, T.J., & Bai, Y. M. (2013). Comorbidity of allergic and autoimmune diseases in patients with autism spectrum disorder: A nationwide population-based study. *Research in autism spectrum disorders*, *7*(2), 205-212.

Cho, I., & Cohen, A. S. (2019). Explaining age-related decline in theory of mind: Evidence for intact competence but com- promised executive function. *PLOS ONE*, 14(9), Article e0222890.

Christensen, A. P., & Golino, H. (2019). Estimating the stability of the number of factors via Bootstrap Exploratory Graph Analysis: A tutorial. *PsyArXiv*, 10.

Christensen, A. P., & Golino, H. (2021). Estimating the stability of psychological dimensions via bootstrap exploratory graph analysis: A Monte Carlo simulation and tutorial. *Psych*, 3(3), 479-500.

Christensen, A. P., & Golino, H. (2021a). Estimating the stability of psychological dimensions via bootstrap exploratory graph analysis: A Monte Carlo simulation and tutorial. *Psych*, 3(3), 479–500. https://doi. org/10.3390/psych3030032

Christensen, A. P., & Golino, H. (2021c). On the equivalency of factor and network loadings. *Behavior Research Methods*, 53, 1563–1580. https://doi.org/10.3758/s13428-020-01500-6

Christensen, A. P., Cotter, K. N., & Silvia, P. J. (2019). Reopening openness to experience: A network analysis of four openness to experience inventories. *Journal of Personality Assessment*, 101(6), 574–588. https:// doi.org/10.1080/00223891.2018.1467428

Christensen, A. P., Kenett, Y. N., Aste, T., Silvia, P. J., & Kwapil, T. R. (2018). Network structure of the wisconsin schizotypy scales–Short forms: Examining psychometric network filtering approaches. Behavior Research Methods, 50(6), 2531–2550. https://doi.org/doi: 10.3758/s13428-018-1032-9

Christensen, J., Grønborg, T. K., Sørensen, M. J., Schendel, D., Parner, E. T., Pedersen, L. H., & Vestergaard, M. (2013). Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. *Jama*, 309(16), 1696-1703.

Chung, Y. S., Barch, D., & Strube, M. (2014). A meta-analysis of mentalizing impairments in adults with schizophrenia and autism spectrum disorder. *Schizophrenia bulletin*, 40(3), 602-616.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55(1), 591–621.

Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment* 6(4): 284.

Clark, L. A., & Watson, D. (2016). Constructing validity: Basic issues in objective scale development. *Psychological Assessment,* 7(3): 309–319.

Clark, R. A., & Delia, J. G. (1976). The development of func- tional persuasive skills in childhood and early adolescence. *Child Development*, 47(4), 1008–1014.

Cole, E. J., Slocombe, K. E., & Barraclough, N. E. (2018). Abilities to explicitly and implicitly infer intentions from actions in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 48, 1712-1726.

Colle, L., Baron-Cohen, S., & Hill, J. (2007). Do children with autism have a theory of mind? A non-verbal test of autism vs. specific language impairment. *Journal of autism and developmental disorders*, 37, 716-723.

Colombi, C., Tancredi, R., Persico, A., & Faggioli, R. (2013). *ADOS-2–Autism Diagnostic Observation Schedule.* Firenze: Hogrefe.

Corrales, M. A., & Herbert, M. (2011). Autism and environmental genomics: synergistic systems approaches to autism complexity. In: Amaral DG, Dawson G, Geschwind DH, eds. Autism spectrum disorders. New York, NY: Oxford University Press, 2011: 875–92.

Corrigan, P. W., & Penn, D. L. (2001). Social cognition and schizophrenia. *American Psychological Association.*

Craig, K. (2017). The history of psychometrics. *Psychometric Testing: Critical Perspectives*, 1-14.

Creak, M. (1961). *Schizophrenia syndrome in childhood*. Cerebral Palsy Bull, 3, 501.

Cristante, F., & Mannarini, S. (2003). *Psicometria* (pp. 1-236). Il Mulino.

Cristante, F., & Mannarini, S. (2003). *Psicometria* (pp. 1-236). Il mulino.

Crop, I. B. M. (2020). IBM SPSS Statistics for Windows (Version 27.0)[Computer software]. IBM Corp.

Dahlgren, S. O., & Trillingsgaard, A. (1996). Theory of mind in non-retarded children with autism and Asperger's syndrome. A research note. *Journal of Child Psychology and Psychiatry*, 37(6), 759-763.

Davis, T. E., Hess, J. A., Moree, B. N., Fodstad, J. C., Dempsey, T., Jenkins, W. S., & Matson, J. L. (2011). Anxiety symptoms across the lifespan in people diagnosed with autistic disorder. *Research in autism spectrum disorders*, *5*(1), 112-118.

De Ayala, R. J. (2013). The theory and practice of item response theory. Guilford Publications.

de Bildt, A., Oosterling, I. J., van Lang, N. D., Sytema, S., Minderaa, R. B., van Engeland, H., ... & de Jonge, M. V. (2011). Standardized ADOS scores: Measuring severity of autism spectrum disorders in a Dutch sample. *Journal of autism and developmental disorders*, 41, 311-319.

De Sanctis, S. (1909). Quadri clinici di Dementia praecox nell'infanzia e nella fanciullezza. *Riv. Ital. di Neuropatol. Psichiatr. de Electtroter*, *2*, 97-104.

Dean, M., Harwood, R., & Kasari, C. (2017). The art of camouflage: Gender differences in the social behaviors of girls and boys with autism spectrum disorder. *Autism*, 21(6), 678-689.

Delobel-Ayoub, M., Saemundsen, E., Gissler, M., Ego, A., Moilanen, I., Ebeling, H., ... & Schendel, D. (2020). Prevalence of autism spectrum disorder in 7–9-year-old children in Denmark, Finland, France and Iceland: a population-based registries approach within the ASDEU project. *Journal of autism and developmental disorders*, 50, 949-959.

Denkla, M. B. (1996b). A theory and model of executive function: a neuropsychological perspective. In G. R. Lyon & N. A. Krasnegor (Eds.), Attention, memory and executive function (pp. 263–278). Baltimore, MD: Paul H. Brookes.

Derntl, B., Finkelmeyer, A., Eickhoff, S., Kellermann, T., Falkenberg, D. I., Schneider, F., & Habel, U. (2010). Multidimensional assessment of empathic abilities: Neural correlates and gender differences. *Psychoneuroendocrinology*, 35(1), 67–82.

Deste, G., Vita, A., Penn, D. L., Pinkham, A. E., Nibbio, G., & Harvey, P. D. (2020). Autistic symptoms predict social cognitive performance in patients with schizophrenia. *Schizophrenia research*, *215*, 113-119.

DeStefano, F., Price, C. S., & Weintraub, E. S. (2013). Increasing exposure to antibody-stimulating proteins and polysaccharides in vaccines is not associated with risk of autism. *The Journal of pediatrics*, 163(2), 561–567.

Dewey, M. (1991). "Living with Asperger's syndrome." In U. Frith (Ed.), Autism and Asperger syndrome (pp. 184–206). Cambridge University Press.

Di Tella, M., Miti, F., Ardito, R. B., & Adenzato, M. (2020). Social cognition and sex: Are men and women really dif- ferent? Personality and Individual Differences, 162, Article 110045.

Dichter, G. S. (2012). Functional magnetic resonance imaging of autism spectrum disorders. *Dialogues in clinical neuroscience*, 14(3), 319-351.

Dietz, P. M., Rose, C. E., McArthur, D., & Maenner, M. (2020). National and state estimates of adults with autism spectrum disorder. *Journal of autism and developmental disorders*, 50, 4258-4266.

Dinstein, I., Gardner, J. L., Jazayeri, M., & Heeger, D. J. (2008). Executed and observed movements have different distributed representations in human aIPS. *Journal of Neuroscience*, *28*(44), 11231-11239.

Dodich, A., Cerami, C., Canessa, N., Crespi, C., Iannaccone, S., Marcone, A., ... & Cappa, S. F. (2015). A novel task assessing intention and emotion attribution: Italian

standardization and normative data of the Story-based Empathy Task. *Neurological Sciences*, 36, 1907-1912.

Doernberg, E., & Hollander, E. (2016). Neurodevelopmental disorders (asd and adhd): Dsm-5, icd-10, and icd-11. CNS spectrums, 21(4), 295-299.

Donadello, I., Spoto, A., Sambo, F., Badaloni, S., Granziol, U., & Vidotto, G. (2017). ATS-PD: An adaptive testing system for psychological disorders. *Educational and psychological measurement*, 77(5), 792-815.

