UNIVERSITÁ DEGLI STUDI DI PADOVA

Dipartimento di Psicologia Generale

Corso di Laurea Magistrale in

Neuroscienze e Riabilitazione Neuropsicologica

Tesi di Laurea Magistrale

Triade Oscura della Personalità: utilizzo dell'indice TF-IDF e *Machine Learning* per rilevare la menzogna a livello di domanda singola nel questionario Dark Triad Dirty Dozen

Dark Triad of Personality: use of TF-IDF index and Machine Learning to spotting deception at the single question level in questionnaire Dark Triad Dirty Dozen

Relatore

Prof. Giuseppe Sartori

Correlatrice

Dr.ssa Giulia Melis

Laureanda: **Lara Petri**

Matricola: **1227931**

Anno Accademico 2021/2022

INDICE

INTRODUZIONE

Tutti gli esseri umani mentono. La menzogna affonda le proprie radici fin dall'infanzia e permea la vita quotidiana di ogni essere umano, soprattutto in determinate situazioni, quando mossa dal desiderio di raggiungere certi obiettivi. È possibile esagerare dei problemi dando volontariamente un'immagine negativa di sé stessi (simulazione - *faking bad*) o moltiplicare qualità e meriti dando volontariamente un'immagine positiva di sé (dissimulazione - *faking good*). Mentre il *faking bad* è caratteristico di specifiche situazioni, il *faking good* è comune in tutti i contesti della vita quotidiana. In questo studio ci concentreremo infatti sul concetto della dissimulazione in un contesto particolarmente rilevante dove vengono usati dei test di personalità: il caso di affido di minori in una condizione di separazione dal proprio *partner*. Nella vita reale, infatti, i rispondenti posti in questa situazione potrebbero manipolare le loro risposte al fine di apparire in una luce migliore di fronte agli occhi di chi ha il compito di valutare i risultati dei loro questionari, per manipolare la decisione del Giudice all'interno del *setting* legale di tale valutazione.

Identificare la dissimulazione, soprattutto in ambito forense dove essa assume grande rilevanza, non è un compito facile. Il metodo tutt'oggi più utilizzato è il metodo clinico tradizionale, basato esclusivamente sull'esperienza e la bravura dell'esaminatore, che tuttavia dimostra una correttezza classificatoria che non differisce molto da quella casuale (Rawling, 1992). Per questo motivo, sono stati costruiti strumenti in grado di identificare eventuali soggetti simulatori o dissimulatori, per evitare una diagnosi di disturbo mentale laddove sia presente soltanto un comportamento simulatorio, o il contrario, ovvero classificare come simulazione segni e sintomi che sono riconducibili ad una reale patologia mentale o, ancora, non rilevare un disturbo poiché nascosto dal soggetto. Esistono infatti diversi strumenti per la detezione della simulazione e della dissimulazione che garantiscono un'accuratezza migliore rispetto al metodo clinico tradizionale: i più comuni e frequentemente utilizzati sono le scale di controllo contenute negli strumenti di *assessment* come l'MMPI-2 o il MCMI-III, che indagano la tendenza dell'individuo al *faking good* o al *faking bad*; inoltre alcuni dei test utilizzati a tale scopo sono stati creati *ad hoc* per il contesto forense, mentre altri sono comunemente utilizzati anche in ambito clinico. Tuttavia, il grande limite di questi metodi è che essi non sono in grado di individuare la simulazione o la dissimulazione a livello del singolo *item* di un questionario, in modo tale da poter distinguere le risposte sincere da quelle menzognere. Infatti, questi strumenti individuano solo una tendenza generale a mentire, alla quale segue – nella maggior parte dei casi – l'invalidazione dell'intero questionario.

L'obiettivo di questa ricerca è quello di valutare l'efficacia di due tecniche innovative, appartenenti al campo del *Computer Science,* per rilevare la menzogna a livello del singolo *item* in un questionario *self-report*. La prima tecnica che proponiamo è il *Term Frequency – Inverse Document Frequency* (TF-IDF; Baeza-Yates & Ribeiro-Neto, 1999), usato in ambito informatico nel recupero delle informazioni come modello standard. La seconda tecnica consiste in un modello di *Machine Learning Self-Attention Based Autoencoders* (SABA), un approccio di *unsupervised deep learning* basato su un *Autoencoder* (Devlin et al., 2019; Vaswani et al., 2017). Tali tecniche sono state applicate al questionario *Dirty Dozen* (Jonason & Webster, 2010), nella sua versione italiana validata da Schimmenti et al., (2017). Lo studio si avvale di un campione composto da 493 soggetti ai quali è stato chiesto di svolgere il questionario due volte: la prima mentendo intenzionalmente allo scopo di ottenere l'affido dei figli in un ipotetico contesto legale e la seconda volta rispondendo in modo onesto.

Nel dettaglio, nel primo capitolo verrà presentato un inciso teorico sul costrutto della triade oscura della personalità, sullo strumento *self-report* utilizzato nel presente studio e sul tema specifico del contesto legale della valutazione dell'idoneità genitoriale in relazione alla dissimulazione e alla *dark triad.* Nel secondo capitolo descriveremo il problema della simulazione e della dissimulazione nella valutazione psicologica in ambito forense e clinico, riportando una breve rassegna delle tecniche disponibili per la detezione della menzogna e infine, descriveremo le tecniche innovative utilizzate a questo scopo, fra cui le due metodologie utilizzate per questa ricerca: l'indice TF-IDF e il modello SABA. Nel terzo capitolo descriveremo dettagliatamente la struttura della ricerca condotta "*Triade oscura della personalità: utilizzo dell'indice TF-IDF e Machine Learning per rilevare la menzogna a livello di domanda singola nel questionario Dark Triad Dirty Dozen*". Nel quarto capitolo verranno mostrate le analisi condotte ed i relativi risultati con l'utilizzo dei soli punteggi grezzi, dell'indice TF-IDF e del modello SABA, proponendone un confronto al fine di valutarne l'accuratezza e l'efficacia. Infine, nel quinto capitolo verranno discussi i risultati e verrà proposta un'interpretazione di quanto emerso, unitamente ai limiti di questa ricerca ed ai possibili futuri sviluppi.

CAPITOLO I

"*Man is least himself when he talks in his own person.*
Give him a mask and he will tell you the truth".
Oscar Wilde

1 TRIADE OSCURA DELLA PERSONALITÁ

Recentemente, all'interno della letteratura scientifica psicologica incentrata sulla tassonomia dei tratti e delle conformazioni di personalità esistenti, l'attenzione di molti ricercatori si è spostata sul "lato oscuro della personalità". In particolare, nel 2002 Paulhus e Williams hanno richiamato l'attenzione sulla "triade oscura della personalità" ("*the Dark Triad of Personality",* termine coniato proprio da Paulhus & Williams, 2002), un costrutto usato per descrivere una costellazione di tre variabili della personalità socialmente avverse, ma non patologiche, le quali sono risultate maggiormente rilevanti fra le variabili della personalità citate da Kowalski (2001), nella sua rassegna sui comportamenti avversivi nelle relazioni interpersonali. Infatti, questi tratti hanno la caratteristica di essere molto informativi nel prevedere comportamenti antisociali o aggressivi in tutti i contesti sociali, come ad esempio sul posto di lavoro, in famiglia e in generale nelle relazioni affettive. Nell'ultimo decennio, l'aumento esponenziale di interesse per il lato oscuro della natura umana è ben evidenziato dalla metanalisi di Furnham et al., del 2013, da cui risultano dozzine di studi sulla *Dark Triad* e, secondo Google Scholar, oltre 350 citazioni, con un aumento esplosivo dal 2002 al 2009 (Jonason & Webster, 2010).

I tre tratti oscuri della personalità studiati da Paulhus e Williams (2002) sono: Machiavellismo, Narcisismo subclinico e Psicopatia subclinica. Nella letteratura sui disturbi della personalità il termine "clinico" e "subclinico" è spesso usato in senso contrapposto (Lebreton et al., 2006). Per fare chiarezza, possiamo considerare come campioni clinici quelli che comprendono individui che sono attualmente sotto il controllo clinico o forense e come campioni subclinici quelli che si riferiscono a distribuzioni continue in campioni di popolazioni più ampie (Furnham et al., 2013). Naturalmente, i campioni subclinici includeranno anche casi clinici non diagnosticati, poiché fanno riferimento ad una gamma più ampia di popolazione rispetto ai campioni clinici (Ray & Ray, 1982).

I costrutti di narcisismo e psicopatia hanno origine dalla letteratura e dalla pratica clinica (Furnham & Crump, 2005), mentre il machiavellismo ha un'eziologia diversa, non clinica.

1.1.1 Machiavellismo

Pensando al machiavellismo facciamo riferimento ad una personalità manipolativa e cinica. Questo tratto di personalità prende il nome dal famoso personaggio Niccolò Machiavelli (1469 – 1527), un consigliere politico della famiglia dei Medici del 1500, il quale credeva che la politica non dovesse basarsi su leggi morali, bensì su leggi economiche e funzionali: "*il fine giustifica i mezzi*" è la celebre frase che riassume questo concetto. Alla base del pensiero di Machiavelli c'è la teorizzazione del concetto di fortuna, una forza irrazionale e imprevedibile che domina la realtà e che per essere conquistata richiede cinismo e spietatezza, nonché di essere un "*gran simulatore e dissimulatore*", cioè la necessità di fingere il male e di fingere il bene. Richard Christie, raccolse e selezionò le affermazioni dai libri originali di Machiavelli, trasformandole nella misura di un tratto di personalità. Christie e Geis (1970) fecero quindi una ricerca per valutare quanto le persone fossero d'accordo o in disaccordo con tali affermazioni, scoprendo che il machiavellismo esiste come un costrutto specifico di personalità; in particolare, i soggetti che erano maggiormente d'accordo con queste affermazioni avevano maggiori probabilità di comportarsi in modo freddo e manipolativo, sia negli studi di laboratorio sia nel mondo reale all'interno della propria vita personale (Christie & Geis, 1970). Studi più recenti (Jones & Paulhus, 2009) hanno voluto includere nei riferimenti indispensabili per comprendere le sfumature di questo costrutto, l'opera del filosofo cinese Sun Tzu intitolata "*L'arte della guerra*" (500 d.C). Sia Machiavelli che Sun Tzu hanno delineato nel machiavellico un'attitudine necessaria per poter arrivare al successo basata sulla strategia, sulla manipolazione, sulla pianificazione e sulla freddezza emozionale, che è il fulcro centrale del tratto del Machiavellismo. Infatti, come evidenziato da uno studio di Petrides et al., (2007) individui con alti tratti di machiavellismo sembrano mostrare bassi livelli di intelligenza emotiva, con specifici deficit nelle abilità di comprensione ed espressione delle emozioni (Petrides et al., 2007). Alcune strategie abitualmente messe in atto da individui che possiedono questo tratto di personalità sono:

a) individuazione di obiettivi a lungo termine e pianificazione meticolosa del futuro: sono da evitare comportamenti che possano impedire il raggiungimento dei propri obiettivi (come sconsideratezza, orgoglio, troppa empatia, troppa moralità o discontrollo della rabbia) ed è fondamentale prepararsi ad ogni possibile evenienza o avversità, in modo tale da poter avere sempre un programma ben studiato da poter seguire (Jones & Paulhus, 2011);
b) controllo degli impulsi come chiave per la vittoria nella vita: da prediligere uno stile di comportamento cauto in modo da poter pensare razionalmente a tutti i costi e i benefici

connessi ad ogni azione, accogliendo il rischio solo quando necessario; come conseguenza le uniche azioni antisociali intraprese dai machiavellici sono tipicamente a basso rischio e ad alto beneficio (Jones & Paulhus, 2011);

c) capacità di adattamento e flessibilità a seconda della situazione: non è funzionale seguire uno schema rigido di comportamento, infatti una strategia per essere efficace deve comprendere onestà e benevolenza, ma se il raggiungimento dell'obiettivo lo richiede è necessario anche saper utilizzare l'inganno (Jones & Paulhus, 2011); Machiavelli nel suo libro scrive "*egli deve rimanere nel buono più a lungo possibile, ma in caso di necessità, egli deve essere pronto e prendere la via del male"* (Machiavelli, 1513);
d) creare una reputazione di sé positiva e creare alleanze strategiche: utilizzo della comunicazione persuasiva con l'obiettivo di ottenere il supporto altrui e suscitare negli altri una giusta dose di paura; non è consigliato l'uso di forza o di malvagità eccessive, poiché creano sfiducia e malcontento negli altri, mentre la manipolazione è preferibile perché è più cauta ma allo stesso tempo più efficace per raggiungere i propri scopi (Jones & Paulhus, 2011).

Lo strumento di misura del Machiavellismo più comunemente usato in letteratura è stato costruito da Christie e Geis ed è il *Mach IV Inventory*: un questionario autovalutativo formato da 20 *item* (di cui 10 si riferiscono al Machiavellismo e 10 al non-Machiavellismo) su scala Likert a 5 punti e composto da tre sottoscale: (a) l'uso dell'inganno nelle relazioni interpersonali, (b) una visione cinica della natura umana e (c) la mancanza di moralità. Il *Mach IV* è la misura più comunemente usata per valutare il machiavellismo nella letteratura della *Dark Triad* (Funham et al., 2013).

1.1.2 Narcisismo

Il termine narcisismo, invece, ha origine dalla figura mitologica di Narciso, un cacciatore famoso per la sua bellezza, incredibilmente crudele, in quanto disdegnava ogni persona che lo amava. Secondo la mitologia, Artemide decise di vendicare le sofferenze dei corteggiatori rifiutati da Narciso. Quindi, a seguito della punizione divina, Narciso si innamorò della sua stessa immagine riflessa in uno specchio d'acqua e consumato da questa passione si uccise, annegando nel lago in cui si era specchiato. Infatti, il termine "narcisista" indica comunemente un tipo di persona che ha troppa ammirazione verso sè stessa. Nel Manuale Diagnostico e Statistico dei Disturbi Mentali (*American Psychiatric Association,* 2013) è presente il Disturbo Narcisistico di Personalità (NPD),

da cui l'origine clinica del termine. Nei criteri diagnostici del DSM-5, per il NPD troviamo indicato un modello persistente di grandiosità, necessità di adulazione e mancanza di empatia, con la presenza di cinque o più delle seguenti:

a) la grandiosità (un'esagerata, infondata sensazione della propria importanza e dei propri talenti);
b) la preoccupazione con fantasie di successi senza limiti, influenza, potere, intelligenza, bellezza o amore perfetto;
c) la convinzione di essere speciali e unici e di doversi associare solo a persone di altissimo livello;
d) un bisogno di essere incondizionatamente ammirati;
e) una sensazione di privilegio;
f) lo sfruttamento degli altri per raggiungere i propri obiettivi (manipolazione);
g) la mancanza di empatia;
h) l'invidia degli altri e la convinzione che gli altri li invidino;
i) l'arroganza e la superbia.

Per comprendere a pieno la natura di questo tratto di personalità, prendiamo in considerazione gli studi di due teorici del Narcisismo: Otto Kernberg e Heinz Kohut (Jones & Paulhus, 2011). Benché i due autori abbiano due approcci diversi, entrambi sottolineano che la caratteristica primaria del narcisismo è la presenza di una percezione di sé grandiosa, dalla quale consegue un'attenzione focalizzata su sé stessi e una grave assenza di interesse per chiunque altro (Kernberg, 1975). I narcisisti metterebbero in atto una continua ricerca di oggetti simbolici per confermare la loro grandiosità, probabilmente perché il loro sé in realtà è instabile e fragile (Kohut, 1968). Strettamente legata alla percezione di sé grandiosa è la tipica attitudine comportamentale dei narcisisti tesa a manipolare gli altri, con l'unico scopo di ottenere ammirazione e approvazione per rinforzare il proprio ego e ad allontanare le persone che si rifiutano di rinforzare la propria grandiosità. Tale senso di grandiosità è anche collegato alla caratteristica dei soggetti narcisisti di sentirsi sempre in diritto di avere o di fare ciò che vogliono (Kernberg, 1975). Raramente i narcisisti sono coinvolti in atti criminali, poiché portano solo a vantaggi pratici o strumentali, che sono poco utili alla loro rigida strategia volta a raggiungere un "nobile e sfuggente" obiettivo identitario (Jones & Paulhus, 2011). Inoltre, il narcisismo correla positivamente con l'intelligenza emotiva, per cui gli individui che possiedono questo tratto sarebbero socialmente consapevoli e abili nel percepire chiaramente le proprie emozioni e anche le emozioni degli altri, spiegando perché soggetti narcisisti vengano considerati tendenzialmente più desiderabili (Petrides et al., 2011).

Il narcisismo subclinico o normale, ha origine dagli studi di Raskin e Hall (1979), i quali hanno tentato di delineare una versione subclinica del disturbo di personalità narcisistico, così come definito dal DSM. A partire dalle caratteristiche del NPD (grandiosità, dominanza, superiorità), sono stati costruiti alcuni *item* che sono stati perfezionati su ampi campioni di studenti e assemblati nel *Narcissistic Personality Inventory* (NPI). Il NPI è un questionario autovalutativo formato da 40 *item* a risposta forzata divisi in 7 subscale che riguardano le sette dimensioni del costrutto: l'autorità, l'esibizionismo, la superiorità, il diritto, lo sfruttamento altrui, l'autosufficienza e la vanità (Raskin & Hall, 1988). Il questionario esiste anche in una sua forma breve: NPI-16 (Ames et al., 2006). Il successo della migrazione dal costrutto clinico del narcisismo a quello subclinico dello stesso è ben supportato da una forte ricerca in letteratura (Morf & Rhoodewalt, 2001).