Drury, V. M., Robinson, E. J., & Birchwood, M. (1998). 'Theory of mind'skills during an acute episode of psychosis and following recovery. *Psychological medicine*, 28(5), 1101-1112.

Dumas, G., Soussignan, R., Hugueville, L., Martinerie, J., & Nadel, J. (2014). Revisiting mu suppression in autism spectrum disorder. *Brain research*, *1585*, 108-119.

Duyzend, M. H., Nuttle, X., Coe, B. P., Baker, C., Nickerson, D. A., Bernier, R., & Eichler, E. E. (2016). Maternal modifiers and parent-of-origin bias of the autism-associated 16p11. 2 CNV. *The American Journal of Human Genetics*, 98(1), 45-57.

Eddy, C. M. (2019). What do you have in mind? Measures to assess mental state reasoning in neuropsychiatric populations. *Frontiers in psychiatry*, 10, 425.

Elison, J. T., Wolff, J. J., Reznick, J. S., Botteron, K. N., Estes, A. M., Gu, H., Hazlett, H. C., Meadows, A. J., Paterson, S. J., Zwaigenbaum, L., & Piven, J. (2014). Repetitive Behavior in 12-Month-Olds Later Classified With Autism Spectrum Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry,* 53(11), 1216–1224. https://doi.org/10.1016/j.jaac.2014.08.004

Epskamp, S. (2015). semPlot: Unified Visualizations of Structural Equation Models. Structural Equation Modeling: A Multidisciplinary Journal, 22(3), 474–483. https://doi.org/10.1080/10705511.2014.937847

Epskamp, S., & Fried, E. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, 23(4), 617–634. https://doi.org/10.1037/met0000167

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network pschometrics: Combining network and latent variable models. *Psychometrika*, 82(4), 904–927. https://doi.org/10.1007/ s11336-017-9557-x

Ercolani, A. P., & Perugini, M. (1997). *La misura in psicologia*: Introduzione ai test psicologici. Milano, Led.

Eusebi, P. (2013). Diagnostic accuracy measures. Cerebrovascular Diseases, 36(4), 267-272.

Evans, B. (2013). How autism became autism: The radical transformation of a central concept of child development in Britain. *History of the Human Sciences*, 26(3), 3-31.

Fairfax, H. (2017). Psychometrics in clinical settings. *Psychometric Testing: Critical Perspectives*, 175-184.

Fernandes, J. M., Cajão, R., Lopes, R., Jerónimo, R., & Barahona-Corrêa, J. B. (2018). Social cognition in schizophrenia and autism spectrum disorders: a systematic review and meta-analysis of direct comparisons. *Frontiers in psychiatry*, 9, 504.

Ferrante, M., Barone, R., Fazio, A., Zerbo, S., Margherita, V., Rizzo, R., Fichera, G. & Fiore, M. (2015). Prevalence and age at diagnosis of Autism Spectrum Disorder in south Italy, 2004–2014: Maria Fiore. *The European Journal of Public Health*, 25(suppl_3), ckv175-099.

Flake, J. K., Davidson, I. J., Wong, O., & Pek, J. (2022). Construct validity and the validity of replication studies: A systematic review. *American Psychologist*, 77(4), 576–588. https://doi.org/10.1037/amp0001006

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. Social Psychological and Personality Science, 8(4), 370-378.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological assessment*, 7(3), 286.

Fodor, J.A. (1983). *The modularity of mind*. MIT press.

Fombonne, E. (2005). The changing epidemiology of autism. *Journal of applied research in intellectual disabilities*, 18(4), 281-294.

Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, 45(1), 30.

Frazier, T. W., & Hardan, A. Y. (2009). A meta-analysis of the corpus callosum in autism. *Biological psychiatry*, 66(10), 935-941.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3), 432-441.

Frith, C. D. (1992).*The cognitive neuropsychology of schizophrenia*. Psychology press.

Frith, C. D. (2004). Schizophrenia and theory of mind. *Psychological Medicine*, 34(3), 385–389.

Frith, C. D., & Corcoran, R. (1996). Exploring 'theory of mind'in people with schizophrenia. *Psychological medicine*, 26(3), 521-530.

Frith, U. (1970). Studies in pattern detection in normal and autistic children. 1. Immediate recall of auditory sequences. *Journal of Abnormal Psychology*, 76, 413–420.

Frith, U. (1971). Spontaneous patterns produced by autistic, normal and subnormal children. In M. Rutter (Ed.), Infantile autism: Concepts, characteristics and treatment (pp. 113–133). London: Churchill Livingstone.

Frith, U. (1989). *Autism: Explaining the enigma*. Oxford, UK: Blackwell.

Frith, U. (1996). Cognitive Explanations of Autism. *Acta Paediatrica*, *85*, 63-68.

Frith, U. (2003). Autism: Explaining the enigma (2nd ed.). Oxford: Blackwell.

Frith, U., & Happé, F. (1994). Autism—beyond theory of mind. Cognition, 50(1–3), 115–132.

Frith, U., & Happé, F. (2005). Autism spectrum disorder. *Current Biology, 15*(19), 786–790

Fuentes, J., Hervás, A., & Howlin, P. (2020). ESCAP practice guidance for autism: a summary of evidence-based recommendations for diagnosis and treatment. *European Child & Adolescent Psychiatry*, 1-24.

Funabiki, Y., & Shiwa, T. (2018). Weakness of visual working memory in autism. *Autism Research*, 11(9), 1245-1252.

Furr, R. M. (2018). Psychometrics: An Introduction (3rd edn.). Thousand Oaks, CA: Sage

Furr, R. M. (2020). Psychometrics in clinical psychological research. In A. G. C. Wright & M. N. Hallquist (Eds.), *The Cambridge handbook of research methods in clinical psychology* (pp. 54–65). Cambridge University Press. https://doi.org/10.1017/9781316995808.008

Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics : an introduction*. Retrieved from http://catdir.loc.gov/catdir/enhancements/fy0808 /2007016663-b.html

Fusar-Poli, L., Brondino, N., Rocchetti, M., Panisi, C.,Provenzani, U., Damiani, S., & Politi, P. (2017). Diagnosing ASD in adults without ID: Accuracy of the ADOS-2 and the ADI-R. *Journal of Autism and Developmental Disorders*, 47(11), 3370–3379.

Gallese, V. (2007). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 659-669.

Gallese, V., Rochat, M. J., & Berchio, C. (2013). The mirror mechanism and its potential role in autism spectrum disorder. *Developmental Medicine & Child Neurology*, *55*(1), 15-22.

Galton, F. (1883). *Inquiries into human faculty and its development*, London, Mac Millan.

Garcia-Garzon, E., Abad, F. J., & Garrido, L. E. (2019). Improving bi-factor exploratory modeling. Methodology.

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at horn's parallel analysis with ordinal variables. *Psychological Methods*, 18(4), 454–74. https://doi.org/10.1037/a0030005

Gatta, M., Spoto, A., Trevisan, G., Benanti, S., & Battistella, P. A. (2015). Attribuzione delle emozioni: Analisi preliminare sulla validità e affidabilità per l'età adolescenziale del test di attribuzione delle emozioni. Gior Neuropsich Età Evol, 35, 1–17.

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of marketing research*, 25(2), 186-192.

Giannini, M., Gori, A., Pannocchia, L., Schimmenti, A., & Caretti, V. (2014). *Principi di psicometria clinica* (pp. 1-308). Raffaello Cortina Editore.

Goldenberg, E. N. (1983). The three faces of evaluation. *Journal of Policy Analysis and Management*, *2*(4), 515-525.

Goldman, A.I. (1989). Interpretation Psychologized. *Mind e Language*, *4*(3), 161-185.

Goldman, A.I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press.

Golino, H. F., & Demetriou, A. (2017). Estimating the dimensionality of intelligence like data using exploratory graph analysis. Intelligence, 62, 54–70. https://doi.org/10.1016/j.intell.2017.02.007

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS ONE*, 12(6), e0174035. https://doi.org/10.1371/journal.pone.0174035

Golino, H. F., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martínez-Molina, A. (2020). Investigating the performance of Exploratory Graph Analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. Psychological Methods, 25(3), 292–320. https://doi.org/10.1037/met0000255

Golino, H., Christensen, A. P., Moulder, R. G. Jr., Kim, S., & Boker, S. M. (2022). Modeling latent topics in social media using dynamic exploratory graph analysis: The case of the right-wing and left-wing trolls in the 2016 US elections. *Psychometrika*, 87(1), 156–187. https://doi.org/10.1007/s11336-021-09820-y

Golino, H., Moulder, R., Shi, D., Christensen, A., Garrido, L., Nieto, M. D., Nesselroadea, J., Sadana, R., Thiyagarajan, J. A., & Boker, S. (2020). Entropy fit indices: New fit measures for assessing the structure and dimensionality of multiple latent variables.

Multivariate Behavioral Research, 56(6), 874–902. https://doi.org/10.1080/00273171.2020.1779642

Gopnik, A., & Wellman, H. M. (1994). *The Theory Theory*. In Hirschfield L. and Gelman S., eds., Mapping the Mind: Domain Specificity in Cognition and Culture. Cambridge University Press, 257-93.

Gopnik, A., Capps, L. & Meltzoff, A.N. (2000). Early Theories of Mind: What the Theory Theory Can Tell Us About Autism? In Baron-Cohen, S., Tager- Flusberger H. e Cohen. D., eds., *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience, 2nd ed*, 50-72. Oxford: Oxford University Press

Gordon, R.M. (1986). Folk Psychology as Simulation. *Mind e Language*, *1*(2), 158-171.

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. *Journal of autism and developmental disorders*, 37, 613-627.

Grafman, J. (1994). *Neuropsychology of the prefrontal cortex*. In: D. W. Zaidel. (Ed.), Neuropsychology (pp. 159–181). San Diego, CA: Academic Press.

Grafman, J., Schwab, K., Warden, D., Pridgen, A., Brown, H. R., & Salazar, A. M. (1996). Frontal lobe injuries, violence, and aggression: A report of the Vietnam Head Injury Study. *Neurology*, 46(5), 1231–1231.

Greydanus, D. E., & Toledo-Pereyra, L. H. (2012). Historical perspectives on autism: Its past record of discovery and its present state of solipsism, skepticism, and sorrowful suspicion. *Pediatric Clinics*, 59(1), 1-11.

Grigorenko, E. L., Han, S. S., Yrigollen, C. M., Leng, L., Mizue, Y., Anderson, G. M., Mulder, E. J., de Bildt, A., Minderaa, R. B., Volkmar, F. R., Chang, J. T., & Bucala, R. (2008). Macrophage migration inhibitory factor and autism spectrum disorders. *Pediatrics*, *122*(2), e438–e445.

Grossi, E., Caminada, E., Goffredo, M., Vescovo, B., Castrignano, T., Piscitelli, D., Valagussa, G., Franceschini, M., & Vanzulli, F. (2021). Patterns of Restricted and Repetitive Behaviors in Autism Spectrum Disorders: A Cross-Sectional Video Recording Study. Preliminary Report. Brain Sciences, 11(6), Articolo 6. https://doi.org/10.3390/brainsci11060678

Grossman, J. B., Klin, A., Carter, A. S., & Volkmar, F. R. (2000). Verbal bias in recognition of facial emotions in children with Asperger syndrome. *The Journal of Child Psychology and Psychiatry and Allied Disciplines,* 41(3), 369- 379.

Grzadzinski, R., Di Martino, A., Brady, E., Mairena, M. A., O'Neale, M., Petkova, E., Lord, C., & Castellanos, F. X. (2011). Examining autistic traits in children with ADHD: does the autism spectrum extend to ADHD?. *Journal of autism and developmental disorders*, *41*(9), 1178–1191.

Gyawali, S., & Patra, B. N. (2019). Trends in concept and nosology of autism spectrum disorder: A review. *Asian journal of psychiatry*, 40, 92-99.

Haker, H., Schneebeli, M., & Stephan, K. E. (2016). Can Bayesian theories of autism spectrum disorder help improve clinical practice?. *Frontiers in psychiatry*, 7, 107.

Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., Miller, J., Fedele, A., Collins, J., Smith, K., Lotspeich, L., Croen, L. A., Ozonoff, S., Lajonchere, C., Grether, J. K., & Risch, N. (2011). Genetic heritability and shared environmental factors among twin pairs with autism. *Archives of general psychiatry*, 68(11), 1095–1102

Happé, F. (1999). Autism: cognitive deficit or cognitive style?. *Trends in Cognitive Sciences*, 3(6), 216–222.

Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of autism and developmental disorders*, 36, 5-25.

Happé, F., & Frith, U. (2014). Annual research review: towards a developmental neuroscience of atypical social cognition. *Journal of Child Psychology and Psychiatry*, 55(6), 553-577.

Happé, F., Cook, J. L., & Bird, G. (2017). The structure of social cognition: In (ter) dependence of sociocognitive processes. *Annual review of psychology*, 68(1), 243-267.

Harrison, A. J., Slane, M. M., Hoang, L., & Campbell, J. M. (2017). An international review of autism knowledge assessment measures. *Autism*, 21(3), 262-275.

Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *The Journal of Psychology*, *10*(2), 249-254.

Hatt N., Colombi C., Saron C., Rogers S. e Rivera S. (2009), Neural Basis of Action and Intention Understanding in Autism and Typical Development. In *International Meeting for Autism Research*.

Hayes, J., Ford, T., Rafeeque, H., & Russell, G. (2018). Clinical practice guidelines for diagnosis of autism spectrum disorder in adults and children in the UK: a narrative review. *BMC psychiatry*, 18, 1-25.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational research methods*, 7(2), 191-205.

Heavey, L., Phillips, W., Baron-Cohen, S., & Rutter, M. (2000). The awkward moments test: A naturalistic measure of social understanding in autism. *Journal of Autism and Developmental Disorders*, 30:, 225–236.

Heller T (1908) Über dementia infantilis. *Zeitschrift fur die Erforschung und Behandlung des Jugenlichen Schuachsinns*, 3: 141–65.

Henry, J. D., Von Hippel, W., Molenberghs, P., Lee, T., & Sachdev, P. S. (2015). Clinical assessment of social cognitive function in neurological disorders. Nature Reviews Neurology, 12(1), 28-39.

Hermelin, B., & O'Connor, N. (1970). Psychological experiments with autistic children. Oxford: Pergamon Press.

Heuer, L., Ashwood, P., Schauer, J., Goines, P., Krakowiak, P., Hertz-Picciotto, I., Hansen, R., Croen, L. A., Pessah, I. N., & Van de Water, J. (2008). Reduced levels of immunoglobulin in children with autism correlates with behavioral symptoms. *Autism research : official journal of the International Society for Autism Research*, *1*(5), 275–283.

Hill, E. L. (2004a). Evaluating the theory of executive dysfunction in autism. *Developmental Review*, 24(2), 189–233.

Hirota, T., & King, B. H. (2023). Autism spectrum disorder: a review. *Jama*, 329(2), 157-168.

Hodges, H., Fealko, C., & Soares, N. (2020). Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation. *Translational pediatrics*, *9* (Suppl 1), S55.

Hollocks, M. J., Lerh, J. W., Magiati, I., Meiser-Stedman, R., & Brugha, T. S. (2019). Anxiety and depression in adults with autism spectrum disorder: a systematic review and meta-analysis. *Psychological medicine*, 49(4), 559-572.

Hong, J. S., Singh, V., Kalb, L., Ashkar, A., & Landa, R. (2021). Replication study of ADOS-2 Toddler Module cut-off scores for autism spectrum disorder classification. *Autism Research*, 14(6), 1284-1295.

Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated?. *Personality and social psychology review*, 14(3), 332-346.

Hossain, M. M., Khan, N., Sultana, A., Ma, P., McKyer, E. L. J., Ahmed, H. U., & Purohit, N. (2020). Prevalence of comorbid psychiatric disorders among people with autism spectrum disorder: An umbrella review of systematic reviews and meta-analyses. *Psychiatry research*, *287*, 112922.

Howlin, P., & Magiati, I. (2017). Autism spectrum disorder: Outcomes in adulthood. *Current opinion in psychiatry*, 30(2), 69-76.

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2004). Adult outcome for children with autism. Journal of child psychology and psychiatry, 45(2), 212-229.

Howlin, P., Moss, P., Savage, S., & Rutter, M. (2013). Social outcomes in mid-to later adulthood among individuals diagnosed with autism and average nonverbal IQ as children. Journal of the American Academy of Child & Adolescent Psychiatry, 52(6), 572-581.

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Hudson, C. C., Hall, L., & Harkness, K. L. (2019). Prevalence of depressive disorders in individuals with autism spectrum disorder: A meta-analysis. *Journal of Abnormal Child Psychology*, *47*(1), 165-175.

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British journal of developmental psychology*, 16(2), 233-253.

Hyman, S. L., Levy, S. E., & Myers, S. M. (2020). Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics*, *145*(1)

Jamison, L., Golino, H., & Christensen, A. P. (2022). Metric invariance in exploratory graph analysis via permutation testing. *PsycArxiv*. https://doi.org/10.31234/osf.io/j4rx9

Jonckheere, A. R. (1954). A Distribution-Free k-Sample Test Against Ordered Alternatives. Biometrika, 41(1/2), 133–145. https://doi.org/10.2307/2333011

Jones, C., Simonoff, E., Baird, G., Pickles, A., Marsden, A., Tregay, J., Happé, F., & Charman, T. (2018). The association between theory of mind, executive function, and the symptoms of autism spectrum disorder. *Autism research : official journal of the International Society for Autism Research*, 11(1), 95–109.