Nonostante la concettualizzazione del narcisismo sublinico, alla base della creazione dell'NPI, consideri il senso di grandiosità come la caratteristica centrale di tale tratto, un crescente numero di ricerche empiriche sta dimostrando l'esistenza di due espressioni fenotipiche del tratto (Cain et al., 2008): da un lato la grandiosità narcisistica e dall'altro l'antitetica vulnerabilità narcisistica, la quale porterebbe l'individuo a sperimentare sensazioni di impotenza, vuoto, bassa autostima e vergogna. Quest'ultima componente sarebbe associata a comportamenti di evitamento a livello relazionale a causa di una ipersensibilità al rifiuto o alla critica. In particolare, le due espressioni fenotipiche del tratto tenderebbero ad alternarsi in modo funzionale: per esempio, quando le strategie rivolte ad accrescere il proprio senso di grandiosità falliscono, i narcisisti potrebbero sperimentare periodi di vulnerabilità, con tendenze ad oscillare più verso una componente del tratto piuttosto che verso l'altra, in base anche alle differenze individuali (Wright et al., 2010).

1.1.3 Psicopatia

Anche il costrutto di psicopatia ha un'origine clinica. Teofrasto, uno degli studenti di Aristotele, fu probabilmente il primo a parlare di psicopatia, riferendosi alle persone affette da questa condizione come "*i senza scrupoli*", descrivendoli come soggetti privi di empatia o coscienza. La storia della psicopatia, però, ha inizio nei primi anni del XIX secolo, quando il medico francese Phillipe Pinel nel 1806 descrisse questa condizione come "*maniaque sans délire*" ("follia senza delirio"), cioè persone che comprendevano il proprio status irrazionale ma continuavano ad agire di conseguenza. Il termine "psicopatia" è stato coniato dallo psichiatra tedesco J.L.A. Koch nel 1888, "*psychopastiche",* che letteralmente significa "anima sofferente" (Kiehl & Hoffman, 2011). Nel

1941, Hervey M. Cleckley, pubblicò "*The Mask of Sanity*": uno studio che fornisce una notevole serie di casi clinici di pazienti, per lo più detenuti, descritti come psicopatici. Cleckley propose sedici caratteristiche per la definizione di psicopatico. Il titolo della sua opera deriva dalla maschera di pseudo normalità che l'autore aveva pensato fosse sottesa a tale disturbo mentale. Tuttavia, una ricerca successiva (1977) sulle caratteristiche attribuite da Cleckley agli psicopatici, concluse che il concetto era stato utilizzato in maniera troppo ampia e dispersiva. Nel 1980, Hare sviluppò la *Psychopathy Checklist* (PCL) basata sulle ipotesi avanzate da Cleckley e successivamente revisionate. Cleckley individuò due diversi fattori interni alla psicopatia: il primo riguarda la tendenza allo sfruttamento, alla manipolazione altrui, alla bassa capacità empatica e il secondo comprende comportamenti autodistruttivi, antisociali, caratterizzati da alta impulsività, ricerca del brivido, basso livello di ansia e facilità alla noia (Hare, 1991; Cleckley, 1976). In particolare, Hare (1996) definisce i soggetti psicopatici come "*predatori intraspecie che usano il fascino, la manipolazione, l'intimidazione e la violenza per controllare il prossimo e soddisfare i propri egoistici bisogni; mancando di morale ed empatia, riescono freddamente a prendere e a fare ciò che vogliono, violando norme e divieti sociali, senza il minimo senso di colpa o rimpianto*". Gli individui caratterizzati da questo tratto di personalità metterebbero in atto comportamenti così maladattivi, impulsivi e disorganizzati da sembrare intrappolati in una dinamica autodistruttiva. Questi soggetti hanno la tendenza a mentire e a mettere in atto comportamenti violenti, rischiosi o irresponsabili anche solo per raggiungere minimi benefici a breve termine o addirittura senza alcun apparente motivo. Attuano rigide strategie comportamentali guidati da un impulso incontrollabile per cui devono avere ciò che vogliono e devono averlo subito, senza alcuna pianificazione o riflessione sui costi per raggiungere i propri obiettivi (Cleckley, 1976) e sembrano incapaci di apprendere dai propri errori (Jones & Paulhus, 2011), Questi soggetti mostrano anche bassi livelli di intelligenza emotiva, mostrando uno specifico deficit nelle abilità di comprensione ed espressione delle emozioni (Petrides et al., 2007). Malgrado questa malevolenza nel loro assetto di personalità, gli individui con queste caratteristiche riescono ad avere un funzionamento che permette loro di vivere in società, tuttavia, si ritrovano a confrontarsi spesso con il sistema giudiziario.

La psicopatia non è riconosciuta nel DSM, ma la sua controparte più vicina è il Disturbo Antisociale di Personalità (Crego & Widiger, 2014). Per la diagnosi di Disturbo Antisociale di Personalità, nel DSM-5, vengono indicate la presenza di tre o più delle seguenti:

a) non conformità alle norme sociali e alla legalità;
b) uso di menzogna o disonestà per profitto o per piacere;

c) incapacità di pianificazione e impulsività (legata alla loro noncuranza verso le conseguenze delle loro azioni);
d) irritabilità e aggressività, con un'elevata frequenza di scontri fisici e assalti;
e) spericolatezza, riguardo le conseguenze delle proprie azioni sugli altri ma anche su se stessi;
f) irresponsabilità lavorativa o finanziaria continuativa;
g) mancanza di rimorso.

Dalla letteratura sulla *Dark Triad*, emerge come la psicopatia venga solitamente misurata con la *Self-Report Psychopathy Scale* (SRP) di Hare (1985) (Furnham et al., 2013), la quale è stata costruita sulla base del *gold standard* per la valutazione della psicopatia in ambito forense, ovvero la *Psychopathy Check List* (Hare, 1991), nel tempo rivisitata e completata attraverso l'affiancamento di un'intervista semistrutturata. Una serie di studi ha confermato la validità di costrutto della scala SRP per la valutazione della psicopatia in campioni subclinici (Forth et al., 1996; Mahmut et al., 2011; Williams et al., 2010). La SRP è andata incontro a diverse revisioni: SRP-I (Hare, 1985); SRP-II (Hare et al., 1989); *Self-Report Psychopathy Scale-III-E* (Williams et al., 2007); e infine, SRP-III (Paulhus et al., 2009). La scala SRP, nella sua terza versione – SRP III - è composta da 64 *item* su una scala a 5 punti ed è stata costruita assemblando gli *item* che differenziano gli psicopatici diagnosticati clinicamente dai non psicopatici (Hare, 1985). Successivamente, è stata validata in campioni di non criminali (Forth et al., 1996). Nella SRP-III la psicopatia è suddivisa in 4 dimensioni: manipolazione interpersonale, affettività ridotta, stile di vita sregolato e comportamenti antisociali (Paulhus et al., in corso di stampa). I punteggi ottenuti alla SRP predicono il comportamento antisociale nelle popolazioni forensi e non forensi (Paulhus et al., 2009). Gli elementi centrali del costrutto includono alti livelli di impulsività e ricerca del brivido, insieme a bassi livelli di empatia (Hare, 1985; Lilienfeld & Andrews, 1996). Ci sono anche altri strumenti utilizzati in letteratura per la misurazione della psicopatia nella *Dark Triad*:

(i) *Psychopathic Personality Inventory* (PPI; Lilienfeld & Andrews, 1996);
(ii) *Levenson Self-Report Psychopathy Scale* (LSRP; Levenson et al., 1995).

Una *review* comparativa ha concluso che i punteggi totali ottenuti alla SRP e al PPI convergono fortemente, mentre la scala LSRP ha più correlazioni con misure del disturbo di personalità antisociale (Hicklin & Widiger, 2005). Comunque, l'adattamento della psicopatia alla sfera subclinica è quello più recente dei tre costrutti (Hare, 1985; Lilienfeld & Andrews, 1996). La migrazione della psicopatia da costrutto clinico a subclinico era stata anticipata da Ray & Ray

(1982). Anche a livello subclinico, la psicopatia è vista come il tratto più malevolo della *Dark Triad* (Rauthmann, 2012).

Da sottolineare come la classificazione psichiatrica sia tradizionalmente categoriale, mentre la valutazione della personalità si basa su modelli dimensionali, come il questionario dei *Big Five* (Costa & McCrae, 1992), utilizzato come prima misura per la valutazione della personalità. Nell'ottica dimensionale, i tratti patologici sono visti come estremi di normalità (Wiggins & Pincus, 1989). La psicopatia, in questa prospettiva, è stata spesso vista come sinonimo di punteggi estremamente bassi sulla gradevolezza e sulla coscienziosità – due delle dimensioni valutate dal *Big Five* (Eysenck & Eysenck, 1985; Miller et al., 2001).

1.2 Dark Triad e principali Modelli Strutturali di Personalità

Data la rilevanza della *Dark Triad* nello studio della personalità normale, non stupisce che essa abbia dei collegamenti con i modelli strutturali di personalità predominanti. I più importanti modelli di personalità sono:

a) *Big Five* - il modello a cinque fattori – (Costa & McCrae, 1991);
b) Modello HEXACO - *Big Six* (Lee & Ashton, 2005).

Il *Big Five* copre cinque ampie dimensioni della personalità (Costa & McCrae, 1991):

1. Estroversione: soggetti con alt livelli di estroversione sono socievoli e attivi; al contrario, soggetti con bassi livelli di estroversione hanno meno interazioni sociali e sono più riservati.
2. Gradevolezza: persone con alti livelli di gradevolezza sono gentili, cooperative, altruiste; al contrario, persone con bassi livelli di gradevolezza sono egoiste, irritabili, intolleranti, impazienti, aggressive.
3. Coscienziosità: soggetti con alti livelli di coscienziosità sono organizzati, persistenti, fiduciosi; al contrario, soggetti con bassi livelli di coscienziosità sono disorganizzati e instabili.
4. Nevroticismo: persone con alti livelli di nevroticismo sono ansiosi, irritabili e stressati; al contrario, persone con bassi livelli di nevroticismo sono calme, meno soggetti ad esperire emozioni forti.
5. Apertura all'esperienza: soggetti con alti livelli di apertura all'esperienza sono curiosi, creativi, disponibili a nuove esperienze; al contrario, soggetti con bassi livelli di apertura all'esperienza sono meno a loro agio con nuove idee ed esperienze in generale.

Le associazioni più rilevanti fra queste cinque dimensioni e i tre membri della triade oscura risultano quelle con gradevolezza e coscienziosità, sul versante negativo dei due tratti (Furnham et al., 2013).

Nel modello HEXACO (Ashton & Lee, 2001; Lee & Ashton, 2005) la sesta dimensione aggiuntiva è stata chiamata Onestà – Umiltà e ricopre il continuum fra comportamento pro-sociale e comportamento antisociale. Tutti e tre i costrutti della *Dark Triad* correlano positivamente con il sesto fattore, sul versante del comportamento antisociale (Lee & Ashton, 2005).

1.3 Correlazioni della triade oscura della personalità

Nonostante le diverse origini di questi tre tratti di personalità subclinica a sfondo malevolo che compongono la *Dark Triad*, essi condividono delle caratteristiche comuni e intercorrelazioni positive nella popolazione normale (Palhus & Williams, 2002) fra:

a) Machiavellismo e Psicopatia - correlazione più forte (>.50) (Furnham et al., 2013);
b) Narcisismo e Psicopatia;
c) Machiavellismo e Narcisismo - correlazione più debole (<.30) (Furnham et al., 2013).

A causa di queste correlazioni positive fra i costrutti, alcuni autori hanno visto i tre tratti di personalità come indistinguibili nei campioni di popolazione normale (McHoskey et al., 1998). Questo ha portato i ricercatori, con il tempo, a combinare i tre tratti della *Dark Triad* in un unico indice globale (Jonason et al., 2010). Tuttavia, applicando all'interno dello stesso campione analisi di regressioni multiple, per determinare i contributi indipendenti di ciascun tratto di personalità, emergono delle differenze tra i tre costrutti (Paulhus & Williams, 2002). Ancora, esistono alcuni studi che dimostrano come alcuni osservatori esterni possono distinguere chiaramente i tre tratti all'interno della *Dark Triad* (Furnham et al., 2013).

Paulhus e Williams (2002) hanno osservato che questi tre tratti oscuri della personalità, misurati in un campione di 245 studenti di Psicologia (65% femmine), non erano equivalenti, ma erano moderatamente intercorrelati, suggerendo quindi di studiarli simultaneamente. I tre costrutti spesso mostrano correlazioni diverse, ma condividono un nucleo comune, che è riconducibile alla manipolazione e all'insensibilità (Furnham et al., 2013): "*tutti e tre condividono caratteristiche socialmente malevole, con tendenze comportamentali verso l'autopromozione, la freddezza emotiva, la doppiezza e l'aggressività*" (Paulhus & Williams, 2002). Paulhus e Williams, nel loro articolo del 2002, hanno cercato di chiarire il concetto di personalità "oscure" all'interno della

normale gamma di funzionamento, cioè a livello subclinico, da cui appunto risultarono le tre variabili più importanti (Machiavellismo, Narcisismo e Psicopatia). I tre costrutti individuati, come già anticipato, sono stati spesso confusi nel tempo, in quanto a livello subclinico di funzionamento essi condividono una somiglianza concettuale e, inoltre, le loro misure comuni si sovrappongono empiricamente. Per separare i membri della triade, Paulhus e Williams (2002) avviarono un programma di ricerca per valutare il grado con cui era possibile distinguere i tre membri fra di loro, sia concettualmente che empiricamente.

I risultati del loro studio mostrarono innanzitutto una differenza di genere: i punteggi dei maschi erano significativamente più alti per tutte e tre le componenti. All'interno del genere, invece, le correlazioni con le variabili esterne erano simili e comunque l'intercorrelazione massima di .50 (fra narcisismo e psicopatia), suggerisce che questi costrutti non possono essere considerati equivalenti nella popolazione normale. L'unico correlato comune nei Big Five con i tre tratti era la bassa gradevolezza. Inoltre:

- psicopatici subclinici si distinguono per un basso nevroticismo - risultato giustificato dal baso livello di ansia tipico del tratto (Hare, 1991);
- machiavellici e psicopatici subclinici hanno bassi punteggi alla coscienziosità;
- narcisisti e psicopatici subclinici risultano positivamente associati all'estroversione e all'apertura all'esperienza;
- narcisisti subclinici riportano basse associazioni positive con le abilità cognitive: i narcisisti tenderebbero a sovrastimare la loro intelligenza; tendenza meno evidente negli psicopatici subclinici e del tutto assente nei machiavellici, i quali infatti hanno una percezione di sé molto realistica e non distorta (Christie & Geis, 1970).

Paulhus e Williams (2002) concludono che la triade, così come misurata nello studio, è composta da costrutti sovrapponibili, ma distinti.

Nella letteratura sulla d*ark triad* sono presenti altri studi che supportano la separazione dei tre tratti di personalità, individuando caratteristiche distintive di ognuno di loro anche in relazione a diversi contesti di vita. Infatti, la valutazione della *Dark Triad* ha enormi potenzialità in molti contesti. Segue una breve esemplificazione di tali relazioni.

1.3.1 Ambiente lavorativo

In relazione al comportamento assunto in ambito lavorativo, O'Boyle et al., (2012) hanno preso in esame gli studi pubblicati tra il 1951 e il 2011 sui tratti di personalità della triade oscura e sul loro impatto sulle prestazioni lavorative e sul comportamento lavorativo controproducente (CWB). Gli autori hanno osservato come tutti e tre i tratti di personalità della triade oscura fossero associati a cattive forme di *leadership* e in generale a cattive prestazioni lavorative (alto CWB) anche per i non *leader* (O'Boyle et al., 2012). In particolare, gli autori hanno evidenziato come riduzioni nella qualità delle prestazioni lavorative fossero costantemente associate ad alti tratti nel machiavellismo e ad alti tratti nella psicopatia. Boddy (2010) si è concentrato sul tratto della psicopatia in ambiente lavorativo, osservando come il clima creato da persone con alti tratti di psicopatia che lavorano nelle aziende sia sostanzialmente tossico, caratterizzato da conflitti, bullismo, aumenti del carico di lavoro e bassi livelli di soddisfazione. Inoltre, i dipendenti di *manager* con alti tratti di psicopatia riceverebbero meno riconoscimenti e ricompense, sperimentando un ambiente di lavoro poco amichevole, con la percezione di alti livelli di ingiustizia e scarsa efficacia comunicativa (Boddy, 2010). Secondo un altro studio, individui con alti livelli di psicopatia non riuscirebbero a svolgere ruoli di alta responsabilità a causa del discontrollo degli impulsi, che potrebbe riversarsi per esempio nella gestione del denaro (Mathieu et al., 2013). Kiazad et al., (2010), invece, si sono concentrati sul machiavellismo, osservando come *manager* con alti tratti di machiavellismo vengano percepiti dai propri dipendenti come avversivi. Amernic et al., (2010) hanno trovato invece una correlazione positiva fra alti tratti di personalità narcisistica e comportamento non etico assunto in ambito lavorativo. Mentre, Furtner et al., (2011) hanno trovato una correlazione positiva tra narcisismo e *leadership* sul posto di lavoro.

1.3.2 Ambiente scolastico

In ambito educativo, invece emerge come la psicopatia sia l'unico predittore indipendente dell'atto di copiare agli esami (Nathanson et al., 2006), mentre il plagio di nei compiti è previsto anche per il machiavellismo (Williams et al., 2010). Il plagio, infatti, a differenza dell'imbroglio in classe che è spesso impulsivo e ad alto rischio, richiede pianificazione, autocontrollo e comporta un livello di rischio più basso (Furnham et al., 2013).

1.3.3 Relazioni amorose e interpersonali

Per quanto riguarda le strategie di accoppiamento e riproduzione, è stato evidenziato dalle ricerche come solo persone con alti tratti di psicopatia mostrerebbero stili di accoppiamento impulsivi (Jones & Paulhus, 2011) e relazioni esclusivamente a breve termine, portando gli individui ad avere promiscuità sessuale, avversione per relazioni a lungo termine e condotte infedeli. Persone con alti tratti di machiavellismo, invece, mostrerebbero massima flessibilità nel loro stile di accoppiamento (Furnham et al., 2013) e sarebbero in grado di modulare il loro approccio relazionale se l'obiettivo è stabilire una relazione a lungo termine e non una relazione strumentale (Jones & Paulhus, 2010).

In relazione al comportamento interpersonale, emerge come tutti e tre i costrutti della triade oscura siano caratterizzati da pregiudizi nei confronti degli immigrati e, più in generale, mostrino un orientamento al dominio sociale (Hodson et al., 2009). È possibile osservare stili interpersonali diversi per i tre membri della triade oscura: in particolare, gli psicopatici avrebbero maggiori probabilità di tatuarsi a scopo intimidatorio (Nathanson et al., 2006) e fare impressioni negative in brevi incontri (Rauthmann, 2012). Coerentemente con gli studi di Christie e Geis (1970) emerge come i machiavellici nutrirebbero maggior cinismo nei confronti degli altri (Rauthmann, 2012). In particolare, uno studio di Aziz (2004) ha trovato come i machiavellici siano molto bravi a manipolare e convincere eventuali acquirenti. Interessante notare come persone presentanti i tre tratti della triade oscura possano essere distinti da caratteristiche facciali, suggerendo così la possibilità di riconoscimento di un segnale di pericolo (Gordon & Platek, 2009; Holzman, 2011). I narcisisti, invece, vedono sè stessi come buoni *leader*, sebbene siano spesso percepiti dagli altri come avversivi (Furtner et al., 2011; Zuroff et al., 2010). L'umorismo, invece, verrebbe usato da tutti e tre i membri della triade oscura come strategia interpersonale (Veselka et al., 2011), con delle differenze: machiavellici e psicopatici sembrano preferire stili di umorismo aggressivo, mentre i narcisisti preferirebbero un tipo di umorismo affiliativo (Martin et al., 2012; Veselka et al., 2010).

1.3.4 Comportamento antisociale

Per quanto riguarda il comportamento antisociale, emerge come gli psicopatici, in confronto ai machiavellici e ai narcisisti, abbiano più probabilità di aver avuto a che fare con il sistema giudiziario (Williams et al., 2001); ciò non stupisce, dal momento che la ricerca sulla psicopatia è

iniziata con gli studi sui criminali recidivi (Cleckley, 1941). Infatti, il tratto della psicopatia ha una robusta capacità predittiva per gli atti delinquenziali. I machiavellici invece, a differenza degli psicopatici, sarebbero più cauti e deliberati nel loro comportamento, non mostrando alcuna associazione con comportamenti aggressivi o vendicativi (Williams et al., 2010). Infine, i narcisisti sono associati ad aggressività solo in conseguenza a provocazione (Williams & Paulhus, 2004).

1.3.5 Moralità

Machiavellismo e psicopatia risultano positivamente associati con un basso livello di moralità, ma con alcune differenze. Infatti, la psicopatia correla negativamente con alte abilità di ragionamento astratto morale, mentre individui con alti tratti di machiavellismo sarebbero capaci di ragionare in termini di moralità sulle loro azioni, adottando anche altre prospettive, ma sceglierebbero egoisticamente di non farlo (Campbell et al., 2009).

1.3.6 Orientamento temporale

Questo costrutto fa riferimento ad un criterio di differenziazione individuale per il quale alcune persone prediligono strategie adattive a lungo termine, mentre altre prediligono strategie orientate all'ottenimento di benefici immediati, o a breve termine. Per quanto riguarda i tre tratti della triade oscura della personalità, punteggi alti di machiavellismo portano gli individui a mettere in atto comportamenti manipolatori ed immorali per raggiungere obiettivi importanti anche se lontani nel tempo, evitando di essere messi a rischio per soddisfazioni immediate ma irrisorie. Al contrario, alti livelli di narcisismo e soprattutto di psicopatia, spingono gli individui a mettere in atto comportamenti manipolatori ed impulsivi per piccoli benefici immediati, senza preoccuparsi delle conseguenze future (Jones & Paulhus, 2011).

1.3.7 Bisogno identitarioù

Infine, il costrutto di "bisogno identitario", cattura la distinzione tra individui che si pongono per lo più obiettivi concreti e di natura strumentale e individui che si pongono obiettivi maggiormente astratti, simbolici e di creazione di significato a partire dal proprio vissuto. Nello specifico, possono essere definiti obiettivi simbolici quegli obiettivi riguardanti l'autostima, lo status sociale e

l'identità, mentre gli obiettivi concreti riguardano beni e servizi (Frankl, 1968). Per quanto concerne i tratti della *dark triad,* è stato evidenziato da alcuni studi come machiavellismo e psicopatia portano a perseguire obiettivi concreti riguardanti ad esempio il sesso, il denaro o il potere, mentre il narcisismo è strettamente legato ad obiettivi astratti ed identitari, come l'ammirazione, il rispetto da parte degli altri o il senso di dominanza (Jones & Paulhus, 2011).

Infine, la ricerca recente si è rivolta anche al lato adattivo, scoprendo contesti in cui i tratti della triade oscura si sono rivelati vantaggiosi (Hogan & Hogan, 2001). Esistono infatti, gli "psicopatici di successo" (Babiak & Hare, 2006; Chatterjee & Hambrick, 2007), così come i "narcisisti di successo" (Chatterjee & Hambrick, 2007), i quali, appunto, hanno successo in determinati contesti grazie alle loro caratteristiche. Secondo Hogan però, la presenza di "tratti oscuri" aiuterebbe le persone ad "andare avanti", ma non necessariamente ad "andare d'accordo" con gli altri sul posto di lavoro.

1.4 Origini della triade oscura della personalità

Negli studi di Pauluhs e Williams (2002) sulle tre componenti della triade oscura, non sono state fatte ipotesi in merito alla loro eziologia; tuttavia, è interessante osservare come i comportamenti tipici della *Dark Triad* siano già evidenti nei giovani di età compresa tra gli 11 e i 17 anni (Lau & Marsee, 2012). Come dimostrato da Vernon et al., (2011), tutti e tre i tratti della triade oscura presentano delle componenti genetiche sottostanti (Petrides et al., 2011; Vernon et al., 2008; Veselka et al., 2011). Solo il machiavellismo condividerebbe una componente maggiormente ambientale (Vernon et al., 2008; Villani et al., 2008) e questo è stato interpretato da Paulhus (2011) come evidenza del fatto che il machiavellismo, più degli altri tratti della triade, possa essere influenzato e modificato dall'esperienza. Esiste anche una teoria evoluzionistica riguardo allo sviluppo della triade oscura nei soggetti: tale teoria risale a Linda Mealey (1995). In questa prospettiva teorica gli individui presenterebbero delle differenze lungo un continuum di strategie riproduttive: quelli che enfatizzano l'accoppiamento sembrano avere una strategia riproduttiva a breve termine, o veloce, mentre quelli che enfatizzano la genitorialità avrebbero una strategia riproduttiva a lungo termine, o lenta (Figueredo, 2007; Rushton, 1985; Jonason et al., 2009). In particolare, gli individui presentanti tratti della triade oscura avrebbero una strategia riproduttiva veloce e sarebbero caratterizzati da deficit nell'autocontrollo, egoismo e manifestazioni antisociali (Furnham et al., 2013).

Sebbene la scelta da parte di Paulhus e Williams (2002) dell'aggettivo "oscuro" (*dark*) per riferirsi alle tre componenti della triade, emerge come in realtà la natura dei tre tratti di personalità non sia solo socialmente avversiva (Hogan & Hogan, 1997), ma ciascun tratto presenta anche la controparte adattiva (Penke et al., 2007). Per esempio, la psicopatia sembra avere dei vantaggi nello stile riproduttivo a breve termine (Jones, 2012). Studi recenti, infatti, suggeriscono che la triade oscura, nel suo insieme, può essere pensata come una strategia sociale a breve termine e di sfruttamento sociale, che potrebbe essersi evoluta per consentire proprio lo sfruttamento in condizioni di suscettibilità dei conspecifici (Jonason et al., 2009).

1.5 Questionari

Di seguito verrà esposto il questionario utilizzato per questo studio insieme ad un'altra variante dello stesso e la sua versione originale.

1.5.1 Dark Triad

Tradizionalmente, la misura standard per valutare la *Dark Triad* è composta da tre diversi strumenti che valutano separatamente ciascuno dei tre tratti indipendenti di personalità:

1. *Mach IV* (Christie & Geis, 1970);
2. *Self-Report Psychopaty Scale* (Hare, 1985);
3. *Narcissistic Personality Inventory* (Raskin & Sala, 1979).

Il questionario completo è molto lungo e richiede molto tempo, essendo composto circa da 100 *item*, a seconda della versione scelta dei tre questionari. Da qui, è risultata la necessità di valutare la *Dark Triad* in modo più semplice ed efficiente. Nel 2010, Jonason e Webster hanno creato una versione molto breve: Dirty Dozen (DD) con solo 12 *item*, cioè 4 *item* per ciascun elemento della triade oscura (Jonason & Webster, 2010). Jones e Paulhus (2014) hanno creato un'altra versione della *Dark Triad* con 27 *item*, chiamata *Short Dark Triad* (SD3).

1.5.2 Dirty Dozen (DD)

Nonostante l'incremento di interesse per la triade oscura, la sua valutazione presentava ancora delle lacune metodologiche. Innanzitutto, con oltre 90 *item* distribuiti su 3 scale, *The Dark Triad* risultava uno strumento di misura inefficiente (Jonason & Webster, 2010), in quanto richiedeva molto tempo e poteva causare affaticamento in alcuni partecipanti. I questionari su larga scala hanno indubbiamente il vantaggio di fornire una pletora di dati, ma d'altro lato, presentano il rischio intrinseco di portare ad errori di risposta derivanti dalla fatica dei partecipanti. Utilizzare strumenti di misura più concisi, invece, ha il vantaggio di eliminare gli *item* ridondanti facendo risparmiare così tempo e fatica, nonché il senso di frustrazione nei partecipanti (Saucier, 1994). Vi è infatti una tendenza crescente nella valutazione psicologica a creare misure concise dei tratti fondamentali della personalità (Burisch, 1984, 1997) che possano essere utilizzati in una varietà di ambiti che richiedono velocità ed efficienza, come: studi di campionamento dell'esperienza, *screening* di massa, test di *prescreening*, studi sul campo, diari quotidiani, studi su popolazioni speciali.. (Jonason & Webster, 2010). Inoltre, ciascuna misura, per ciascun costrutto, presentava propri *biases* di risposta e specifiche limitazioni:

a) *Mach IV* (Christie & Geis, 1970), usato per misurare il machiavellismo, può essere influenzato dalla desiderabilità sociale (Wilson et al., 1996), inoltre l'affidabilità interna non supera .70, un valore considerevolmente basso per una scala di 20 *item* (Carmines & Zeller, 1979).
b) NPI (Raskin & Terry, 1988), usato per misurare il narcisismo, è formato da una serie di domande dicotomiche che possono risultare problematiche (Comrey, 1973).
c) L'utilizzo di diverse scale di valutazione per i tre tratti complica la capacità di misurare la triade oscura, in quanto richiede che i punteggi di ogni scala vengano standardizzati (Jonason et al., 2009).

Per questi motivi, Jonason e Webster (2010), hanno sviluppato e validato una misura più breve della *Dark Triad*: *The Dirty Dozen* (DD). Inizialmente i due autori, a partire dalle misure tradizionali dei singoli tratti della triade, hanno creato 22 *item;* da questo numero iniziale sono stati estratti solo 4 *item* per ogni tratto della triade, in base alle relazioni più forti con i fattori primari dei tratti. Quindi, la DD consta di 12 *item* totali, 4 per ciascuna subscala della triade oscura (Machiavellismo, Narcisismo e Psicopatia). Per ogni *item* della scala bisogna indicare il proprio grado di accordo con le affermazioni utilizzando una scala Likert a 7 punti, in cui 1 implica un forte disaccordo e 7 implica completo accordo. Questo nuovo strumento ha mantenuto la flessibilità necessaria a

misurare i tre costrutti indipendenti fra loro, ma correlati, migliorando la sua efficienza e riducendo il numero di *item* dell'87% (da 91 a 12 *item*). Inoltre, il *Dirty Dozen* consente a tutti e tre i costrutti di essere misurati utilizzando la stessa scala di misura, riducendo così i limiti derivanti dalla necessità di standardizzazione dei punteggi per diverse scale di misura presenti nel questionario originale.

Per creare e validare la DD, Jonason & Webster (2010) hanno realizzato quattro studi, coinvolgendo 1085 partecipanti e tenendo conto del fatto che il nuovo questionario con il numero di *item* ridotti avrebbe dovuto avere le stesse correlazioni che presentava il questionario originale (*Dark Triad*), ovvero:

a) correlazioni negative con la gradevolezza (Paulhus & Williams, 2002);
b) correlazioni positive con strategie riproduttive a breve termine (Jonason et al., 2009);
c) correlazioni positive con l'aggressività (Bushman & Baumeister, 1998; Paulhus & Williams, 2002);
d) punteggi più alti negli uomini in tutti e tre i tratti di personalità rispetto alle donne (Jonason et al., 2010; Jonason et al., 2009).

In particolare, negli studi 1, 2 e 4 sono state esaminate l'affidabilità strutturale del questionario, la validità convergente e la validità discriminante, mentre nello studio 3 è stata esaminata l'affidabilità *test-retest*. In conclusione, dallo studio condotto sono state evidenziate buone proprietà psicometriche del DD, con una buona validità convergente (con NPI, Big Five, strategie riproduttive e aggressività) e con una buona validità discriminante (con l'autostima). Il questionario, inoltre, si è dimostrato affidabile nel tempo, attraverso una serie di studi *test-retest*. Sono quindi state identificate prove coerenti con il fatto che il *Dirty Dozen* misura effettivamente il costrutto latente della triade oscura della personalità misurata con la *Dark Triad*. Infine, sono state osservate le stesse differenze di genere note dalla *Dark Triad*, per cui gli uomini otterrebbero punteggi più elevati a tutti e tre i costrutti, rispetto alle donne (Pauhlus & Williams, 2002; Jonason et al., 2009). Una possibile spiegazione potrebbe essere quella che gli uomini beneficerebbero maggiormente dallo sfruttamento sociale (D.M Buss & Duntley, 2008), mentre sfruttare gli altri comporterebbe un costo maggiore per le donne, poiché le donne dipenderebbero di più dalle reti sociali di quanto non dipendano gli uomini (Jonason et al., 2008).

Di seguito è riportato l'indice *alpha* di Cronbach per il questionario *Dirty Dozen* e le sue subscale. Si specifica che tale indice statistico viene utilizzato per valutare l'affidabilità di un questionario e

può assumere valori compresi tra 0 e 1: valori elevati dell'*alpha* – quindi vicini ad 1 – indicano che è presente un'elevata affidabilità all'interno della dimensione indagata.

- α DD = .83
- α psicopatia = .63
- α machiavellismo = .72
- α narcisismo = .79

I due autori hanno anche misurato le correlazioni tra le scale della DD e gli strumenti utilizzati fino a quel momento per misurare i singoli tratti della triade ed esse si sono rivelate modeste:

- la scala del Machiavellismo correla positivamente con il Mach-IV con r = .34;
- la scala della Psicopatia correla positivamente con il SRP-III con r = .42;
- la scala del Narcisismo correla positivamente con il NPI con r = .46

Per tutte le correlazioni il livello di significatività è pari a p< .01, quindi secondo gli autori DD come misura singola della *dark triad* risulta essere efficiente nella valutazione dei tre tratti, offrendo i due vantaggi della brevità della somministrazione e dell'utilizzo di un'unica scala Likert come metodo di risposta (Jonason & Webster, 2010).

Ricerche successive, tuttavia, hanno criticato la *Dirty Dozen* soprattutto per la sua eccessiva brevità: probabilmente la riduzione degli *item* ha portato alla rimozione di contenuti essenziali per la rappresentazione dei tratti e quindi ad un più basso livello di validità di costrutto. Il fatto che le scale della DD non rispecchiano tutti gli aspetti della *dark triad* è dimostrato anche dalle correlazioni con le misure standard dei tratti (sopra descritte) che sono modeste con il Narcisismo e la Psicopatia e basse con il Machiavellismo (Miller et al., 2012). Inoltre, la scala DD del Machiavellismo ha mostrato forti correlazioni positive con misure di orientamento a breve termine, risultato che ha sollevato molte critiche poiché in completa contraddizione con la concettualizzazione originaria del tratto (Jonason & Tost, 2010).