Joreskog, K. G., & Sorbom, D. (1989). *LISREL 7 user's reference guide*. Scientific Software.

Jurek, L., Balthazar, M., Gulati, S., Novakovic, N., Núñez, M., Oakley, J., & O'Hagan, A. (2021). Response (minimum clinically relevant change) in ASD symptoms after an intervention according to CARS-2: consensus from an expert elicitation procedure. *European child & adolescent psychiatry*, 1-10.

Kamata, A., & Vaughn, B. K. (2004). An introduction to differential item functioning analysis. Learning Disabilities: A Contemporary Journal, 2(2), 49–69.

Kamp-Becker, I., Ghahreman, M., Heinzel-Gutenbrunner, M., Peters, M., Remschmidt, H., & Becker, K. (2013). Evaluation of the revised algorithm of Autism Diagnostic

Observation Schedule (ADOS) in the diagnostic investigation of high-functioning children and adolescents with autism spectrum disorders. *Autism*, 17(1), 87-102.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous child*, *2*(3), 217-250.

Keller, R. (Ed.). (2019). Psychopathology in Adolescents and Adults with Autism Spectrum Disorders. Springer Nature.

Keller, R., & Bari, S. (2019). Psychosis and ASD. In *Psychopathology in adolescents and adults with autism spectrum disorders* (pp. 51-65). Springer, Cham.

Keller, R., Bari, S., Aresi, A., Notaro, L., Bianco, F., & Pirfo, E. (2015). Diagnosing autism spectrum disorders in adulthood. *Journal of Psychopathology*, *21*, 13-18.

Keller, R., Bugiani, S., Fantin, P., & Pirfo, E. (2011). Neuroni specchio e autismo. *Giornale Italiano di Psicopatologia*, 17, 404-412.

Keller, R., Chieregato, S., Bari, S., Castaldo, R., Rutto, F., Chiocchetti, A., & Dianzani, U. (2020). Autism in adulthood: clinical and demographic characteristics of a cohort of five hundred persons with autism analyzed by a novel multistep network model. *Brain sciences*, 10(7), 416.

Khan, M. R. A., & Brandenburger, T. (2020). ROCit: Performance assessment of binary classifier with visualization. *R Package Version*, *2*(1), 1–21. https://CRAN.Rproject.org/package=ROCit

Kilford, E. J., Garrett, E., & Blakemore, S. J. (2016). The development of social cognition in adolescence: An integrated perspective. *Neuroscience & Biobehavioral Reviews*, 70, 106-120.

King, C., Merrick, H., & Le Couteur, A. (2020). How should we support young people with ASD and mental health problems as they navigate the transition to adult life including access to adult healthcare services. *Epidemiology and psychiatric sciences*, 29, e90.

Kirkovski, M., Enticott, P. G., Hughes, M. E., Rossell, S. L., & Fitzgerald, P. B. (2016). Atypical neural activity in males but not females with autism spectrum disorder. *Journal of autism and developmental disorders*, 46, 954-963.

Kline, P. (1996). *Manuale di psicometria: Come costruire, valutare e applicare un test psicologico*. Astrolabio.

Klopfer, B. (1936). *Rorschach research exchange*. Rorschach Institute.

Kluever, R. C., Smith, D. K., Green, K. E., Holm, C. B., & Dimson, C. (1995). The WISC-III and Raven coloured progressive matrices test: A pilot study of relationships. https://files.eric.ed.gov/fulltext/ED390938.pdf

Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. CBE—Life Sciences Education, 18(1), rm1.

Kolvin, I. (1971). Studies in the childhood psychoses I. Diagnostic criteria and classification. *The British Journal of Psychiatry*, 118(545), 381–384.

Lai, M. C., & Baron-Cohen, S. (2015). Identifying the lost generation of adults with autism spectrum conditions. *The Lancet Psychiatry*, *2*(11), 1013-1027.

Lai, M. C., Kassee, C., Besney, R., Bonato, S., Hull, L., Mandy, W., Szatmari, P., & Ameis, S. H. (2019). Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. *The lancet Psychiatry*, *6*(10), 819–829.

Lai, M. C., Lombardo, M. V., Ruigrok, A. N., Chakrabarti, B., Auyeung, B., Szatmari, P., Happé, F., Baron-Cohen, S., & MRC AIMS Consortium (2017). Quantifying and exploring camouflaging in men and women with autism. *Autism*, 21(6), 690–702.

Lai, M.C., Lombardo, M.V., & Baron-Cohen, S. (2014) Autism. *Lancet*, 383: 896–910.

Lauritzen, S. L. (1996). *Graphical models* (Vol. 17). Oxford: Clarendon Press.

Lee, K. S., Chung, S. J., Thomas, H. R., Park, J., & Kim, S. H. (2019). Exploring diagnostic validity of the autism diagnostic observation schedule-2 in south Korean toddlers and preschoolers. *Autism Research*, 12(9), 1356-1366.

Leitner, Y. (2014). The co-occurrence of autism and attention deficit hyperactivity disorder in children–what do we know?. *Frontiers in human neuroscience*, *8*, 268.

Leslie, A. M. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition*, 50(1-3), 211-238.

Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., Jones, E., Jones, R. M., Pickles, A., State, M. W., Taylor, J. L., & Veenstra- VanderWeele, J. (2020). Autism spectrum disorder. *Nature reviews. Disease primers*, *6*(1), 5.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview- Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders.*Journal of autism and developmental disorders*, *24*(5), 659-685

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). Autism diagnostic observation schedule, (ADOS-2) modules 1-4. Los Angeles, California: Western Psychological Services

Lord, F.M. & Novick, M.R. (1968). *Statistical theories of metal test scores.* Reading: Addison Wesley

Love, N., Ruff, G., & Geldmacher, D. S. (2015). Social cogni- tion in older adults: A review of neuropsychology, neuro- biology, and functional connectivity. *Medical & Clinical Reviews*, 1, Article 6.

Loveland, K. A., Pearson, D. A., Tunali-Kotoski, B., Ortegon, J.,& Gibbs, M. C. (2001). Judgments of social appropriateness by children and adolescent with autism. *Journal of Autism and Developmental Disorders,* 31:, 367–376.

Lubbe, D. (2019). Parallel analysis with categorical variables: Impact of category probability proportions on dimensionality assessment accuracy. *Psychological Methods*, 24(3), 339–351. https://doi.org/http://dx.doi.org/10.1037/met0000171

Luce, R. D., & Krumhansl, C. L. (1988). *Measurement, scaling, and psychophysics*. In R.C. Atkinson – R.J. Herrnstein- G. Lindzey – R.D. Luce (eds.). *Stevens'handbook of Experimental Psychology, vol.1: perception and motivation,* New York, Wiley and Sons.

Lucidi, F. (2013). *Problemi di Attendibilità e Validità nelle Prove Standardizzate per la Valutazione del Sistema Educativo*.https://www.fondazioneagnelli.it/wp-content/uploads/2017/08/Lucidi_2013__Problemi_di_attendibilita_e_validita_nelle_prove_standardizzate.pdf.

Lüdemann, L., Grieger, W., Wurm, R., Wust, P., & Zimmer, C. (2006). Glioma assessment using quantitative blood volume maps generated by T1-weighted dynamic contrast-enhanced magnetic resonance imaging: A receiver operating characteristic study. *Acta Radiologica*, 47(3), 303–310.

Lugnegård, T., Hallerbäck, M. U., & Gillberg, C. (2011). Psychiatric comorbidity in young adults with a clinical diagnosis of Asperger syndrome. *Research in developmental disabilities*, *32*(5), 1910-1917.

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., ... & Lord, C. (2009). The Autism Diagnostic Observation Schedule—Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of autism and developmental disorders*, 39, 1305-1320.

Maat, A., Fett, A. K., Derks, E. & Group Investigators. (2012). Social cognition and quality of life in schizophrenia. *Schizophrenia Research*, 137(1–3), 212–218.

MacDonald, P., & Paunonen, S. V. (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and psychological measurement*, 62(6), 921-943.

Madonna, R., Valenti, M., Borrelli, G., Cerbo, R., De Lellis, M., Sprovera, G., Massaro, M., D'Alessandro, A., Marinucci, M. C., Tiberti, S., & Di Orio, F. (1998). Epidemiological

monitoring of psychological, neurological and sensory handicaps in the 0-24-year-old Abruzzo population (Italy). Preliminary results. *Epidemiol Psichiatr Soc*, 7(3): 188-96

Maenner, M. J. (2023). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2020. MMWR. Surveillance Summaries, 72.