Il questionario abbreviato *Dirty Dozen* è stato anche tradotto e validato in italiano da Schimmenti et al., (2017) ed utilizza una scala a 5 punti. Questa versione in italiano rappresenta specificamente il test che è stato utilizzato nel presente studio. Si rimanda all'Appendice A per il questionario integrale.

1.5.3 Short Dark Triad (SD3)

Jones & Paulhus (2014), considerando la criticità della DD come unico strumento disponibile per la valutazione della *dark triad*, hanno ideato un nuovo strumento *self-report* capace di misurare tale costrutto: *Short Dark Triad* (SD3). La SD3 consta di 27 *item* divisi equamente nelle tre subscale con modalità di risposta basata su una scala Likert a 5 punti che misura il grado di accordo con le affermazioni esposte. Le tre scale hanno mostrato delle intercorrelazioni positive di effetto moderato, in particolare: il Machiavellismo è correlato positivamente con la Psicopatia, r = .50, e con il Narcisismo, r = .18; la Psicopatia è correlata positivamente con il Narcisismo, r = .34. Tutti e tre i valori sono significativi con $p < .001$. Le scale hanno mostrato anche un'affidabilità modesta con $\alpha = .71$ per il Machiavellismo, $\alpha = .77$ per la Psicopatia e $\alpha = .74$ per il Narcisismo. Uno dei limiti della SD3, riscontrato in ricerche successive a quella originale, riguarda il fatto che la scala del Narcisismo sembrerebbe rappresentare solo la dimensione della Grandiosità narcisistica e non quella della Vulnerabilità, al contrario della DD che copre in maniera moderata entrambe le dimensioni (Maples et al., 2014).

1.6 Dark Triad e propensione alla menzogna

È opportuno specificare che recenti ricerche suggeriscono che sia abbastanza facile per i partecipanti alterare volontariamente i punteggi ai test di personalità, in particolare dissimulando (*faking good* – concetto che verrà approfondito nel seguente capitolo) (Winkelspecht et al., 2006). Infatti, diversi studi hanno mostrato che i tratti della triade oscura sono correlati alla propensione alla bugia. In particolare, nello studio di Jonason et al., (2014) è emerso come soggetti con alti tratti di machiavellismo e di psicopatia fossero più propensi a mentire in diversi contesti e come alti tratti di psicopatia fossero correlati con la sensazione di emozioni positive associate alla menzogna. Ancora, uno studio di Book et al., (2006) ha trovato una correlazione significativa fra psicopatia e *faking good,* ma non con il *faking bad.* Infatti, tratti psicopatici sarebbero collegati ad alti tratti di desiderabilità sociale. Riguardo al machiavellismo, Geis & Moon (1981) hanno trovato una relazione significativa fra machiavellismo e abilità di fingere; in particolare, soggetti machiavellici sarebbero più bravi a simulare e più convincenti rispetto a soggetti con bassi tratti di machiavellismo. Heggestad (2012) ha evidenziato come soggetti machiavellici sarebbero maggiormente disposti a manipolare gli altri, sentendosi meno legati alle regole ed essendo più

propensi a comportamenti falsi come conseguenza di bassi o inesistenti livelli di stress legati a tali comportamenti (MacNeil & Holden, 2006).

1.7 Il contesto civile dell'Affido dei Minori

Uno dei problemi più rilevanti riscontrati nella società contemporanea è il continuo aumento del numero delle separazioni coniugali; esse possono essere di due tipi: Consensuale o Giudiziale.

Nella separazione consensuale i coniugi si accordano circa aspetti futuri della loro relazione, quasi sempre in materia di abitazione, beni e affidamenti dei figli; se il loro accordo non danneggia la salute psicofisica di quest'ultimi, il tribunale rende effettivo il loro accordo. Nella separazione giudiziale, invece, i coniugi non sono riusciti a trovare un accordo e quindi l'ultima parola spetterà al Giudice Istruttore. La custodia dei figli rappresenta spesso il nodo centrale dell'esacerbarsi del conflitto nei casi di separazione giudiziale e, infatti, il Giudice Istruttore, per riuscire a raccogliere il maggior numero di informazioni utili a prendere una decisione nel miglior interesse del minore, spesso dispone una Consulenza Tecnica d'Ufficio circa l'affidamento. In questo ambito, si può affermare che la consulenza psicologica ha come obiettivo centrale l'accertamento di quanto sostenuto da uno dei due genitori riguardo la non idoneità dell'altro ad ottenere o mantenere l'affido condiviso, considerato la regola generale in base al diritto di bi-genitorialità dei figli, opposto all'eccezione dell'affido esclusivo (Gulotta, 2011).

Il Consulente Tecnico d'Ufficio (CTU) nominato dal Giudice deve presentare una relazione nella quale esprime la sua valutazione del caso: all'interno della relazione assume particolare importanza la parte sulla valutazione delle capacità genitoriali degli ex coniugi. Per quanto riguarda questa parte, gli aspetti di cui si deve tener conto secondo Gulotta (2011) sono:

- caratteristiche dell'interazione genitore-figlio;
- motivazione alla genitorialità;
- caratteristiche di personalità, grado di empatia del genitore e storia psicosociale del singolo genitore;
- capacità di comprensione dei bisogni del figlio;
- capacità organizzativa e pratica del genitore (risorse materiali, di tempo e di spazio).

Un altro autore che si è occupato di questo tema è Camerini (2006), il quale propone quattro criteri che il CTU dovrebbe utilizzare per definire uno stile genitoriale sano:

1. Criterio dell'accesso: riguarda in che misura un genitore (generalmente quello con cui il figlio vive abitualmente) interpone difficoltà all'altro genitore nella frequentazione del figlio. Richiama quindi il diritto alla bi-genitorialità del minore,
2. Criterio della competenza genitoriale dei due coniugi: riguarda la qualità della relazione di attaccamento.
3. Criterio dell'attenzione ai bisogni reali dei figli.
4. Criterio della funzione riflessiva: intesa come la capacità da parte di ciascuno dei genitori di attivare riflessioni e di elaborare significati partendo dal proprio punto di vista, per arrivare a mettersi nei panni del figlio e attivare il comportamento più adeguato alla sua tutela. Richiama quindi la capacità di mentalizzazione.

Il criterio dell'accesso è molto importante in quanto legato ad alcune situazioni di estremo conflitto genitoriale in cui potrebbe insorgere la cosiddetta Sindrome di Alienazione Genitoriale (PAS) (Gardner, 1988). Nella PAS un genitore (il genitore alienante) manipola il figlio e lo induce a partecipare alla "campagna di denigrazione e ipercritica" a discapito dell'altro genitore (il genitore alienato). Essa non può essere intesa come una patologia del bambino indotta da uno dei genitori, ma deve essere considerata una patologia relazionale che riguarda almeno tre soggetti (i genitori e un figlio), la quale, però, costituisce un importante fattore di rischio per lo sviluppo di problemi comportamentali o emotivi, ma anche vere e proprie patologie psichiatriche nel bambino (Gulotta, 2011). Alcuni indicatori, sottolineati da Gardner (1988), utili a riconoscere tale sindrome a partire da una valutazione del bambino sono: presenza di razionalizzazioni deboli o superficiali per giustificare il proprio odio nei confronti del genitore alienato; mancanza di normale ambivalenza e senso di colpa nei confronti del genitore alienato; supporto immediato manifestato verso il genitore alienante; affermazione della propria indipendenza di pensiero nella scelta di rifiutare il genitore alienato; utilizzo di frasi e scenari presi in prestito dal genitore alienante; estensione dell'ostilità verso tutta la famiglia del genitore alienato. Statisticamente il genitore alienante è più frequentemente la madre, ma a prescindere dal genere, il genitore alienato solitamente è considerato il "responsabile" della fine della relazione. Non a caso, spesso il comportamento di manipolazione del figlio da parte del genitore alienante è una sorta di vendetta verso l'ex coniuge, con l'obiettivo di creare sofferenza nell'altro, oppure, nel peggiore dei casi, di ottenere un tipo di affidamento esclusivo. Nei casi di PAS grave, infatti, è possibile che il bambino arrivi ad accusare falsamente il genitore alienato di abusi sessuali, sotto suggerimenti espliciti del genitore alienante (Gulotta, 2011).

1.7.1 Valutazione dell'Idoneità Genitoriale e Dissimulazione

Generalmente, nella consulenza tecnica in sede di separazione o di divorzio la valutazione delle capacità genitoriali avviene sia tramite il colloquio clinico sia tramite test psicologici che possono essere specifici - quindi test forensi esplicitamente dedicati ad individuare il genitore più adatto per l'affidamento - oppure aspecifici, quindi non creati per questo ambito, ma adattabili ad esso (Gulotta, 2011).

Un esempio di test specifico è la batteria ACCESS (*A Comprehensive Custody Evaluation Standard System*) di Barry Bricklin & Elliot (1995), formata da 5 test specifici differenti: i primi due test sono usati sui bambini, mentre gli altri tre sono rivolti ai genitori e consentono di individuare (a) le reazioni dei genitori nelle situazioni stressogene di accudimento dei figli (PASS - *Parent Awareness Skills Survey*); (b) il profilo del bambino alla luce delle conoscenze del genitore nei vari settori della vita del minore, permettendo al consulente di valutare l'accuratezza con cui il genitore percepisce il bambino (PPCP - *Parent Perception of Child Profile*); (c) le abilità genitoriali con minori al di sotto dei 5 anni d'età (APSIP - *Assessment of Parenting Skills: Infant and Preschoolers*).

Per quanto riguarda i test aspecifici, in sede di consulenza si utilizzano spesso test psicologici di ambito clinico, quali ad esempio: il *Minnesota Multiphasic Personality Inventory*–2 (MMPI–2), e il *Millon Clinical Multiaxial Inventory* (MCMI-III) (Gulotta, 2011) che verranno discussi nel capitolo II.

Come anticipato, il contesto della valutazione psicologica per l'affido dei minori in seguito ad una separazione è il contesto per eccellenza in cui si riscontrano delle strategie di dissimulazione. Poiché è in gioco la relazione con i propri figli, è logico presumere che le persone sottoposte ad una valutazione per la custodia dei minori siano più che motivate a presentarsi come competenti e responsabili, riducendo al minimo qualsiasi indizio della presenza di problematiche di ogni tipo. Oltre a questa motivazione più semplice e "genuina" dietro la forte tendenza alla dissimulazione, ce ne possono essere altre più "oscure", legate soprattutto al voler essere il genitore presso cui il minore abita stabilmente al fine di: ottenere l'assegno di mantenimento dall'ex coniuge; ottenere una sorta di vendetta sull'ex coniuge che ha voluto la separazione e che quindi ha provocato sofferenza; voler semplicemente usare il proprio figlio come arma per ferire l'ex coniuge. Appare, quindi, evidente la necessità di dover usare strumenti valutativi capaci di controllare il più possibile la variabile della dissimulazione: proprio per questo, l'MMPI-II e il MCMI-III, in quanto dotati di scale di controllo della simulazione e della dissimulazione (vedi capitolo II), sono due degli strumenti più utilizzati in questo campo. In un'indagine nazionale sulle pratiche di valutazione in

ambito di custodia di minori, Ackerman & Ackerman (1997) hanno riferito che l'MMPI-2 è stato lo strumento più utilizzato nella valutazione degli adulti in questi casi. Per quanto riguarda il MCMI-III, benché ci siano meno ricerche riguardo il suo utilizzo nel *setting* forense, anch'esso risulta ampiamente utilizzato nel contesto della valutazione per la custodia dei minori (Ackerman & Ackerman, 1997).

1.8 Genitorialità e Dark Triad

Benché spesso nei quesiti che il Giudice pone al consulente in una causa di affido sia presente la richiesta di una valutazione della personalità, essa non è centrale nella capacità genitoriale: non rientra tra i diritti del figlio o del minore quello di avere genitori con personalità "armonica". La personalità va considerata come variabile importante solo nel momento in cui influisce in maniera fortemente disadattiva sui comportamenti e sulle capacità psicologiche legate alla genitorialità. Non è detto che i tratti di personalità più "negativi", come ad esempio quelli compulsivi o istrionici, vadano a interferire nell'esercizio della genitorialità e lo stesso vale con determinati disturbi di personalità. In alcuni studi è emerso come genitori esaminati in ambito di custodia di minori, tendessero ad ottenere punteggi elevati alla scala del Narcisismo nel MCMI-III, ma ciò non è necessariamente un segno di patologia di personalità, in quanto moderate elevazioni potrebbero indicare sani livelli di fiducia in sé stessi, socialità e tratti adattivi (Millon et al., 1997). Lampel (1999), invece ha suggerito che i genitori, nel contesto di decisione del regime di affidamento, potrebbero pensare che la promozione di sé stesi come "persone popolari, sicure di sé stesse, socievoli, gregarie, meticolose e coscienziose" sia vantaggioso e socialmente accettabile (Lampel, 1999).

Alla luce di queste considerazioni, recentemente molti studiosi si sono interrogati sulla capacità dei tre tratti subclinici di personalità più malevoli che costituiscono la *dark triad* di poter costituire un fattore di rischio per il disgregamento del nucleo familiare e ovviamente per la crescita dei minori in un ambiente non sereno e poco funzionale. Non bisogna dimenticare, inoltre, che i tre tratti sono caratterizzati dalla tendenza alla manipolazione altrui, la quale potrebbe facilmente avere luogo nel contesto di contesa per l'affido dei minori, contesto in cui ci sono tanti vantaggi e svantaggi in gioco. Nello studio di Glenn e collaboratori (2020) emerge come:

- un genitore con un alto tratto del narcisismo non spenderebbe abbastanza tempo ed attenzioni per suo figlio o sua figlia, rispetto a quello che spenderebbe per sé stesso;

- un genitore machiavellico, i cui comportamenti sociali sono costantemente manipolatori, potrebbe manipolare i suoi figli, portando a risultati emotivi come senso di tradimento e mancanza di fiducia nel genitore da parte del figlio o della figlia;
- un genitore con il tratto della psicopatia mostrerebbe una mancanza di risposta emotiva per i figli e potrebbe mettere in atto una forma dura di genitorialità completamente priva di empatia e preoccupazione per il proprio figlio.

Nello stesso studio sono stati messi in relazione i tratti della *dark triad* con gli stili di genitorialità descritti da Baumrind (1966), che sono:

- lo stile autorevole: caratterizzato da una tendenza a responsabilizzare e consultare il proprio figlio o figlia, fissando contemporaneamente limiti fermi;
- lo stile autoritario: segnato da una tendenza ad esercitare il proprio potere sul bambino, spesso in modo duro;
- lo stile permissivo: caratterizzato dalla tendenza a consentire ai bambini di fare ciò che vogliono e ottenere ciò che chiedono, quindi da noncuranza.

I risultati dello studio mostrano come la genitorialità autorevole sia in genere negativamente correlata ai tratti della *dark triad*, mentre la genitorialità autoritaria e quella permissiva siano positivamente correlate ad essi (Glenn et al., 2020).

Nello studio di Clemente et al., (2020), invece, gli autori hanno studiato la relazione tra i tratti della *dark triad* e la tendenza dei genitori a mentire ed usare i figli sia durante il periodo di contesa per l'affido degli stessi con le rispettive valutazioni sia nel periodo successivo alla risoluzione della contesa. I risultati mostrano come tutti e tre i tratti siano dei buoni indicatori della tendenza a voler manipolare le decisioni del Giudice durante una contesa per la custodia dei figli e nello specifico:

- il machiavellismo è un buon indicatore della tendenza a mentire nel *setting* legale per trarre vantaggi concreti dalla separazione, come il collocamento del figlio presso la propria casa o l'affido di mantenimento;
- il narcisismo, a causa probabilmente del forte orgoglio, da una parte è un indicatore di conflitti per la decisione del regime di visite al bambino dopo il divorzio da parte dell'altro genitore non convivente, dall'altra predice un disimpegno nei confronti della cura del figlio nel caso in cui il genitore narcisista non sia quello presso cui il bambino vive stabilmente;
- la psicopatia è un predittore di divorzio e di cattiva relazione matrimoniale, caratterizzata soprattutto da manipolazioni e aggressività.

Infine, un secondo studio di Clemente & Espinosa (2021) ha esaminato la capacità predittiva dei tratti oscuri della personalità per i comportamenti di vendetta in seguito ad uno sbaglio (solitamente tradimento o aggressione) commesso dal *partner*. I risultati mostrano che i tratti della *dark triad* sono predittori efficaci della vendetta e della sua pianificazione, in particolare, la relazione è più forte in primo luogo per la psicopatia e successivamente per il machiavellismo, mentre il narcisismo non sembra essere un predittore significativo.

Da quanto emerge negli studi sopra citati, appare evidente che la presenza nei genitori dei tratti della *dark triad* influenza negativamente la vita dei figli sia in maniera diretta, a causa dello stile genitoriale e della loro strumentalizzazione per i propri fini, sia in maniera indiretta, a causa del contesto di disagio emotivo in cui vivono - sia prima che dopo il divorzio - a causa delle tensioni tra i genitori e dei loro comportamenti vendicativi. Quindi, in generale, machiavellismo, psicopatia e narcisismo influenzano negativamente gli stili e le capacità genitoriali (Baumrind, 1996; Glenn et al., 2020) e i rapporti tra i partner prima e dopo il divorzio (Clemente & Espinosa, 2021), ma aumentano anche la tendenza dei genitori a negare aspetti negativi di sé e a voler manipolare la decisione del Giudice all'interno del *setting* legale della valutazione per l'affido di minori.

Quindi, la valutazione dei tratti della *dark triad* all'interno del contesto della valutazione dell'idoneità genitoriale nelle cause di affido dovrebbe essere implementata sia per riuscire a prevedere eventuali comportamenti dissimulatori e manipolatori dei genitori nei confronti delle decisioni giudiziarie sia per tenere conto della presenza di questi tratti in un genitore nel momento in cui si stabilisce il migliore tipo di affido ed il collocamento del bambino.

In conclusione, come dimostrato da questi studi, la simulazione – e in particolare la dissimulazione - ai testi di personalità è un serio problema (Winkelspecht et al., 2006). Generalmente, i test sono considerati strumenti di valutazione più affidabili rispetto alle interviste cliniche, perché producono risultati accurati e oggettivi. Purtroppo, a causa dell'alto rischio di dissimulazione a tali questionari (*fake good*) questa affidabilità è messa in crisi. Risulta quindi importante minimizzare tale criticità nei test di personalità studiando nuove tecniche di detezione della dissimulazione, la quale risulta una tematica poco approfondita rispetto alla simulazione nei questionari.

CAPITOLO II

"La menzogna ha mille volti e un campo indefinito"

Liliana Dell'Osso

2 SIMULAZIONE E DISSIMULAZIONE

2.1 Prevalenza e concetto di simulazione

Tutti mentiamo. Con *mentire* si intende "alterare la verità, dire il falso con piena consapevolezza." [1] Mentire e ingannare sono comportamenti sociali con cui tutti prima o poi abbiamo a che fare (Bass & Hallingan, 2007). La menzogna è un comportamento che ha da sempre affascinato scrittori e poeti come Shakespeare "*All the world's a stage, and all the men and women merely players*" [2] (*"As you like it",* Atto II Scena VII; 1623), Pirandello con il suo concetto di "maschera" dietro la quale ogni uomo si nasconde per scelta o necessità, ma anche gli psicologi. In ambito psicologico è stato dimostrato come generalmente una persona menta almeno due volte al giorno (Vrij, 2000).

Naturalmente, non tutte le bugie sono uguali: esse possono essere dette a fin di bene – le cosiddette bugie bianche – senza grandi ripercussioni; oppure possono essere dette con fini manipolatori – le cosiddette bugie nere – queste portano a conseguenze negative e ben peggiori. In uno studio di Feldman et al., (2002), ai partecipanti veniva chiesto di presentarsi ad uno sconosciuto in dieci minuti di tempo. Dalle analisi delle presentazioni è emerso come in media, in soli dieci minuti una persona raccontasse almeno tre bugie; bugie non solo verbali, ma anche comportamentali, come un sorriso o un cenno d'assenso fatti senza sincerità (Feldman et al., 2002).

Da dove deriva la menzogna? Secondo Griffith & Mc Daniel (2006), la menzogna affonda le proprie radici nel'evoluzione e possiede delle caratteristiche adattive. Infatti, gli organismi che sono capaci di ingannare il predatore guadagnano un vantaggio competitivo e hanno maggiori possibilità di sopravvivere; per esempio, l'opossum inganna i potenziali predatori fingendosi morto, oppure altri animali raggiungono lo stesso obiettivo imitando l'ambiente circostante (*camouflage*). La necessità di ingannare per vivere è vera non solo per gli animali, ma anche per gli esseri umani. Come anticipato nel primo capitolo, Machiavelli (1513), per esempio, consigliò al suo principe ideale di essere sia volpe che leone: il leone è forte e può difendersi dai lupi, mentre la volpe è

[1] https://www.treccani.it/vocabolario/simulare/.
[2] Tutto il mondo è un palcoscenico e tutti gli uomini e tutte le donne sono solo dei semplici attori.

astuta e non cade nelle trappole. In particolare, secondo Machiavelli il principe deve essere "*gran simulatore e dissimulatore*".

Mentire è un comportamento complesso che comprende processi di inibizione e rievocazione, processi di *working memory* e di teoria della mente. Fingere è una capacità che si sviluppa fin dall'infanzia (Bass & Hallingan, 2007). Non è ancora completamente chiaro quali siano le strutture e i circuiti cerebrali implicati nella menzogna, ma sappiamo che sono diversi rispetto a quelli implicati nella verità e che rispetto a questi presentano una maggiore attivazione (Langleben et al., 2005). Come descritto nel libro "La verità sulla menzogna: dalle origini alla post verità" (Dell'Osso & Conti, 2017), già "*Socrate pur proclamandosi amico della verità, aveva intuito che era più sapiente colui che mente sapendo di mentire, rispetto a colui che è capace di dire soltanto il vero*". Di fatto, questa sua intuizione aveva anticipato di circa 25 secoli ciò che poi è stato documentato tramite le tecniche di *neuroimaging,* ovvero una maggiore attività cerebrale in chi mente. La definizione di "menzogna" rimane però un termine troppo generale per lo scopo del presente elaborato. Infatti, è possibile fingere qualsiasi cosa, in qualsiasi contesto, per obiettivi diversi. Nello specifico, la simulazione si verifica nel caso in cui si vogliano ottenere dei guadagni esterni (Stracciari et al., 2010); per esempio, nell'ambito di un colloquio di lavoro la persona può dire o fare qualcosa che la ponga in maniera tale da fare bella figura, per aumentare la possibilità di ottenere il lavoro. In particolare, in ambito forense questi guadagni esterni insiti nell'atteggiamento simulatorio sono numerosi, a partire dal tentativo di evitare interrogatori e sottrarsi alla partecipazione ai processi per un'incapacità di intendere e di volere (simulando per esempio un disturbo: alcuni disturbi risultano più facili da simulare rispetto ad altri, sulla base di quanto i sintomi siano "intuitivi"), o sfruttando tale fenomeno per ottenere una pena minore o misure alternative rispetto alla detenzione, nel contesto penitenziario.

Proprio perché la menzogna fa parte del comportamento umano, negli ultimi 30 anni si è assistito ad un progressivo e costante interesse da parte dei ricercatori per il costrutto della simulazione, specialmente nell'ambito della neuropsicologia forense. Se consideriamo il periodo che va dal 1990 al 2000, le maggiori riviste di neuropsicologia forense hanno pubblicato 120 lavori sulla simulazione, su un totale di 139 studi, ovvero l'86% di questi (Sweet et al., 2002). Uno dei motivi principali dell'aumento di interesse nei confronti di tale tematica, è la presenza sempre più marcata della simulazione in ambito medico-legale. In uno studio, Rogers et al., (1994) hanno stimato la prevalenza generale della simulazione in *setting* forensi e in *setting* clinici, corrispondente rispettivamente al 15,7% e al 7,4%. Sebbene in letteratura siano presenti molteplici studi che cercano di quantificare la presenza della simulazione in ambito forense ed assicurativo, è importante

sottolineare che stimarne la precisa prevalenza è tutt'altro che facile. In primo luogo, essa sarà sempre una sottostima del fenomeno, in quanto non è possibile risalire alla percentuale di simulatori che non sono stati riconosciuti come tali. In secondo luogo, tale difficoltà deriva dalla definizione stessa del costrutto di simulazione e su quale parte del significato di tale costrutto viene posta l'attenzione, a discapito degli altri.

Quando si fa riferimento al concetto di simulazione, solitamente si tiene in considerazione la produzione intenzionale di sintomi fisici e psicologici falsi e non l'esagerazione, accentuazione o prolungamento di tali. In inglese esiste un termine specifico che descrive l'atteggiamento del fingersi malato: "*malinger".* Etimologicamente significa "fingere malattia per sfuggire al dovere" e sembra derivare dal francese "*malingrer"* ("soffrire"), una parola in gergo che probabilmente significava "fingere di essere malato". Quindi, con il termine "simulazione" facciamo riferimento al significato di *malinger,* ovvero al significato di simulazione di malattia. Tale simulazione può andare in due diverse direzioni: *faking good* e *faking bad.*

2.1.1 Faking good e faking bad

Il *faking bad* (o *malingering*) consiste nell'esagerazione intenzionale di problemi fisici o mentali, oppure nel caso in cui non vi sia alcun disturbo e questo venga creato dal nulla, ponendosi quindi in una luce più negativa della realtà (Sartori et al., 2010). Diversamente dal *faking good,* di cui parleremo di seguito, questo costrutto ha ottenuto maggiori attenzioni a causa dei suoi costi sociali (per esempio in termini di risarcimento assicurativo) che sono più facilmente riconoscibili (Mazza et al., 2010). Esempi di contesti forensi in cui ci si imbatte maggiormente in una simulazione di questo genere possono essere, come già accennato: il caso in cui l'imputato cerca di assicurarsi una pena meno pesante o sfuggire totalmente alla pena sfruttando una condizione di infermità mentale; oppure, nel caso in cui egli voglia ottenere una pensione di disabilità o un risarcimento danni. Il *faking bad* può verificarsi anche in un contesto forense civile, per esempio simulando un disturbo per ottenere un risarcimento dopo un incidente.

Il *faking good* può anche essere definito *dissimulazione* ed è inteso come "la tendenza a dare una descrizione di sé eccessivamente positiva" (Paulhus et al., 2002) o a nascondere un disturbo psicologico. Paulhus et al., (2002), nel loro studio identificarono quattro diverse sfaccettature della dissimulazione: *self-deceptive enhancement* (SDE) e *agency management* che riflettono un *bias* egoistico e *self-deceptive denial* e *communion management* che riflettono invece un *bias* morale. Di

queste componenti, solamente *agency management* e *communion management* sono da considerarsi atti deliberati e consci (Bensch et al., 2019). Tale fenomeno di dissimulazione è riscontrabile sia in contesti forensi che in situazioni di vita quotidiana. Per quanto riguarda i contesti forensi, può verificarsi la condizione per cui una persona considerata pericolosa per la società potrebbe dissimulare la propria pericolosità per ottenere la libertà; in un contesto civile invece, una persona potrebbe nascondere un disturbo psicologico per ottenere l'affido dei minori in caso di separazione tra i genitori. Infatti, in tale contesto entrambi i coniugi hanno come obiettivo la custodia dei figli e il loro scopo è dimostrare al Giudice di essere migliori rispetto al proprio *partner* per ottenere l'affidamento degli stessi. Questo è il caso a cui farà riferimento il suddetto elaborato. Relativamente all'ambito forense appena descritto, numerosi studi dimostrano che i genitori in tali contesti tendono a porsi sotto una luce migliore sottostimando alcuni sintomi psicologici e ottenendo punteggi elevati alle scale di controllo per la desiderabilità sociale, come la scala L del MMPI-II di cui parleremo in seguito (Bathursgt et al., 1997). In uno studio italiano (Roma et al., 2014) è stato dimostrato che le donne tenderebbero maggiormente al *fake good* in tale contesto, in confronto agli uomini; probabilmente a causa di variabili culturali che vedono le donne come le figure di riferimento per le cure genitoriali. Per quanto riguarda i contesti di vita quotidiana per esempio, la dissimulazione si verifica spesso in ambito lavorativo dove il candidato tende a dare una buona impressione allo scopo di ottenere il lavoro. Mazza et al., (2020) hanno riportato che il comportamento dissimulatorio è diffuso in molti contesti, che vanno dal 30% al 50% in ambito lavorativo per quanto riguarda la selezione del personale e fino al 30% in ambito forense. Relativamente all'ambito forense in particolare, Stracciari et al., (2010) anni prima riportano percentuali simili: 30% delle cause civili e circa il 20% delle cause penali. Come vedremo più avanti, mentre la simulazione - nella sua accezione di *fake bad* – è inclusa nel Manuale Diagnostico e Statistico dei Disturbi Mentali (DSM-5), la dissimulazione (*fake good*) non vi è inclusa.

Come appena menzionato, il costrutto di simulazione trova posto anche nella quinta edizione del DSM (*Diagnostic and Statistical Manual of Mental Disorders* – American Psychiatric Association, 2013). Qui, viene definito come "*la presentazione o produzione volontaria di sintomi psichici o fisici esagerati. I sintomi sono prodotti per perseguire uno scopo che è riconoscibile attraverso la comprensione della situazione dell'individuo piuttosto che attraverso la sua psicologia"* (American Psychiatric Association, 2013; p.726). Suddetto Manuale riporta che se qualsiasi combinazione dei seguenti punti fosse presente in un paziente, il clinico dovrebbe considerare l'ipotesi di simulazione:

1. Il paziente è coinvolto in un contesto medico-legale (perizie di accertamento in merito ad un disturbo mentale o se ci sono cause legali che lo riguardano).

2. Una marcata discrepanza tra la disabilità o lo stress riportato dal paziente e le osservazioni oggettive.
3. Il paziente manifesta una mancanza di cooperazione durante una valutazione diagnostica e nell'adempiere alle prescrizioni dei trattamenti ad esso imposti.
4. Il paziente presenta un disturbo antisociale di personalità.

Le due caratteristiche fondamentali cui fare riferimento per una diagnosi differenziale sono: la produzione volontaria e consapevole di sintomi da parte del soggetto e la presenza di incentivi o vantaggi esterni. Inoltre, è importante specificare che secondo il DSM-5 la simulazione non è considerata come una patologia a sé stante con una diagnosi in senso stretto, bensì rientra nelle "*Ulteriori condizioni che possono essere oggetto di attenzione clinica".* Anche se il *faking good* non è incluso nel DSM-5, secondo Stracciari et al., (2010), esso dovrebbe seguire le stesse regole valide per il costrutto di simulazione: in un contesto forense si può parlare di dissimulazione solo se è intenzionale e motivata da importanti incentivi o vantaggi esterni.

Secondo una definizione data da Miller (2015), la simulazione si configura come una variabile continua dove non si può parlare di "presenza" o "assenza" di essa in maniera dicotomica, poiché tale variabile può essere modulata in diversi gradi e a seconda dell'obiettivo postosi (Sartori et al., 2016). Vi sono infatti tre diverse forme di simulazione riconosciute in letteratura:

a) Simulazione "pura": la completa fabbricazione di una sintomatologia senza alcun sintomo reale.
b) Simulazione "parziale": esagerazione grossolana di sintomi già esistenti, in modo da renderli peggiori di quello che realmente sono, per dare l'idea di una severità del disturbo maggiore del reale – o prolungamento dei sintomi che nel tempo si sono attenuati o che addirittura sono scomparsi.
c) Falsa ripresa oppure falso miglioramento di sintomi ancora presenti.

Ancora, Stracciari et al., (2010) distinguono altre due tipologie di simulazione:

1. La simulazione generalizzata: si presenta quando vi è un atteggiamento simulatorio generale nel quale vengono creati o amplificati dei sintomi appartenenti a diverse aree psicopatologiche, come ad esempio la depressione, l'ansia, i deficit cognitivi.
2. La simulazione specifica: costituita dalla descrizione di aspetti patologici più o meno puntuali riguardanti un particolare disturbo.

Un ultimo aspetto da considerare è il *coaching.* Tale termine si riferisce alle situazioni in cui un esperto, come un avvocato o un medico, addestra il proprio assistito periziando prima di una valutazione psicodiagnostica eseguita dalla controparte, suggerendo delle risposte che possano influenzare e modificare la perizia e rendere difficoltosa la detezione della simulazione (Stracciari et al., 2010). Similmente, il *coaching* può verificarsi quando i genitori addestrano i propri figli a simulare o esagerare dei disturbi con fini esterni come dei sostegni economici (Aronson et al., 2001).

Riassumendo, il *faking bad* e il *faking good* possono essere intesi come simulazione e dissimulazione. Una stessa persona può percorrere entrambe le direzioni a seconda degli incentivi cui si trova di fronte. Ad esempio, un traumatizzato cranico può avere gravi problematiche nelle attività quotidiane o al lavoro per ricevere un compenso assicurativo (simulazione - *faking bad*), ma "fa sparire" i sintomi nel momento in cui deve riprendere la patente che gli è stata ritirata (dissimulazione - *faking good*) (Stracciari et al., 2010).

Alla luce di quanto esposto è inevitabile che la simulazione abbia un'altissima e preoccupante prevalenza in ambito psicologico e in particolare, in campo forense. Anche se è difficile avere una stima precisa del fenomeno (Conroy & Kwartner; 2006), in letteratura ci sono diversi dati e generalmente si può identificare un *range* che seppur non preciso rende l'idea della vastità del fenomeno: tra il 20% e il 40% (Greeve et al., 2009; Mittenberg et al., 2003; Fishbein et al., 1999).

2.2 Simulazione e patologie psichiatriche: diagnosi differenziale

In questo paragrafo saranno analizzate le condizioni cliniche da escludere per poter parlare di simulazione. In ambito forense, i vantaggi ottenibili attraverso la simulazione sono molteplici ed è quindi estremamente importante porre molta attenzione alla sua valutazione, al fine di evitare sia di diagnosticare un disturbo mentale laddove vi è solo una volontà simulatoria, sia al fine di classificare come simulazione segni e sintomi che sono invece riconducibili ad una reale patologia mentale. È importante precisare che è possibile parlare di simulazione solamente se vengono escluse eventuali diagnosi di psicopatologia e disturbi psichiatrici con cui essa potrebbe confondersi (Stracciari et al., 2010). Di seguito, descriviamo la differenza fondamentale tra la simulazione e le patologie che presentano una sintomatologia e comportamenti ad essa paragonabili.

Come anticipato nel precedente paragrafo, le due caratteristiche che differenziano la simulazione da altre condizioni sono:

1. L’intenzionalità consapevole del soggetto nella produzione dei sintomi.
2. La presenza di incentivi o vantaggi esterni.