Mair, P., Hatzinger, R., & Maier, M. J. (2021). eRm: Extended Rasch Modeling. 1.0-2. https://cran.r-project.org/package=eRm

Mandy, W., & Lai, M. C. (2016). Annual Research Review: The role of the environment in the developmental psychopathology of autism spectrum condition. *Journal of Child Psychology and Psychiatry*, 57(3), 271-292.

Mannion, A., & Leader, G. (2013). Comorbidity in autism spectrum disorder: A literature review. *Research in Autism Spectrum Disorders*, *7*(12), 1595-1616.

Mari, M., Castiello, U., Marks, D., Marraffa, C., & Prior, M. (2003). The reach-to-grasp movement in children with autism spectrum disorder. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 358(1430), 393–403.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological bulletin*, 103(3), 391.

Martinez, G., Mosconi, E., Daban-Huard, C., Parellada, M., Fananas, L., Gaillard, R., Fatjo-Vilas, M., Krebs, M. O., & Amado, I. (2019). "A circle and a triangle dancing together": Alteration of social cognition in schizophrenia compared to autism spectrum disorders. Schizophrenia research, 210, 94–100.

Masedu, F., Vagnetti, R., Pino, M. C., Valenti, M., & Mazza, M. (2021). Comparison of visual fixation trajectories in toddlers with autism spectrum disorder and typical development: a Markov chain model. *Brain sciences*, 12(1), 10.

Massara, G. P., Di Matteo, T., & Aste, T. (2016). Network filtering for big data: Triangulated maximally filtered graph. *Journal of Complex Networks*, 5(2), 161–178.

Matsui, M., Arai, H., Yonezawa, M., Sumiyoshi, T., Suzuki, M., & Kurachi, M. (2009). The effects of cognitive rehabilita- tion on social knowledge in patients with schizophrenia. *Applied Neuropsychology*, 16(3), 158–164.

Matsui, M., Sumiyoshi, T., Yuuki, H., Kato, K., & Kurachi, M. (2006). Impairment of event schema in patients with schizo- phrenia: Examination of script for shopping at supermarket. *Psychiatry Research*, 143(2–3), 179–187.

Mazza, M., & Valenti, M. (2019). Cognizione Sociale e autismo. Modelli comportamentali, implicazioni cliniche e valutazione. Franco Angeli Editore, Serie di Psicologia

Mazza, M., Attanasio, M., Pino, M. C., Masedu, F., Tiberti, S., Sarlo, M., & Valenti, M. (2020). Moral decision-making, stress, and social cognition in frontline workers vs. population groups during the COVID-19 pandemic: An explorative study. *Frontiers in Psychology*, 11, 3228

Mazza, M., Lucci, G., Pacitti, F., Pino, M. C., Mariano, M., Casacchia, M., & Roncone, R. (2010). Could schizophrenic subjects improve their social cognition abilities only with observation and imitation of social situations?. *Neuropsychological rehabilitation*, 20(5), 675-703.

Mazza, M., Mariano, M., Peretti, S., Masedu, F., Pino, M. C., & Valenti, M. (2017). The role of theory of mind on social information processing in children with autism spectrum disorders: A mediation analysis. *Journal of Autism and Developmental Disorders*, 47(5), 1369-1379.

Mazza, M., Pino, M. C., Keller, R., Vagnetti, R., Attanasio, M., Filocamo, A., ... & Valenti, M. (2022). Qualitative differences in attribution of mental states to other people in autism and schizophrenia: what are the tools for differential diagnosis?. *Journal of Autism and Developmental Disorders*, 52(3), 1283-1298.

Mazza, M., Pino, M. C., Mariano, M., Tempesta, D., Ferrara, M., De Berardis, D., Masedu, F., & Valenti, M. (2014). Affective and cognitive empathy in adolescents with autism spectrum disorder. *Frontiers in human neuroscience*, 8, 791.

Mazza, M., Pino, M. C., Vagnetti, R., Filocamo, A., Attanasio, M., Calvarese, A., & Valenti, M. (2021b). Intensive intervention for adolescents with autism spectrum disorder: comparison of three rehabilitation treatments. *International journal of psychiatry in clinical practice*, *25*(1), 28–36.

McEvoy, R. E., Rogers, S. J., & Pennington, B. F. (1993). Executive function and social communication deficits in young autistic children. *Journal of child psychology and psychiatry*, *34*(4), 563-578.

McKenzie, K., Milton, M., Smith, G., & Ouellette-Kuntz, H. (2016). Systematic review of the prevalence and incidence of intellectual disabilities: current trends and issues. *Current Developmental Disorders Reports*, 3, 104-115.

McLaurin, W. A., & Farrar, W. S. (1973). Validities of the progressive matrices tests against IQ and grade point average. *Psychological Reports*, 32(3), 803–806.

McNeish, D. M. (2015). Using lasso for predictor selection and to assuage overfitting: A method long overlooked in behavioral sciences. *Multivariate behavioral research*, 50(5), 471-484.

Medda, J. E., Cholemkery, H., & Freitag, C. M. (2019). Sensitivity and specificity of the ADOS-2 algorithm in a large German sample. *Journal of Autism and Developmental Disorders*, 49, 750-761.

Mehling, M. H., & Tassé, M. J. (2016). Severity of autism spectrum disorders: Current conceptualization, and transition to DSM-5. *Journal of autism and developmental disorders*, 46, 2000-2016.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, 50(9), 741.

Metz, C. E. (1978). Basic principles of ROC analysis. Seminars in Nuclear Medicine, 8(4), 283–298. https://doi.org/10.1016/ S0001-2998(78)80014-2 PMID: 112681

Minshew, N. J., & Keller, T. A. (2010). The nature of brain dysfunction in autism: functional brain imaging studies. *Current opinion in neurology*, 23(2), 124-130.

Montes, G., & Halterman, J. S. (2006). Characteristics of school-age children with autism. *Journal of Developmental & Behavioral Pediatrics*, *27*(5), 379- 385

Morrison, K. E., Pinkham, A. E., Kelsven, S., Ludwig, K., Penn, D. L., & Sasson, N. J. (2019). Psychometric evaluation of social cognitive measures for adults with autism. *Autism Research*, *12*(5), 766-778.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of autism and developmental disorders*, 36, 27-43.

Mutluer, T., Aslan Genç, H., Özcan Morey, A., Yapici Eser, H., Ertinmaz, B., Can, M., & Munir, K. (2022). Population-based psychiatric comorbidity in children and adolescents with autism spectrum disorder: A meta-analysis. *Frontiers in Psychiatry*, 13, 856208.

Nah, Y. H., & Poon, K. K. (2011). The perception of social situ- ations by children with autism spectrum disorders. *Autism*, 15(2), 185–203.

Narzisi A, Posada M, Barbieri F, Chiericoni N, Ciuffolini D, Pinzino M, Muratori F. Prevalence of Autism Spectrum Disorder in a large Italian catchment area: a school-based population study within the ASDEU project. *Epidemiol Psychiatr Sci*. 2018;29:e5. https://doi.org/10.1017/S2045796018000483.

Nawaz, S., Lewis, C., Townson, A., & Mei, P. (2023). What does the strange stories test measure? Developmental and within-test variation. *Cognitive Development*, 65, Article 101289.

Newschaffer, C. J., Croen, L. A., Daniels, J., Giarelli, E., Grether, J. K., Levy, S. E., Mandell, D. S., Miller, L. A., Pinto-Martin, J., Reaven, J., Reynolds, A. M., Rice, C. E., Schendel, D., & Windham, G. C. (2007). The epidemiology of autism spectrum disorders. *Annual review of public health*, 28, 235–258.

Nippold, M. A., Ward-Lonergan, J. M., & Fanning, J. L. (2005). Persuasive writing in children, adolescents, and adults. Language, Speech, and Hearing Services in Schools, 36(2), 125–138.

o'Halloran, L., Coey, P., & Wilson, C. (2022). Suicidality in autistic youth: A systematic review and meta-analysis. Clinical Psychology Review, 93, 102144.

Okada, K., Hashimoto, K., Iwata, Y., Nakamura, K., Tsujii, M., Tsuchiya, K. J., Sekine, Y., Suda, S., Suzuki, K., Sugihara, G., Matsuzaki, H., Sugiyama, T., Kawai, M., Minabe, Y., Takei, N., & Mori, N. (2007). Decreased serum levels of transforming growth factor-beta1 in patients with autism. *Progress in neuropsychopharmacology & biological psychiatry*, *31*(1), 187–190

Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: toward a brief form for research and applied settings. *Frontiers in psychology,* 6, 1503.