Diversamente, i disturbi mentali (quali il disturbo somatoforme, disturbi dissociativi e altre patologie) non presentano alcuna intenzionalità né incentivi esterni individuabili (Conroy & Kwartner, 2006). Prendendo in considerazione tali punti, Rogers (1997) ha identificato e standardizzato il concetto di simulazione e gli stili di risposta relati ad essa, da escludere nel momento in cui si procede ad una valutazione forense.

- *Simulazione*

Si riferisce alla produzione intenzionale di sintomi fisici o psicologici falsi o grossolanamente esagerati (American Psychiatric Association, 1994) motivata da incentivi esterni. La produzione dei sintomi deve essere volontaria altrimenti dovrebbero essere considerati i disturbi somatoformi. La motivazione deve essere esterna, nel caso contrario andrebbe considerato il disturbo fittizio.

- *Disturbi di conversione e somatizzazione*

La quarta edizione del DSM li definisce come una conversione non volontaria di disturbi mentali in deficit cognitivi e sensoriali, con lo scopo di ottenere sollievo dall’ansia ad essi associata (American Psychiatric Association, 2000). Fanno quindi parte di un gruppo di diagnosi che si basano sulla produzione non intenzionale di sintomi fisici. Si differenziano dalla simulazione in quanto essi sono tipicamente limitati al piano fisico; inoltre, non sono intenzionali, a differenza della simulazione dove si osservano obiettivi esterni e intenzionalità (Slick & Sherman, 2012). Alcuni esempi di tali disturbi sono: il disturbo da ansia di malattia e il disturbo di conversione.

- *Disturbi fittizi*

Nel DSM-5 questo disturbo viene definito come una condizione caratterizzata da produzione o simulazione intenzionale, in sé stessi o altri, di sintomi fisici o psicologici, al fine di acquisire il ruolo di malato (American Psychiatric Association, 2013). Come per la simulazione vi è quindi un’intenzionalità nel mostrare una particolare sintomatologia psichiatrica (Stracciari et al., 2010). Nel caso di disturbo fittizio, però, c’è la volontà di mantenere il ruolo di malato e quindi la motivazione si può definire interna piuttosto che oggettivamente esterna (Slick & Sherman, 2012), quindi non vi sono vantaggi esterni che portano il soggetto a simulare.

Secondo Stracciari et al., (2010) altre condizioni che richiedono attenzione in ambito di valutazione forense sono le seguenti:

- *Disturbi dissociativi*

Tale tipologia di disturbi si presenta in modo simile ai disturbi di conversione e somatizzazione, con la differenza che i disturbi dissociativi presentano sintomi di natura neuropsicologica come disorientamento, amnesia, pseudodemenza e deficit di ragionamento e comprensione, non associabili ad un'oggettiva lesione cerebrale (Stracciari et al., 2010); mentre i secondi presentano sintomi di natura fisica. Anche in questo caso vi è un'assenza di intenzione nella produzione della sintomatologia da parte del soggetto. Per tale motivo, si pone nuovamente l'accento sulla necessità di valutare attentamente l'intenzionalità consapevole da parte del soggetto, come criterio diagnostico necessario per poter parlare di simulazione.

- *Altre categorie diagnostiche*

In letteratura sono presenti altri termini diagnostici che non possiedono una definizione uniforme e standardizzata; si tratta di condizioni che spaziano tra la simulazione, i disturbi fittizi e le forme dissociative o somatoformi vere e proprie. Le più citate sono le seguenti:

a. *Sindrome di Münchaausen*[3]: il soggetto si procura in maniera intenzionale lesioni fisiche con lo scopo di ricevere attenzioni mediche, mantenendo uno stile di vita incentrato sull'ospedalizzazione e sulle cure mediche (Meadow, 1982; Stracciari et al., 2010). Molto spesso tale sintomatologia è accompagnata da pseudologia fantastica, ovvero la tendenza a raccontare bugie esagerate e sintomatologie fantastiche a cui il soggetto stesso può arrivare realmente a credere. Generalmente viene considerata una delle forme più gravi del disturbo fittizio e come questo si differenzia dalla simulazione per la motivazione che al contrario risulta interna.
b. *Sindrome di Münchaausen per procura*: in questa condizione il perpetrante induce disturbi in un'altra persona. Nello specifico, si tratta di sintomi di tipo fisico in assenza di finalità esterne prevalenti, come un vantaggio economico. Da considerare la possibilità di simulazione per procura, motivata quindi da incentivi esterni.

[3]Deriva dal nome del Barone di Münchhausen (Freiherr Karl Friedrich Hieronymus von Münchhausen, 1720-1797). Questo nobile tedesco era noto per raccontare storie inverosimili e fantasiose su se stesso. Nel 1951, il medico britannico Richard Asher descrisse per primo un tipo di autolesionismo, in cui il soggetto si inventava storie, segni e sintomi di malattia. Ricordando il Barone, Asher chiamò questo disturbo "sindrome di Münchhausen" (https://it.wikipedia.org/wiki/Sindrome_di_M%C3%BCnchhausen).

c. *Sindrome di Ganser*: fu descritta per la prima volta da Ganser nel 1898 (Allen & Poste, 1994) sulla base dell'osservazione dei detenuti. Conosciuta anche come "parodia della demenza", "pseudodemenza psicogena" o "stato crepuscolare isterico" (Catanesi, 1995). È osservata prevalentemente in ambiente carcerario, in particolare in detenuti in attesa di esecuzione (Stracciari et al., 2010) e consiste nel crollo generalizzato delle funzioni cognitive superiori, con risposte irragionevoli, amnesia autobiografica lacunare, puerilismo isterico. Al contrario, vi sarebbe un apparente mantenimento dell'orientamento, della comprensione e della coscienza. In questo caso non si può parlare di simulazione ma di un vero e proprio disturbo che rientrava nei disturbi dissociativi presenti nel DSM-IV (Drob&Meehan, 2000).

In ambito neuropsicologico forense è quindi fondamentale attuare una diagnosi differenziale al fine di escludere nella valutazione peritale eventuali forme di psicopatologia e, se necessario, utilizzare i giusti strumenti per identificare una possibile simulazione da parte del soggetto esaminato. In sintesi, i due fattori da tenere in considerazione nella diagnosi differenziale per la simulazione sono:

1. L'intenzionalità consapevole di inventare o peggiorare dei sintomi di disturbi, che possono essere anche già presenti.
2. La presenza di incentivi esterni e individuabili, che nelle altre diagnosi, a differenza della simulazione, presentano motivazioni che si possono definire come interne (per esempio come nei disturbi fittizi e come per la Sindrome di Münchaausen).

Inoltre, si sottolinea l'importanza di svolgere un'accurata anamnesi considerando la storia del paziente e inquadrando il più possibile la situazione in cui egli si trova, al fine di individuare la presenza di eventuali vantaggi economici. Infine, si ricorda che in psichiatria e in psicologia clinica nessuna procedura valutativa può considerarsi completamente obiettiva in quanto tutte si basano sulla partecipazione di un soggetto osservato, che è a sua volta attivo e influenza le risposte e gli esiti delle indagini in base a infinite possibili variabili situazionali.

2.3 Logiche alla base della detezione della simulazione di psicopatologia e di deficit cognitivi

I tentativi di smascherare la menzogna affondano le proprie radici fin dall'antichità, come descritto per esempio nel mito di Ulisse, il quale per evitare di partecipare alla Guerra di Troia, si finse pazzo

arando la sabbia del mare. Poiché Palamede non credeva realmente alla sua follia, per smascherarlo mise il figlio di Ulisse di fronte all'aratro. Ulisse, per non uccidere suo figlio, si fermò immediatamente e così Palamede dimostrò che egli stava solo simulando la sua follia. Tornando in tempi più recenti, a fronte di quanto descritto nei paragrafi precedenti, emerge chiaramente la necessità di avere a disposizione delle tecniche e degli strumenti standardizzati per la detezione della simulazione. Essa, infatti risulta essere particolarmente impegnativa soprattutto in sede di valutazioni psichiatriche, psicopatologiche e neuropsicologiche, durante le quali i sintomi comportamentali, che per loro natura sono molto facili da simulare, risultano essere maggiormente presenti. È possibile quindi fare una differenziazione tra la simulazione di psicopatologia e quella di deficit cognitivi, in cui il soggetto per esempio vuole dimostrare di possedere un Q.I. minore o di avere difficoltà di memoria (Haines & Norris, 1995). Disturbi come ansia e depressione possono essere diagnosticati sulla base di quanto viene riferito dal soggetto e sono quindi particolarmente manipolabili (Monaro et al., 2018). Per questo e per l'importanza di limitare il fenomeno della simulazione, da sempre si sono studiate diverse strategie per arginare tale problema (Sartori et al., 2016). In questo paragrafo verrà fornita una rassegna delle logiche alla base delle tecniche maggiormente utilizzate nel campo dell'identificazione di comportamenti simulatori in ambito forense, seguendo la linea tracciata da Stracciari et al., (2010) nel Manuale "*Neuropsicologia forense"* (Stracciari et al., 2010).

2.3.1 Metodo della correlazione anatomo-clinica

Tale tecnica consiste nel valutare il grado di coerenza fra il dato oggettivo, ovvero il dato neuroradiologico lesionale ottenuto attraverso le tecniche di neuroimmagine (TAC, RM) e il quadro cognitivo funzionale deficitario dell'esaminato (il dato soggettivo). Questo metodo può essere utilizzato solamente nel caso in cui i disturbi riportati dal soggetto in esame posseggano un correlato anatomo-cerebrale riconosciuto in letteratura. Non è perciò applicabile nel caso di sintomatologia psichiatrica. La logica alla base sta nel valutare il deficit espresso e le funzioni residue con particolare attenzione all'oggettivazione dei sintomi, cioè ricercando le basi neurologiche misurabili oggettivamente che supportino quanto viene dichiarato dalla persona. Si possono osservare contraddizioni sia di tipo qualitativo che di tipo quantitativo. Le prime riguardano il fatto che la lesione riscontrata generalmente non corrisponde ai sintomi dichiarati, mentre le seconde prendono in considerazione il fatto che i sintomi sembrano essere di severità maggiore rispetto a quelli corrispondenti alla lesione (Stracciari et al., 2010). Questo metodo è

utilizzato soprattutto per la valutazione del danno psichico in seguito a trauma cranico lieve (Stracciari et al., 2010).

2.3.2 Analisi della psicosintomatologia

In questa categoria si possono includere tutte quelle tecniche e quei metodi di detezione della simulazione che abbiano come scopo una qualche analisi della psicosintomatologia dichiarata dal soggetto. Tra queste troviamo la macrocategoria della "*psicopatologia incoerente*" (Stracciari et al., 2010). Essa consiste nella valutazione dei sintomi propri del disturbo in questione, riportati in letteratura, e nel loro confronto con quelli dichiarati dalla persona, al fine di ricercare discordanze, interpretabili come plausibili indizi di simulazione.

Sempre all'interno di questa categoria possiamo trovare il "*metodo dell'analisi dei sintomi rari*": si tratta dell'identificazione di una serie di sintomi riportati dal soggetto esaminato che apparentemente sembrano descrivere sintomi di disturbi psichiatrici, ma che in realtà sono infrequenti nella popolazione psichiatrica di riferimento (Conroy & Kwartner, 2006). Infatti, spesso il soggetto tende a descrivere un'ampia gamma di sintomi che non appartengono ad una specifica categoria di disturbo, poiché non possiede una rappresentazione mentale chiara del quadro sintomatologico associato tipicamente ad una determinata patologia. Per questo motivo, molto frequentemente i simulatori tendono a riportare durante le valutazioni i sintomi più "ovvi" e comuni, ovvero che spesso sono associati ad uno stato di malattia mentale in generale, anche se essi non riguardano il disturbo mentale che si sta cercando di simulare (Sartori et al., 2017). Un esempio è costituito dalle allucinazioni visive, che sono molto meno frequenti di quelle uditive, ma che vengono segnalate molto spesso dai simulatori (Stracciari et al., 2010). Tale strategia è alla base anche di alcuni questionari, come il *Structured Inventory of Malingering Symptomatology* e il *Structured Interview of Reported Symptoms* – che verranno approfonditi in seguito – i quali contengono dei sintomi non plausibili ma che sembrano poterlo essere e che solitamente il simulatore segnala, con lo scopo di mostrarsi sotto una luce quanto più possibile patologica, quando in realtà questi sintomi sono considerati rari nella psicopatologia reale.

Un'altra logica è quella dell'*Indiscriminant Symptom Endorsement* o anche della combinazione improbabile di sintomi. Con tale definizione si fa riferimento al fatto che spesso il simulatore ha la tendenza a riferire una vasta serie di sintomi, piuttosto che pochi e mirati, non avendo in mente una specifica diagnosi. In altre parole, spesso il soggetto simula tanto e in diversi campi. È importante

quindi esaminare questa possibile sovraesposizione sintomatologica, indice di probabile simulazione, rispetto invece ad una sintomatologia più specifica (Conroy & Kwartner, 2006). Infatti, i simulatori tendono a riportare un numero elevato di sintomi poiché credono che agendo in tale modo possano aumentare la probabilità di essere identificati come affetti da un disturbo mentale reale (Sartori et al., 2017). Allo stesso modo può capitare che il simulatore esponga una combinazione di sintomi che non si riscontra in nessuna diagnosi psicopatologica.

Un'altra logica ancora è quella della bizzarria e surrealtà dei sintomi riferiti. Spesso vengono dichiarati dei sintomi che potrebbero sembrare plausibili in psicopatologia, perché gravi e apparentemente psicotici, ma che risultano in realtà troppo bizzarri e non dimostrabili scientificamente. Ad esempio, uno tra gli *item* del SIMS riguarda la perdita "a guanto" di sensibilità della mano: per quanto possa sembrare un sintomo psicotico, questa perdita di sensibilità non può avvenire su tutta la mano, bensì in maniera laterale, seguendo cioè la direzione dei nervi che attraversano il braccio e raggiungono la mano. Quindi, chiunque dichiari una perdita totale di sensibilità della mano con molta probabilità sta simulando.

Segue poi la logica dei sintomi ovvi. Questa si basa sul fatto che un simulatore tenderà a riportare facilmente tutti quei sintomi che, per conoscenza comune, si sanno appartenere alla popolazione psichiatrica, ma difficilmente riuscirà a riferire quelli specifici di una particolare diagnosi, meno conosciuti se non dai clinici e della popolazione psichiatrica stessa (Conroy & Kwartner, 2006). La maggior parte di queste descrizioni sintomatologiche simulate si basa infatti su rappresentazioni stereotipate di malattie mentali. Ad esempio, sintomi negativi come il comportamento catatonico sono raramente simulati; al contrario, sintomi quali allucinazioni e deliri vengono più frequentemente simulati (Cornell & Hawk, 1989).

Infine, una distinzione importante riguarda la differenza tra sintomi (riferiti dal soggetto) e segni (oggettivamente osservabili). Accade infatti frequentemente che quanto affermato dal soggetto simulatore non corrisponda ai segni che un clinico può vedere ed analizzare (Conroy & Kwartner, 2006).

2.3.3 Symptom Validity Testing (SVT)

Detto anche metodo del livello casuale di scelta o metodo della scelta forzata (Binder & Pankratz, 1987). Tale tecnica viene utilizzata nella detezione di simulazione di deficit neuropsicologici, con lo scopo di attestare la veridicità del sintomo (Stracciari et al., 2010). In particolare, fa riferimento ad

un test di ricordo seguito da riconoscimento a scelta forzata. Nella prima fase vengono presentati uno alla volta una serie di stimoli che dovranno essere memorizzati dal soggetto in esame. La seconda fase invece, consiste nel riferire quale stimolo sia stato precedentemente presentato, attraverso la scelta tra due diversi *item*: lo stimolo già visto e un distrattore. La prestazione del soggetto può variare dal 100% (*performance* perfetta in cui il soggetto sceglie sempre l'*item* familiare) al 50% (espressione di una scelta puramente casuale). Una prestazione sotto il 50% è indice di probabile simulazione, poiché questa viene interpretata come riconoscimento dell'*item* già visto e intenzionale scelta del distrattore, con l'intento di dimostrare un livello di capacità cognitive più basso di quello realmente posseduto (Bianchini et al., 2001).

Per questi test non è richiesta nessuna taratura particolare, in quanto essi sono incentrati su un semplice, intrinseco e sempre valido principio matematico (Stracciari et al., 2010). Infatti, l'applicazione di questa logica permette di utilizzare la statistica binomiale per calcolare la probabilità che l'esaminato sia un simulatore, attraverso l'analisi matematica del suo punteggio al test: minore è il punteggio grezzo ottenuto dal soggetto, più elevata sarà l'accuratezza della diagnosi della simulazione. Infine, il metodo del SVT può essere applicato a qualsiasi test neuropsicologico che preveda una risposta a scelta forzata e può essere applicato anche per la detezione della simulazione dei disturbi sensoriali come cecità, sordità e parestesie tattili.