Osterhaus, C., & Bosacki, S. L. (2022). Looking for the light- house: A systematic review of advanced theory-of-mind tests beyond preschool. *Developmental Review*, 64, 101021.

Ozonoff, S., & Iosif, A. M. (2019). Changing conceptualizations of regression: What prospective studies reveal about the onset of autism spectrum disorder. *Neuroscience & Biobehavioral Reviews*, 100, 296-304.

Ozonoff, S., & Jensen, J. (1999). Brief report: specific executive function profies in three neurodevelopmental disorders. *Journal of Autism and Developmental Disorders*, 29(2), 171–177.

Ozonoff, S., & McEvoy, R. E. (1994). A longitudinal-study of executive function and theory of mind development in autism. *Development and Psychopathology*, 6(3), 415–431.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991a). Executive function deficits in high-functioning autistic individuals —relationship to theory of mind. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 32(7), 1081–1105.

Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., Bryson, S., Carver, L. J., Constantino, J. N., Dobkins, K., Hutman, T., Iverson, J. M., Landa, R., Rogers, S. J., Sigman, M., & Stone, W. L. (2011). Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics*, 128(3), e488–e495

Ozonoff, S., Young, G. S., Landa, R. J., Brian, J., Bryson, S., Charman, T., ... & Iosif, A. M. (2015). Diagnostic stability in young children at risk for autism spectrum disorder: a baby siblings research consortium study. *Journal of Child Psychology and Psychiatry*, 56(9), 988-998.

Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian approaches to autism: Towards volatility, action, and behavior. *Psychological bulletin*, 143(5), 521.

Pegado, F., Hendriks, M. H., Amelynck, S., Daniels, N., Bulthé, J., Masson, H. L., . . .de Beeck, H. O. (2018). Neural rep- resentations behind "social norm" inferences in humans. *Scientific Reports*, 8(1), 1–11.

Pellicano, E. (2010). Individual differences in executive function and central coherence predict developmental changes in theory of mind in autism. *Developmental psychology*, *46*(2), 530.

Pellicano, E., Maybery, M., Durkin, K., & Maley, A. (2006). Multiple cognitive capabilities/deficits in children with an autism spectrum disorder: "weak" central coherence and its relationship to theory of mind and executive control. *Development and Psychopathology*, 18(1), 77–98.

Penn, D. L., Ritchie, M., Francis, J., Combs, D., & Martin, J. (2002). Social perception in schizophrenia: The role of context. *Psychiatry Research*, 109(2), 149–159.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 37(1), 51–87.

Pennington, B.F., Rogers, S.J., Bennetto, L., Griffith, E.M., Reed, D.T., & Shyu, V. (1997), *Validity Tests of the Executive Dysfunction Hypothesis of Autism*. In Russell J., ed., *Autism as an Exectuive Disorder* (pp. 143-178), Oxford: Oxford University Press.

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory-of-mind development for children with deafness or autism. *Child Development*, 76(2), 502–517.

Philip, R. C., Dauvermann, M. R., Whalley, H. C., Baynham, K., Lawrie, S. M., & Stanfield, A. C. (2012). A systematic review and meta-analysis of the fMRI investigation of autism spectrum disorders. *Neuroscience & Biobehavioral Reviews*, 36(2), 901-942.

Pickup, G. J., & Frith, C. D. (2001). Theory of mind impairments in schizophrenia: symptomatology, severity and specificity. *Psychological medicine*, 31(2), 207-220.

Pilkonis, P. A., Choi, S. W., Reise, S. P., Stover, A. M., Riley, W. T., & Cella, D. (2011). Item Banks for Measuring Emotional Dis- tress from the Patient-Reported Outcomes

Measurement Infor- mation System (PROMIS®): Depression, Anxiety, and Anger. Assessment, 18, 263–283.

Pinkham, A. E., Morrison, K. E., Penn, D. L., Harvey, P. D., Kelsven, S., Ludwig, K., & Sasson, N. J. (2020). Comprehensive comparison of social cognitive performance in autism spectrum disorder and schizophrenia. *Psychological Medicine*, 50(15), 2557-2565.

Pino, M. C., & Mazza, M. (2016). The use of "literary fiction" to promote mentalizing ability. PLOS ONE, 11(8), Article e0160254. https://doi-org.univaq.clas.cineca.it/10.1371/journal.pone.0160254

Pino, M. C., Le Donne, I., Vagnetti, R., Tiberti, S., Valenti, M., & Mazza, M. (2024). Using the Griffiths Mental Development Scales to Evaluate a Developmental Profile of Children with Autism Spectrum Disorder and Their Symptomatologic Severity. *Child Psychiatry & Human Development,* 55(1), 117–126. https://doi.org/10.1007/s10578-022-01390-z

Pino, M. C., Mariano, M., Peretti, S., D'Amico, S., Masedu, F., Valenti, M., & Mazza, M. (2020a). When do children with autism develop adequate social behaviour? Cross-sectional analysis of developmental trajectories. *European Journal of Developmental Psychology*, 17(1), 71-87.

Pino, M. C., Mazza, M., Mariano, M., Peretti, S., Dimitriou, D., Masedu, F., Valenti, M., & Franco, F. (2017). Simple Mindreading Abilities Predict Complex Theory of Mind: Developmental Delay in Autism Spectrum Disorders. *Journal of autism and developmental disorders*, 47(9), 2743–2756.

Pino, M. C., Spoto, A., Mariano, M., Granziol, U., Peretti, S., Masedu, F., ... & Vidotto, G. (2018). Formal psychological assessment for autism spectrum disorder diagnosis: A new methodology to build an adaptive testing system. *The Open Psychology Journal*, 11(1).

Pino, M. C., Vagnetti, R., Masedu, F., Attanasio, M., Tiberti, S., Valenti, M., & Mazza, M. (2020). Mapping the network of social cognition domains in children with autism spectrum disorder through graph analysis. Frontiers in Psychiatry, 11, 579339.

Pino, M.C., Mariano, M., Peretti, S., D'Amico, S., Masedu, F., Valenti, M. & Mazza, M. (2018). When do Children with Autism Develop Adequate Social Behaviour? Cross-sectional Analysis of Developmental Trajectories. *European Journal of Developmental Psychology*, 1-17.

Pons, P., & Latapy, M. (2006). Computing communities in large networks using random walks. *J. Graph Algorithms Appl.*, 10(2), 191–218.

Postorino, V., Kerns, C. M., Vivanti, G., Bradshaw, J., Siracusano, M., & Mazzone, L. (2017). Anxiety disorders and obsessive-compulsive disorder in individuals with autism spectrum disorder. *Current psychiatry reports*, *19*(12),

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, 1(4), 515-526.

Preskill, J. (2018). Quantum shannon entropy. In J. Preskill (Ed.), Quantum information (p. 94). *Cambridge University Press*. https://arxiv.org/pdf/1604.07450.pdf

Prieto, L., Alonso, J., & Lamarca, R. (2003). Classical test the- ory versus Rasch analysis for quality of life questionnaire reduction. *Health and Quality of Life Outcomes*, 1(1), 1–13.

Prior, M. R. (1979). Cognitive-abilities and disabilities in infantile-autism—review. *Journal of Abnormal Child Psychology*, 7(4), 357–380.

Prior, M. R., Gajzago, C. C., & Knox, D. T. (1976). Epidemiological-study of autistic and psychotic children in 4 eastern states of Australia. Australian and New Zealand Journal of Psychiatry, 10(2), 173–184.

Prior, M., Sartori, G., & Marchi, S. (2003). Cognizione sociale e comportamento. UPSEL Domeneghini..

Raimo, S., Cropano, M., Roldán-Tapia, M. D., Ammendola, L., Malangone, D., & Santangelo, G. (2022). Cognitive and affective theory of mind across adulthood. *Brain Sciences*,12(7), Article 899.

Rajendran, G., & Mitchell, P. (2007). Cognitive theories of autism. *Developmental review*, 27(2), 224-260.

Ramachandran, V.S., & Oberman, L.M. (2006). Broken Mirrors: A Theory of Autism. *Scientific American*, *295*(5).

Rao, P. A., & Landa, R. J. (2014). Association between severity of behavioral phenotype and comorbid attention deficit hyperactivity disorder symptoms in children with autism spectrum disorders. *Autism*, *18*(3), 272-280.

Rasch, G. (1960). Probabilistic models for some intelligence tests and attainment tests. *Danish Institute for Educational Research, Copenhagen, Denmark.*

Ratto, A. B., Kenworthy, L., Yerys, B. E., Bascom, J., Wieckowski, A. T., White, S. W., Wallace, G. L., Pugliese, C., Schultz, R. T., Ollendick, T. H., Scarpa, A., Seese, S., Register-Brown, K., Martin, A., & Anthony, L. G. (2018). What About the Girls? Sex-Based Differences in Autistic Traits and Adaptive Skills. *Journal of autism and developmental disorders*, 48(5), 1698–1711

Raven, J. (1983). The progressive matrices and mill hill vocabulary scale in western societies. In Human assessment and cultural factors (pp. 107-114). Boston, MA: Springer US.

Raven, J. C., & Court, J. H. (1938). *Raven's progressive matrices*. Western Psychological Services.

Raven, J.C., Raven, J. E., & Court, J. H. (2000). Standard progressive matrices. Psychology Press.

Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate behavioral research*, 14(4), 403-414.

Ricks, D. M., & Wing, L. (1976). Language, communication and the use of symbols. In L. Wing (Ed.), Early childhood autism. New York: Pergamon.

Ritvo, R. A., Ritvo, E. R., Guthrie, D., Ritvo, M. J., Hufnagel, D. H., McMahon, W., Tonge, B., Mataix-Cols, D., Jassi, A., Attwood, T., & Eloff, J. (2011). The Ritvo Autism Asperger Diagnostic Scale-Revised (RAADS-R): a scale to assist the diagnosis of Autism Spectrum Disorder in adults: an international validation study. *Journal of autism and developmental disorders*, *41*(8), 1076–1089.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). PROC: An open- source package for R and S+ to analyze and compare ROC curves. BMC *Bioinformatics*, 12(1), 1–8.

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of autism and developmental disorders*, *31*(2), 131-144.

Roeyers, H., Buysse, A., Ponnet, K., & Pichal, B. (2001). Advancing advanced mind-reading tests: Empathic accuracy in adults with a pervasive developmental disorder. *Journal of child psychology and psychiatry*, 42(2), 271-278.

Ronald, A., & Hoekstra, R. A. (2011). Autism spectrum disorders and autistic traits: a decade of new twin studies. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 156(3), 255-274.

Ronald, A., Simonoff, E., Kuntsi, J., Asherson, P., & Plomin, R. (2008). Evidence for overlapping genetic influences on autistic and ADHD behaviours in a community twin sample.*Journal of Child psychology and Psychiatry*, *49*(5), 535-542.

Roncone, R., Ventura, J., Impallomeni, M., Falloon, I. R. H., Morosini, P. L., Chiaravalle, E., & Casacchia, M. (1999). Reliability of an Italian standardized and expanded Brief

Psychiatric Rating Scale (BPRS 4.0) in raters with high vs. low clinical experience. *Acta Psychiatrica Scandinavica*, 100(3), 229–236.

Rosen, N. E., Lord, C., & Volkmar, F. R. (2021). The Diagnosis of Autism: From Kanner to DSM-III to DSM-5 and Beyond. *Journal of Autism and Developmental Disorders*, 1-18.

Rosenberg, R. E., Law, J. K., Yenokyan, G., McGready, J., Kaufmann, W. E., & Law, P. A. (2009). Characteristics and concordance of autism spectrum disorders among 277 twin pairs. *Archives of pediatrics & adolescent medicine*, 163(10), 907-914.

Rosenberger, L. A., Pfabigan, D. M., Lehner, B., Keckeis, K., Seidel, E. M., Eisenegger, C., & Lamm, C. (2019). Fairness norm violations in anti-social psychopathic offenders in a repeated trust game. *Translational Psychiatry,* 9(1), 1–10.

Russell, J., Jarrold, C., & Hood, B. (1999). Two intact executive capacities in children with autism: Implications for the core executive dysfunctions in the disorder. *Journal of Autism and Developmental Disorders*, *29*(2), 103-112.

Rutter, M. (1972). Childhood schizophrenia reconsidered. *Journal of Autism and Childhood Schizophrenia*, 2(3), 315–337.

Rutter, M. (2011). Research review: Child psychiatric diagnosis and classification: concepts, findings, challenges and potential. *Journal of Child Psychology and Psychiatry*, 52(6), 647-660.

Saffin, J. M., & Tohid, H. (2016). Walk like me, talk like me: The connection between mirror neurons and autism spectrum disorder. *Neurosciences Journal*, *21*(2), 108-119.

Sahl, A. C., Rognlien, H. F., Andreassen, O. A., Melle, I., Ueland, T., & Vaskinn, A. (2023). Theory of mind in schizophrenia: A comparison of subgroups with low and high IQ. *Nordic Journal of Psychiatry*, 77(4), 329–335.

Samson, F., Mottron, L., Soulières, I., & Zeffiro, T. A. (2012). Enhanced visual functioning in autism: An ALE meta-analysis. *Human brain mapping*, 33(7), 1553-1581.

Sanchack, K., & Thomas, C. A. (2016). Autism spectrum disorder: Primary care principles. *American family physician*, 94(12), 972-979.

Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Larsson, H., Hultman, C. M., & Reichenberg, A. (2014). The familial risk of autism. *Jama*, 311(17), 1770-1777.

Sartori, R. (2008). *Psicologia psicometrica*. LED, Milano

Sartori, R., & Pasini, M. (2007). Quality and quantity in test validity: how can we be sure that psychological tests measure what they have to?. *Quality & quantity*, *41*, 359-374.

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research*, 45(1), 73–103.

Sasson, N. J., Pinkham, A. E., Carpenter, K. L., & Belger, A. (2011). The benefit of directly comparing autism and schizophrenia for revealing mechanisms of social cognitive impairment. *Journal of Neurodevelopmental Disorders*, 3(2), 87-100.

Scattoni, M. L., Fatta, L. M., Micai, M., Sali, M. E., Bellomo, M., Salvitti, T., ... & Puopolo, M. (2023). Autism spectrum disorder prevalence in Italy: a nationwide study promoted by the Ministry of Health. *Child and Adolescent Psychiatry and Mental Health*, 17(1), 125.

Schisterman, E. F., Perkins, N. J., Liu, A., & Bondell, H. (2005). Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology,* 16(1), 73–81.

Schmidt, M. F., Rakoczy, H., Mietzsch, T., & Tomasello, M. (2016). Young children understand the role of agreement in establishing arbitrary norms—but unanimity is key. *Child Development*, 87(2), 612–626.

Schopler. E., Van Bourgondien, M.E., Wellman, G.J., Love, S.R. (2010). *The Childhood Autism Rating Scale*. 2nd edn. Los Angeles, CA: Western Psychological Services.

Sebanz, N., Knoblich, G., & Prinz, W. (2005). How two share a task: corepresenting stimulus-response mappings. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1234.

Shanton, K., & Goldman, A. (2010). Simulation theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(4), 527-538.

Sharma, S. R., Gonda, X., & Tarazi, F. I. (2018). Autism spectrum disorder: classification, diagnosis and therapy. *Pharmacology & therapeutics*, 190, 91-104

Sheldon, W. (1942). *The Varieties of Temperament*, New York, Harper and Row.

Shumway, S., Thurm, A., Swedo, S. E., Deprey, L., Barnett, L. A., Amaral, D. G., ... & Ozonoff, S. (2011). Brief report: Symptom onset patterns and functional outcomes in young children with autism spectrum disorders. *Journal of autism and developmental disorders*, 41, 1727-1732.

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, 47(8), 921-929.

Šimundić, A. M. (2009). Measures of diagnostic accuracy: basic definitions. *eJIFCC*, 19(4), 203.

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). Vineland adaptive behavior scales (2nd ed.). Circle Pines, MN: American Guidance Service.

Strunz, S., Westphal, L., Ritter, K., Heuser, I., Dziobek, I., & Roepke, S. (2015). Personality pathology of adults with autism spectrum disorder without accompanying intellectual impairment in comparison to adults with personality disorders. *Journal of Autism and Developmental Disorders*, *45*(12), 4026-4038.

Stuart, E. A., King, G., Imai, K., & Ho, D. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8), 1–28.

Tabachnick, B. et Fidell, LS (2007). Using multivariate statistics (5e éd.).

Tager-Flusberg, H., & Sullivan, K. (1994). A second look at second-order belief attribution in autism. *Journal of autism and developmental disorders*, 24(5), 577-586.

Takumi, T., Tamada, K., Hatanaka, F., Nakai, N., & Bolton, P. F. (2020). Behavioral neuroscience of autism. *Neuroscience & Biobehavioral Reviews*, 110, 60-76.