2.3.4 Floor Effect Strategy

Detto anche metodo del livello minimo, è applicato a test di livello, in cui viene richiesto di dimostrare una certa *performance* cognitiva. La *Floor Effect Strategy* (Rey, 1958; Roger set al., 1993) si basa sulla logica che anche pazienti con demenza medio-grave mantengono un livello di prestazione minimo in test semplici. Sulla base di questa idea, un soggetto che mostra una prestazione troppo scadente in test molto facili, è plausibilmente un mentitore. All'inizio della somministrazione, il clinico presenta la prova inducendo nel soggetto il pregiudizio di maggiore difficoltà del reale, che dovrebbe provocare una prestazione nulla o molto scadente, al di sotto del livello minimo atteso (Stracciari et al., 2010). In realtà il test è molto semplice e potrebbe essere svolto anche da pazienti con gravi problemi mentali. Tale metodo può essere applicato a diverse prove neuropsicologiche, come il *Dot Counting Test* (Rey, 1994; Lezak, 1995) o il *Test of Memory Malingering* (TOMM) (Tombaugh, 1996), che verranno presentati nel proseguo del presente capitolo. Una buona misura di tale metodo è il *15-Item Memory Test* di Rey, un test breve e molto

semplice per la valutazione della memoria a breve termine (Rogers et al., 1993) che verrà descritto nei paragrafi successivi.

2.3.5 Violazione di una legge scientifica

Un'altra logica utile per la detezione della simulazione riguarda la violazione di una legge scientifica. Com'è noto dalla letteratura, vi sono delle leggi scientifiche che regolano e spiegano alcuni aspetti che possono essere valutati dai testi neuropsicologici. In questo caso la prestazione del soggetto in esame viene paragonata alla prestazione standard della popolazione (sana o genuinamente malata). L'emergere di una contraddizione tra il *pattern* deficitario mostrato dal soggetto esaminato e il corrispondente *pattern* atteso, ovvero ciò che ci si aspetta nei soggetti normali o nei gruppi clinici genuini, può essere interpretato come un indizio di simulazione. Un esempio può essere un test a difficoltà crescente: se la prestazione non peggiora gradualmente, come ci si aspetterebbe vista la difficoltà crescente degli *item*, ma avviene in maniera casuale o addirittura diminuisce, probabilmente il soggetto sta simulando. Altre "leggi neuropsicologiche" basate su evidenze scientifiche sono le seguenti: (a) i punteggi ad una prova di riconoscimento sono sempre migliori di quelli ad una prova di rievocazione; (b) le parole ad alta frequenza sono meglio ricordate di quelle a bassa frequenza; (c) le parole concrete si ricordano meglio rispetto alle parole astratte (effetto concretezza); (d) il ricordo libero peggiora quando preceduto da un compito di interferenza (effetto interferenza) (Stracciari et al., 2010).

2.3.6 Metodo degli scenari

Questo metodo è stato ideato da Stenberg e colleghi (1995) per misurare l'intelligenza pratica, in quanto essa sarebbe in grado di prevedere il livello di adattamento di un soggetto alla vita quotidiana. Il metodo degli scenari consiste nella presentazione al soggetto di una serie di specifiche situazione di vita quotidiana; ad esse fanno seguito dei comportamenti più o meno adatti alla situazione descritta che l'esaminato ha il compito di ordinare seguendo una determinata logica spiegata dall'esaminatore. La consegna può ad esempio riguardare il differente livello di dolore che provocano le suddette azioni (dalla meno dolorosa alla più dolorosa), oppure una scala di difficoltà nello svolgimento. In seguito, viene confrontato l'ordine fornito dal periziando con quello del paziente con patologia genuina. Se l'ordine differisce, è probabile che il soggetto stia simulando,

infatti un soggetto che non soffre del disturbo in esame molto difficilmente ordinerà i comportamenti nella stessa maniera di un soggetto realmente malato (Stracciari et al., 2010). Questo metodo è stato applicato alla detezione della depressione simulata (Sartori et al., 2000), al danno psichico da lutto simulato (Biron & Sartori, 2002) e alla detezione della simulazione del colpo di frusta (Sartori et al., 2003).

Come anticipato, le logiche sopraesposte possono essere applicate a molteplici test neuropsicologici, alcuni di questi verranno trattati di seguito.

2.4 Strumenti per la detezione della simulazione di psicopatologia e deficit cognitivi

Vediamo di seguito alcuni strumenti e metodi per la detezione della simulazione della psicopatologia e dei deficit cognitivi.

2.4.1 Metodo tradizionale

Tale metodo consiste nell'applicazione di regole cliniche ed epidemiologiche in contesti forensi. Le strategie appartenenti a questa categoria sono basate per esempio sulla detezione delle discrepanze relative al profilo psicometrico dell'esaminando e sull'analisi qualitativa dei sintomi riportati, considerando la loro incidenza nella popolazione psichiatrica (Sartori et al., 2017). Allo scopo di identificare comportamenti simulatori, il metodo tradizionale (detto anche metodo clinico-idiografico) prevede l'analisi qualitativa della sintomatologia dell'esaminando basata sull'esperienza e l'abilità clinica dell'esaminatore, il quale tramite queste facoltà sarebbe in grado di distinguere sintomi veri da sintomi falsi (Stracciari et al. 2010).

Nella tradizione psichiatrico-forense italiana questo approccio è quello dominante, sebbene non sia così attendibile. Alcuni studi, tra i quali quello di Faust et al., (1988), rilevano la difficoltà nel riconoscimento di un comportamento simulatorio. Nella ricerca appena citata, i risultati dimostrano come nessuno dei neuropsicologi coinvolti sia riuscito a riconoscere i simulatori di lesioni cerebrali dai soggetti che possedevano realmente le lesioni cerebrali (Faust et al., 1988). Secondo la letteratura, infatti, tale metodo di classificazione e valutazione non sembra differire significativamente da quello di una scelta casuale. Gulotta (2002), a tal proposito afferma che "*giudizi fatti da psicologi non si rivelano più validi di quelli fatti da studenti neolaureati; i clinici che possono vantare una maggiore esperienza non sono più accurati dei clinici che hanno*

un'esperienza minore". Da queste evidenze si evince come sia importante lo sviluppo di strumenti utilizzati per la detezione della simulazione nell'ambito della psicologia (Faust et al., 1988).

Hust, nel 1940, affermò che ci sono solamente due evenienze in cui la simulazione è certa. La prima si verifica quando il simulatore crede di non essere osservato e si comporta in maniera genuina dimostrando così una differenza sostanziale con quanto dichiarato; la seconda quando lo stesso confessa. La simulazione certa è quindi un'evenienza rara (Stracciari et al., 2010).

Di seguito distinguiamo gli strumenti utilizzati per la simulazione di psicopatologia da quelli utilizzati per i deficit cognitivi.

2.4.2 Test e questionari

Nella seguente rassegna facciamo una prima distinzione tra (a) gli strumenti che vengono usati principalmente nella normale prassi clinica e psichiatrica e (b) gli strumenti creati ad hoc per la detezione della sintomatologia inattendibile e per i comportamenti simulatori in ambito forense (Stracciari et al., 2010).

Tra i primi troviamo il *Minnesota Multiphasic Personality Inventory-II,* il *Millon Clinical Multiaxial Inventory-II* e il *Personality Assessment Inventory.* Questi test presentano oltre alle scale cliniche, anche delle scale di controllo, le quali permettono di individuare se le risposte date dal soggetto tendano alla desiderabilità sociale (*faking good*) oppure ad un'inesistente o eccessiva psicopatologia (*faking bad*). Di seguito li analizziamo uno alla volta.

- *Minnesota Multiphasic Personality Inventory-2* (MMPI-2)

È stato creato intorno al 1940 dagli autori Hathaway e McKinley, con l'obiettivo di valutare la presenza di problemi psichiatrici nella popolazione e in ambito ospedaliero. È diventato rapidamente uno degli strumenti standardizzati più somministrato nella valutazione dei disturbi di personalità (Butcher, 2010) e risulta essere il test più utilizzato in ambito forense per la detezione della simulazione di psicopatologia (Rogers et al., 2001). Di tale test ne esistono due versioni: quella per adulti (MMPI-II) (Butcher et al., 2001) e quella per adolescenti (MMPI-A) (adattamento italiano di Sirigatti & Pancheri, 2001). Bisogna sottolineare che il suddetto test non permette la rilevazione della simulazione a livello del singolo sintomo, bensì l'atteggiamento simulatorio generale (Stracciari et al., 2010). L'MMPI è composto da 567 *item* dicotomici a risposta vero/falso

e fornisce indicazioni sulla personalità del soggetto e sulla corrispondenza del profilo di risposte con diversi quadri nosografici psichiatrici (Stracciari et al., 2010). Gli *item* sono suddivisi in 10 scale cliniche e viene inoltre valutata e interpretata la combinazione delle due o tre scale che presentano il punteggio più elevato (Mathiesen & Einarsen, 2001):

1. Hs: ipocondria.
2. D: depressione.
3. Hy: hysteria.
4. Pd: deviazione psicopatica.
5. Mf: mascolinità/femminilità.
6. Pa: paranoia.
7. Pt: psicoastenia.
8. Sc: schizofrenia.
9. Ma: ipomaniacalità.
10. Si: introversione sociale.

La particolarità dell'MMPI-2 è la presenza di scale di controllo e di indici che hanno l'obiettivo di valutare l'attitudine del soggetto nei confronti del test, l'attendibilità della sua compilazione e gli eventuali comportamenti simulatori e dissimulatori (Butcher, 2010; Stracciari et al., 2010). Di seguito le presentiamo le scale principali nel dettaglio:

- F (*Frequency*): i sintomi contenuti in questa scala sono sintomi rari e bizzarri; un punteggio elevato (80 punti) può essere indice di esagerazione della gravità dei problemi, tendenza ad ammettere un ampio *range* di sintomi e anche la loro finzione totale (la scala F misura quindi il *fake* bad).
- L (*Lie*): sono presenti una serie di *item* che descrivono piccole scorrettezze comportamentali, infrazioni che chiunque è disposto ad ammettere. Questa scala misura quindi la desiderabilità sociale. Un punteggio alto può essere interpretato come la tendenza del soggetto a nascondere problematiche personali o più in generale a dare un'impressione migliore di sé al reale (la scala L misura quindi il *fake good*).
- K (*Correction*): valuta l'attitudine difensiva del soggetto nei confronti del test e la sua tendenza a minimizzare i problemi. Un alto punteggio può essere interpretato come un atteggiamento difensivo nel tentativo di mettersi in luce migliore rispetto al reale; un basso punteggio indica un tentativo di mostrare molti segni psicopatologici. La scala può essere quindi intesa come una misura del meccanismo di negazione (Butcher, 2010).

- Indice F-K (*Dissimulation Index*): viene ricavato dalla differenza tra il punteggio grezzo alla scala F e alla scala K. In generale, soggetti che riportano molti sintomi hanno alti punteggi alla scala F e bassi punteggi alla scala K, quindi l'indice F-K avrà un valore positivo. Al contrario, soggetti che mostrano un atteggiamento difensivo hanno basi punteggi alla scala F e alti punteggi alla scala K, con un indice F-K negativo. L'accuratezza dell'indice nell'individuare la simulazione è intorno al 90% (Stracciari et al., 2010).

- *Millon Clinical Multiaxial Inventory-III* (MCMI-III)

La terza edizione del MCMI è stata messa a punto nel 1997 (Millon & Davis, 1997). Questo test è formato da 175 *item,* divisi in 14 scale di personalità e 10 scale di sindromi cliniche, più 4 scale di controllo chiamate "*modifier*". Queste ultime servono per individuare la modalità di risposta e il grado di apertura e collaborazione del soggetto per identificare la simulazione e la dissimulazione (Millon & Davis, 1997). In particolare:

- Scala X (*Disclosure Level*): indaga il livello di apertura al test, dimostrando se il soggetto sia cooperativo o meno, se riporta i sintomi, anche quelli più gravi. È simile alla scala K dell'MMPI-2 e può essere considerata una scala che individua sia la simulazione che la dissimulazione.
- Scala Y (*Desirability Gauge*): riguarda l'attitudine di risposta in termini di desiderabilità sociale, quindi misura la resistenza del soggetto a mostrare aspetti di sé poco desiderabili socialmente; la sincerità delle risposte diminuisce con l'aumentare del punteggio (indaga *fake good*). Questa scala è simile alla scala F dell'MMPI-2.
- Scala Z (*Debasement*): indica eventuali esagerazioni o sottostime delle difficoltà comportamentali, emozionali e quotidiane, da parte del soggetto esaminato (Craig, 2014). È un indice di drammatizzazione dei sintomi e quindi generalmente riflette tendenze alla simulazione, contrarie a quelle della scala Y.

Un basso punteggio alla scala X ed un alto punteggio alla scala Y indicano una tendenza del soggetto alla dissimulazione (*fake good*); mentre un alto punteggio sia alla scala X che alla scala Z indicano la tendenza del soggetto a simulare (*fake bad*).

- *Personality Assessment Inventory* (PAI)

Il PAI è stato sviluppato da Leslie Morey nel 1991 ed è composto da 334 *item*. Come gli altri questionari sopraesposti, oltre ad essere utilizzato per la valutazione dei disturbi clinici, permette di indagare i tentativi di simulazione attraverso l'analisi delle risposte fornite dal soggetto (Rogers et al., 2013). La lettura degli *item* del PAI risulta più facile rispetto ad altri test, infatti sembra richiedere un livello di scolarità basso (Edens et al., 2000), di conseguenza viene utilizzato in ambito forense soprattutto nella somministrazione ai detenuti che spesso presentano un livello educativo di basso livello. Di seguito vengono elencati alcuni indici di validità specifici:

- *Negative Impression Scale* (NIM): utile per identificare l'esagerazione dei sintomi e la possibilità di simulazione (Conray & Kwartner, 2006).
- *Malingering Index* (MAL): valutato tramite l'analisi di diverse caratteristiche distribuite all'interno del questionario che permettono di individuare un comportamento di *faking bad,* quindi tentativi di simulazione di disturbi psicologici (Edens et al., 2000).
- *Positive Impression Management* (PIM) e *Defensiveness Index:* permettono una valutazione del grado di difensività nei confronti del test e della sottostima dei problemi presentati (Edens et al., 2000).

Questi test, come già specificato, sono utilizzati nella normale prassi clinica e psichiatrica. Vi sono però anche degli strumenti creati ad hoc per l'identificazione di possibili comportamenti simulatori in ambito forense. Tra i più utilizzati citiamo il SIRS e il SIMS.

- *Structured Interview of Reported Symptoms* (SIRS)

Introdotto da Rogers et al., (1992) è lo strumento standard di riferimento per la valutazione della simulazione a carattere psichiatrico, con elevati coefficienti di validità e attendibilità (Stracciari et al., 2010). Indaga un'ampia gamma di psicopatologie autentiche e di sintomi la cui veridicità è molto improbabile. Non solo misura sintomi rari, improbabili o assurdi, ma anche l'atteggiamento difensivo, quindi è in grado di rilevare uno stile di risposta legato alla dissimulazione e quindi influenzato dalla desiderabilità sociale. È progettato per rilevare gli otto stili di risposta più comunemente associati alla simulazione (Stracciari et al., 2010). Esso è un'intervista strutturata composta da 172 domande suddivise in 13 scale: 8 scale primarie e 5 scale supplementari; 32 domande vengono ripetute in modo tale da verificare la consistenza nelle risposte (Stracciari et al., 2010). Le 8 scale primarie identificano criticamente gli eventuali tentativi di simulazione ed i loro

punteggi vengono classificati nelle seguenti categorie: “rispondenti onesti”, “indeterminati”, “probabili simulatori”, “simulatori certi” (Conroy & Kwartner, 2006). Le 8 scale sono le seguenti:

- *Rare Symptoms* (RS): in analogia con la scala F del MMPI-II, contiene i sintomi plausibili ma generalmente molto rari nella popolazione psichiatrica.
- *Symptom Combinations* (SC): riguarda le combinazioni di sintomi che risultano rare nella popolazione psichiatrica.
- *Improbable and Absurd Symptoms* (IA): contiene i sintomi che risultano assurdi e molto improbabili anche nella popolazione psichiatrica.
- *Blatant Symptoms* (BL): riguarda i sintomi eclatanti, ovvero i sintomi che anche se plausibili per la popolazione psichiatrica, tendono ad essere quantitativamente riferiti in modo maggiore dai simulatori.
- *Subtle Symptoms* (SU): contiene sintomi reali e diffusi ma che tendono ad essere trascurati dai simulatori per il contenuto apparentemente poco grave.
- *Selectivity of Symptoms* (SEL): di solito il soggetto simulatore dimostra un *range* troppo vasto di sintomi rispetto a quello che realmente si potrebbe riscontrare nel vero malato, il quale generalmente conosce e riferisce solamente una parte dei problemi psicologici e dei sintomi.
- *Severity of Symptoms* (SEV): indaga l’eccessivo numero di sintomi severi riportati dal simulatore rispetto alla popolazione psichiatrica.
- *Reported versus Observed Symptoms* (RO): riguarda i sintomi che vengono riportati dal simulatore ma su cui l’esaminatore può avere un riscontro diretto attraverso l’osservazione dei comportamenti (Stracciari et al., 2010).

Di seguito elenchiamo le 5 scale supplementari (Rogers et al., 2015):

- *Direct Appraisal of Honesty* (DA).
- *Defensive Symptoms* (DS).
- *Overly Specified Symptoms* (OS).
- *Symptom Onset and Resolution* (SO).
- *Inconsistency of Symptoms* (INC).

Per quanto riguarda l’interpretazione dei punteggi, oltre ai metodi suggeriti dal manuale del test, sono stati pubblicati in letteratura alcuni *cut-off* aggiuntivi che hanno permesso di identificare i simulatori con gradi di accuratezza molto elevati, raggiungendo infatti il 97% (Stracciari et al., 2010).