Taniai, H., Nishiyama, T., Miyachi, T., Imaeda, M., & Sumi, S. (2008). Genetic influences on the broad spectrum of autism: Study of proband-ascertained twins. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 147(6), 844-849

Team, R. C. (2023). R: A language and environment for statistical computing. [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Team, RStudio. (2023). RStudio: Integrated Development for R [Computer software]. RStudio, PBC. http://www.rstudio.com/

Terman, L. M. (1916). *The measurement of intelligence* (Vol. 191). Boston: Houghton Mifflin Company.

Terpstra, T. J. (1952). The asymptotic normality and consistency of kendall's test against trend, when ties are present in one ranking. *Indagationes Mathematicae* (Proceedings), 55, 327–333. https://doi.org/10.1016/S1385-7258(52)50043-X

Thorndike, E. L. (1918). *The nature, purposes, and general methods of measurements of educational products*. In G. M. Whipple (ed.) *The Seventeenth Yearbook of the National Society for the Study of education*, 16-24.

Thorndike, R. M., & Lohman, D. F. (1990). *A century of ability testing*. Riverside Publishing Company.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology,* 58(1), 267-288.

Tick, B., Bolton, P., Happé, F., Rutter, M., & Rijsdijk, F. (2016). Heritability of autism spectrum disorders: a meta-analysis of twin studies. *Journal of Child Psychology and Psychiatry*, 57(5), 585-595.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological methods*, 16(2), 209.

Tin, L., Lui, S., Ho, K., Hung, K., Wang, Y., Yeung, H., Wong, T. Y., Lam, S. M., Chan, R., & Cheung, E. (2018). High-functioning autism patients share similar but more severe impairments in verbal theory of mind than schizophrenia patients. Psychological medicine, 48(8), 1264–1273.

Trembath, D., Vivanti, G., Iacono, T., & Dissanayake, C. (2015). Accurate or assumed: Visual learning in children with ASD. *Journal of autism and developmental disorders*, 45, 3276-3287.

Vagnetti, R., Pino, M. C., Masedu, F., Peretti, S., Le Donne, I., Rossi, R., ... & Mazza, M. (2020). Exploring the social cognition network in young adults with autism spectrum disorder using graph analysis. *Brain and Behavior*, 10(3), e01524.

Valenti, M., Mazza, M., Arduino, G. M., Keller, R., Le Donne, I., Masedu, F., ... & Scattoni, M. L. (2023). Diagnostic assessment, therapeutic care and education pathways in persons with autism spectrum disorder in transition from childhood to adulthood: the Italian National Ev. A Longitudinal Project. *Annali dell'Istituto Superiore di Sanità*, 59(4).

Valenti, M., Mazza, M., Masedu, F. & Vagnetti, R. (2021). *Lezioni di metodologia epidemiologica.* Murgo Editore.

Valenti, M., Vagnetti, R., Masedu, F., Pino, M. C., Rossi, A., Scattoni, M. L., Monica, M., Di Giovanni, C., Attanasio, M., Filocamo, A., Le Donne, I., & Siracusano, M. (2019). Register-based cumulative prevalence of Autism Spectrum Disorders during childhood and adolescence in Central Italy. *Epidemiology Biostatistics and Public Health*, 16(4), e13226.

van der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. Psycho- logical Review, 113(4), 842–861. https://doi.org/10.1037/0033-295X.113.4.842

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321-327.

Vogel, A., Jørgensen, K., & Larsen, I. U. (2022). Normative data for emotion hexagon test and frequency of impairment in behavioral variant frontotemporal dementia, Alzheimer's disease and Huntington's disease. *Applied Neuropsychology: Adult,* 29(1), 127–132.

Volkmar, F. R., Bregman, J., Cohen, D. J., & Cicchetti, D. V. (1988). DSM-III and DSM-III-R diagnoses of autism. *The American journal of psychiatry*, 145: 1404–1408.

Von Neumann, J. (1927). Wahrscheinlichkeitstheoretischer aufbau der quantenmechanik. *Nachrichten von Der Gesellschaft Der Wissenschaften Zu Göttingen, Mathematisch-Physikalische Klasse*, 1927, 245–272.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, 75(2), 523–541.

Wilkinson, M. R., & Ball, L. J. (2012). Why studies of autism spectrum disorders have failed to resolve the theory theory versus simulation theory debate. *Review of Philosophy and Psychology*, 3(2), 263-291.

Williams, D. L., Goldstein, G., Carpenter, P. A., & Minshew, N. J. (2005). Verbal and spatial working memory in autism. *Journal of autism and developmental disorders*, 35, 747-756.

Williams, J. H., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience & Biobehavioral Reviews*, *25*(4), 287-295.

Wing, L. (1981). Asperger's syndrome: a clinical account. *Psychological medicine*, 11(1), 115-129.

Wing, L. (1992). Manifestations of social problems in high-functioning autistic people. In *High-functioning individuals with autism* (pp. 129-142). Boston, MA: Springer US.

Wing, L. (1997). The history of ideas on autism: legends, myths and reality. *Autism*, *1*(1), 13-23.

Wong, V., McGrew, J., & Ruble, L. (2020). Predicting the outcomes of parents of transition-age youth or young adults with ASD. Journal of autism and developmental disorders, 50(8), 2723-2739.

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for Asperger syndrome using the AQ: A preliminary study of its diagnostic validity in clinical practice. *Journal of autism and developmental disorders*, *35*(3), 331-335.

World Health Organization. (1992). The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. World Health Organization.

World Health Organization. (2018). International classification of diseases for mortality and morbidity statistics (11th Revision).https://icd.who.int/browse11/l-m/en

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. Rasch Measurement Transactions, 8, 370–371.

Zalla, T., & Korman, J. (2018). Prior knowledge, episodic control and theory of mind in autism: toward an integrative account of social cognition. *Frontiers in psychology*, 9, 752.

Zambianchi, E. (1994). *Elementi di psicofisica*, Roma, Borla.

Zeidan, J., Fombonne, E., Scorah, J., Ibrahim, A., Durkin, M. S., Saxena, S., ... & Elsabbagh, M. (2022). Global prevalence of autism: A systematic review update. *Autism research*, 15(5), 778-790.

Ziegler, M., & Hagemann, D. (2015). Testing the Unidimensionality of Items. *European Journal of Psychological Assessment*, 31, 231–237.

Ziv, Y., & Sorongon, A. (2011). Social information process- ing in preschool children: Relations to sociodemographic risk and problem behavior. *Journal of Experimental Child Psychology*, 109(4), 412–429.

Ziv, Y., Hadad, B. S., & Khateeb, Y. (2014). Social information processing in preschool children diagnosed with autism spectrum disorder. *Journal of autism and developmental disorders*, 44(4), 846-859.

Lista abbreviazioni

A-ToM = Advanced-Theory of Mind

ABA = Applied Behaviour Analysis

ADHD = Disturbo da Deficit Attentivo e Iperattività

ADOS-2 = Autism Diagnostic Observation Schedule – Second Edition

APA = American Psychiatric Association

AQ = Autism Spectrum Quotient

AS = Affetto Sociale

ASD = Autism Spectrum Disorder

CARS2= Childhood Autism Rating Scale-Second Edition

CFA = Confermative Factor Analysis

CFI = Comparative fit index

CI = confidence interval

CRR = Comportamenti ristretti e ripetitivi

CTT = Classical Test Theory

DFI = Differential Item Functioning

DOC = Disturbo Ossessivo/Compulsivo

DSM = Diagnostic and Statistical Manual for Mental Disorder

EAT = Emotion Attribution Task

EFA = Explorative Factor Analysis

EGA = Exploratory Graph Analysis

FE = Funzioni Esecutuve

IA-CST = Intentions Attribution -Comic Strip Task

ICC = Item Charateristic Curve

ICD = International Classification of Diseases

IIF = Item Information Function

IRT = Item Response Theory

M-CHAT = Modified Checklist for Autism in Toddlers

PDD = Pervasive Developmental Disorders (Disturbi Pervasivi dello Sviluppo)

QI = Quoziente Intellettivo

RAADS-r = Ritvo Autism Asperger Diagnostic Scale-Revised

RMSEA = Root mean square error approximation

ROC = Receiver Operating Characteristic curve

SC = Social Cognition

SRMR= Standardized Root Mean Square Residual

SSD = Schizophrenic Spectrum Disorder

SST = Social Situation Test

ST = Simulation Theory

TCT = Teoria Classica dei Test

TD = Typical Development

TEACCH = Treatment and Education of Autistic and related Communication handicapped Children

TLI =Tucker-Lewis Index

ToM = Teory of Mind (Teoria della Mente)

TT= Theory Theory

VABS-II = Vineland Adaptive Behavior Scale

WCC = Weak Central Coherence (Teoria della Coerenza Centrale Debole)

WHO = World Health Organization