- *Structured Inventory of Malingered Symptomatology* (SIMS)

Venne introdotto da Smith e Burger nel 1997 ed è un questionario *self-report* specifico per la detezione della simulazione di sintomatologie psichiatriche e cognitive, composto da 75 *item* a risposta dicotomica vero/falso riguardanti sintomi e affermazioni implausibili. La sua costruzione si basa sulla logica secondo cui il simulatore tenderà a prediligere sintomi bizzarri, rari o atipici (Van Impelen et al., 2014), mentre in un profilo psichiatrico veritiero dovrebbe essere riscontrabile solamente una piccola quantità di *item* molto atipici. È costituito da 5 sottoscale, ciascuna contenente 15 *item*:

- *Low Intelligence* (LI).
- *Affective Symptoms* (AF).
- *Neurologic Impairment* (N).
- *Psychosis* (P).
- *Amnestic Symptoms* (AM).

Le scale P, AF, N e AM valutano la presenza di sintomi rari, mentre la scala LI è composta da *item* che ci si aspetta vengano sbagliati solamente da coloro che presentano determinati deficit cognitivi (Heinz & Purisch, 2006). Anche se queste scale non permettono una discriminazione tra simulatori e onesti più efficace rispetto al punteggio totale del SIMS, riescono a fornire una migliore valutazione qualitativa inerente al tipo di sintomi che vengono maggiormente simulati (Smith & Burger, 1997). Gli autori Van Impelen et al., (2014) hanno dimostrato che questo questionario risulta essere sufficientemente robusto nei confronti di un soggetto addestrato alle risposte da dare. Per quanto riguarda lo *scoring*, il test fornisce un punteggio totale per la probabile simulazione di disturbi psichici la cui sensibilità è circa del 97% (Glenn & Gary, 1997). Un esempio di *item* contenuto nel questionario, relativo a patologie psichiatriche, è "*Quando la mia depressione diventa troppo grave, esco a fare una passeggiata o a fare un po' di moto per ridurre la tensione":* ad un clinico esperto appare evidente che un paziente con depressione grave non metterebbe mai in atto un comportamento simile.

- *Balanced Inventory of Desirable Responding* (BIRD)

Per quanto riguarda i test di personalità, uno dei metodi più usati per la detezione della dissimulazione è l'inclusione delle scale di desiderabilità sociale. Secondo la definizione data da Paulhus sulla desiderabilità sociale, questa è "*la tendenza a dare descrizioni di sé stessi positive*" (Paulhus, 2002). Il BIRD è una scala che misura la desiderabilità sociale, introdotta da Paulhus (1991) e comprende 40 *item,* 20 per ciascuna delle due subscale, che sono le seguenti:

- *Self-deceptive Enhancement* (SDE): è una tendenza inconscia a fornire un'immagine di sé più positiva della realtà, fino ad arrivare a credere all'immagine di sé fornita, con lo scopo di proteggere la propria autostima. Questo tipo di dissimulazione è correlato positivamente al narcisismo e a forme rigide di presunzione (Bobbio & Manganelli, 2011).
- *Impression Management* (IM): è la tendenza abituale e consapevole a presentarsi in modo migliore e socialmente desiderabile di fronte agli altri. Questo stile di risposta è positivamente correlato alla dissimulazione e ad alti punteggi nelle scale di validità e detezione della menzogna dei questionari *self-report* (Bobbio & Manganelli, 2011).

Nella letteratura riguardante lo stile di risposta legato alla desiderabilità sociale, ampio spazio hanno avuto le ipotesi circa il suo meccanismo di funzionamento, le quali sono ad oggi principalmente tre. Le prime ipotesi potrebbero spiegare più l'*Impression Management*, l'ultima sembra spiegare più il *Self-Deceptive Enhancement*:

- la dissimulazione si attua durante l'ultima fase della produzione delle risposte al questionario: gli individui, dopo aver recuperato le informazioni in memoria riguardanti sé stessi, necessarie a rispondere agli *item*, attuano una valutazione, selezione o modificazione delle stesse, in virtù della desiderabilità sociale. Questo meccanismo di manipolazione della propria immagine porta gli individui ad impiegare più tempo nel rispondere al questionario (Holtgraves, 2004);
- la dissimulazione viene messa in atto ancor prima del recupero delle informazioni: gli individui non si preoccupano di cercare informazioni riguardanti sé stessi in memoria, ma producono delle risposte basate soltanto su ciò che sarebbe socialmente preferibile. In questo caso, il tempo necessario per rispondere al questionario dovrebbe risultare ridotto. Questo meccanismo è proposto nelle più moderne concettualizzazioni dell'*Impression Management* (Holtgraves, 2004);
- la dissimulazione agisce più in profondità e distorce direttamente il meccanismo di recupero delle informazioni, tanto che si può parlare di *bias* o errore di recupero. Gli individui che devono rispondere ad un questionario, guidati dal desiderio di trovare informazioni di sé

socialmente desiderabili, recuperano dalla memoria, quasi involontariamente, solo informazioni che confermano un'immagine di sé positiva. Anche in questo caso, il tempo di risposta al questionario può risultare ridotto poiché non vi è un completo processo di recupero delle informazioni (Holtgraves, 2004).

Infine, illustriamo alcuni degli strumenti più utilizzati per l'identificazione della simulazione dei deficit cognitivi. Per quanto possano essere simulati diversi deficit, generalmente il disturbo maggiormente simulato in ambito forense è quello della memoria (Sartori et al., 2016), per questo tali strumenti sono incentrati prevalentemente su questo disturbo.

- *15-item Memory Test*

È stato introdotto da Rey nel 1958 ed è uno dei test più usati in quanto è molto semplice e rapido nella sua somministrazione. In particolare, si basa sulla logica del *Floor Effect* (metodo del livello minimo). Al soggetto viene presentato un *set* di 15 stimoli visivi (lettere, numeri e forme geometriche) da memorizzare in 10 secondi e da rievocare in una seconda fase, riportandoli su un foglio bianco (Lezak, 1995). Il compito risulta essere molto semplice perché la disposizione degli elementi segue una logica ripetitiva: molti soggetti normali infatti riescono a riprodurre facilmente tutti i 15 elementi. È stato proposto un cut-off di 9/15 stimoli correttamente memorizzati, un punteggio più basso è indice di una possibile volontà simulatoria (Stracciari et al., 2010). È importante che nelle istruzioni venga posta enfasi sul numero 15, per dare l'idea dell'elevata quantità di stimoli da ricordare. In realtà gli *item* vengono poi presentati in sequenze da 3 stimoli ciascuna, logiche e facili da memorizzare (Boone et al., 2002). Tuttavia, diverse ricerche sperimentali hanno dimostrato una scarsa specificità del test, con un elevato tasso di falsi positivi e una scarsa sensibilità, con troppi falsi negativi. Esiste anche una versione modificata del test, composta da 16 stimoli raggruppati in 4 righe di 4 *item*, che sembra avere caratteristiche migliori rispetto alla versione originale (Iverson et al., 1991).

- *Test of Memory Malingering* (TOMM)

Introdotto da Tombaugh nel 1996, è uno strumento molto utilizzato per la detezione della simulazione dei disturbi di memoria in contesti clinici e forensi. Rientra nelle caratteristiche del

STV (*Symptom Validity Test*). È una prova di memoria di riconoscimento a scelta forzata su materiale pittorico e si divide in tre prove, di cui la terza è opzionale. Nella prima prova vengono mostrate al soggetto 50 immagini da memorizzare (in forma di disegno di oggetti comuni); durante la seconda fase vengono presentate delle coppie di immagini – di cui una è un distrattore, quindi un'immagine non vista in precedenza (Ree set al., 1998) - e il soggetto deve indicare quale tra le due ha visto nella prima fase. Questo doppio compito può essere svolto una seconda volta, seguito da una fase di riconoscimento differito, dopo 15 minuti, al fine di smascherare i soggetti che credono che, a causa del periodo di latenza, la prestazione possa essere inferiore a quella della seconda fase di test, arrivando così a compiere un numero maggiore di errori. In realtà, dati sperimentali mostrano un ulteriore incremento della *performance* media, dovuto probabilmente alla visione aggiuntiva degli stimoli *target* (Stracciari et al., 2010). Il punteggio grezzo del TOMM può essere interpretato applicando due diverse procedure logiche: secondo la logica SVT un punteggio inferiore a 18/50 è indice di simulazione (Stracciari et al., 2010); secondo la logica del *Floor Effect* il soggetto viene ritenuto simulatore con un *cut-off* di 45/50. La logica del SVT risulta più affidabile e sensibile – con un basso tasso di falsi positivi – rispetto alla logica del *Floor Effect* (Stracciari et al., 2010).

- *Rey Dot Counting Test*

Introdotto da Rey nel 1941, è un test in cui vengono presentati ad un soggetto 12 cartoncini diversi su cui sono rappresentati dei punti in un numero variabile da 7 a 28, che il soggetto deve contare il più velocemente possibile e dirne la quantità. Questi punti possono seguire due logiche di raggruppamento: (a) sono sparsi in tutto il foglio in modo casuale, (b) vengono rappresentati in modo raggruppato. Viene registrato quindi il tempo di risposta. L'idea alla base del test è che il tempo impiegato a contare i punti raggruppati sia inferiore al tempo necessario a contare i punti nella condizione casuale. Questo accade perché nella condizione in cui i punti sono raggruppati il colpo d'occhio permette una maggiore rapidità nel conteggio. L'assenza di questo effetto cognitivo e quindi una velocità maggiore nel contare i punti sparsi rispetto a quelli raggruppati, è molto probabilmente un indice di simulazione. In particolare, il *cut-off* è di 180 secondi totali per il conteggio dei punti non raggruppati e 130 per quelli disposti in ordine casuale (Stracciari et al., 2010). Il test presenta un'alta specificità ma una bassa sensibilità, avendo un elevato tasso di falsi negativi (Stracciari et al., 2010).

2.4.3 Approcci recenti alla detezione della simulazione

Nei paragrafi precedenti, sono state descritte alcune tecniche che sono utilizzate per la detezione di un atteggiamento simulatorio generale da parte del soggetto esaminato. Spesso però, in un contesto forense vengono riportati dei sintomi specifici come per esempio l'amnesia per il crimine commesso. In questi casi, in cui è necessario testare delle sintomatologie specifiche, possono essere utilizzati degli strumenti innovativi che permettono quindi la detezione della simulazione di specifici sintomi.

Recentemente, per esempio, sono state introdotte delle tecniche basate sui tempi di reazione che rilevano latenze più lunghe e percentuali più elevate di errore quando viene messo in atto un atteggiamento simulatorio. L'aumento dei tempi di reazione è associato ai processi cognitivi aggiuntivi coinvolti nel processo di menzogna. L'ipotesi alla base di tali tecniche è che un compito cognitivo con maggiori richieste dovrebbe richiedere anche un tempo maggiore (Sartori et al., 2017). Un esempio di tecnica basata sui tempi di reazione è *l'Autobiographical Implicit Association Test* (aIAT) (Sartori et al., 2008) che verrà descritto di seguito.

- *Autobiographical Implicit Association Test* (aIAT)

Questo test è stato sviluppato da Sartori et al., (2008) ed è una tecnica che permette di verificare se un ricordo autobiografico specifico è presente nella memoria di un soggetto. Esso è una variante dell'*Implicit Association Test* (IAT) (Greenwald & McGhee, 1998): uno strumento di misura indiretto che in base ai tempi di latenza delle risposte stabilisce l'associazione tra concetti.

Nello IAT il soggetto deve classificare degli *item* presentati in sequenza secondo un ordine casuale, tramite due risposte motorie diverse (schiacciare il più velocemente possibile due diversi tasti in base agli *item* presentati). Gli *item* appartengono a quattro concetti: due concetti *target* e due dimensioni di un attributo (positiva/negativa). La logica alla base è che se due concetti sono associati nella mente del soggetto, allora i tempi di classificazione risulteranno molto più rapidi (condizione congruente) rispetto alla condizione in cui i due *item* richiedono risposte differenti (condizione incongruente) perché non sono associati nella mente del soggetto. La differenza, nei tempi di reazione, fra la condizione incongruente e quella congruente, viene chiamata "effetto IAT" (Stracciari et al., 2010).

Nello aIAT, invece, la prova si compone di cinque parti, chiamate blocchi, dove tre sono categorizzazioni semplici (1, 2, 4) e due sono categorizzazioni combinate (3 e 5); il terzo e il quinto blocco sono i più importanti del test in quanto la misurazione dei tempi di reazione di questi due blocchi indica quali eventi autobiografici sono veri e quali falsi. Lo aIAT, a differenza dello IAT tradizionale, al posto di parole e immagini usa come concetti delle frasi che descrivono fatti e sintomi dell'esaminato, con lo scopo di identificarne la presenza o meno di questi nella memoria autobiografica tramite l'analisi dei tempi di reazione. Le quattro categorie utilizzate in questo test sono le seguenti:

- frasi sempre "vere";
- frasi sempre "false";
- frasi "colpevoli (ovvero riferite ai sintomi che si sospettano essere simulati);
- frasi "innocenti".

Il soggetto, come nello IAT tradizionale, deve classificare le differenti tipologie di frasi premendo due tasti nel minor tempo possibile. La condizione "congruente" si verifica quando i tempi di reazione sono minori (maggiore velocità), la condizione "incongruente" si verifica quando i tempi di reazione sono maggiori (velocità minore). Le frasi che sono associate più velocemente con le frasi vere quando condividono la stessa risposta motoria saranno quelle a cui corrisponde la traccia di memoria autobiografica (Stracciari et al., 2010). Studi sperimentali hanno dimostrato l'efficacia di tale metodo nella detezione della simulazione del colpo di frusta e della depressione, con una precisione nella classificazione intorno al 92% (Stracciari et al., 2010). Inoltre, l'accuratezza dell'aIAT come strumento in grado di distinguere fra ricordi autobiografici veri e falsi è stata confermata e quantificata con una percentuale di circa il 91% (Sartori et al., 2008). È importante sottolineare però che anche le tecniche basate sulla misurazione dei tempi di reazione presentano dei limiti. Esse infatti, misurano solamente la latenza delle risposte, perciò all'esaminato è necessario controllare solamente questo parametro per falsificare la prova (Sartori et al., 2017).

Altri approcci recenti alla detezione della simulazione vengono illustrati da alcuni studi pionieristici che hanno dimostrato come sia possibile identificare automaticamente un mentitore applicando modelli di *Machine Learning* (di cui parleremo in seguito) all'analisi cinematica di una risposta motoria in un compito di doppia scelta (Monaro et al., 2016). Gli indici cinematici, come i tempi di reazione, possono essere utilizzati per il riconoscimento della simulazione. È stato osservato che il tracciamento del movimento della mano fornisce una ricostruzione in tempo reale dei processi

mentali sottostanti un determinato compito, inclusi quelli coinvolti nella produzione della simulazione (Monaro et al., 2016).

Per esempio, Duran et al., (2010) hanno utilizzato il *controller* della console "*Nintendo Wii*" per analizzare la traiettoria del movimento della mano durante un compito in cui i soggetti erano istruiti a mentire. I partecipanti dovevano rispondere ad alcune frasi presentate in maniera veritiera o falsificata in base a un *cue* visivo. I ricercatori hanno osservato che le risposte false differivano da quelle vere in alcuni parametri come l'*onset* della risposta motoria, il tempo totale necessario per la risposta, la traiettoria del movimento e in alcuni parametri cinetici come la velocità e l'accelerazione. Hibbeln et al. (2014) hanno studiato il movimento del *mouse* in un contesto di frode assicurativa *online*. È stato osservato che il movimento del *mouse* dei simulatori è più distante rispetto a quello dei non simulatori, che il movimento dei primi è più lento e che il loro tempo di risposta incrementa. Inoltre, aumenterebbero i "*clicks*" del tasto sinistro del mouse nella condizione di simulazione. L'applicazione del *mouse tracker* è stata utilizzata anche per smascherare le autodichiarazioni di false identità (Monaro et al., 2016). I risultati indicano che analizzare il movimento del *mouse* per discriminare un'autodichiarazione d'identità falsa da una veritiera ha un'accuratezza molto alta, di circa il 95%.

Uno dei vantaggi dell'utilizzo delle tecniche innovative quali la misurazione dei tempi di risposta e il *mouse tracking* è che raccogliere tale tipologia di dati è poco costoso e non richiede una strumentazione aggiuntiva rispetto agli elementi di un normale *computer* come il *mouse* e la tastiera. Inoltre, esse non richiedono una particolare esperienza da parte dell'esaminatore, rendendole utilizzabili anche in una fase di *screening*. Sicuramente però, tali tecniche presentano anche dei limiti. In primo luogo, in alcuni casi risulta difficile controllare tutte le caratteristiche del movimento analizzato e stabilirne la relazione con la detezione della menzogna. Inoltre, il *mouse tracking* può raggiungere un'accuratezza molto elevata per un sintomo specifico ma non per il *range* di sintomi che compongono una determinata patologia. Infatti, una delle maggiori sfide nell'identificazione della simulazione non consiste solamente nel decidere se un paziente ha una particolare patologia o meno, ma anche stimare la severità della malattia, identificare pazienti che esagerano una sintomatologia o la aggravano (Sartori et al., 2017). Come riportato da Sartori et al., (2017), per la detezione accurata dei casi di simulazione gli esaminatori dovrebbero considerare risorse multiple di dati indipendenti che devono includere tecniche di misurazione, validate e strutturate, progettate specificatamente per la detezione della simulazione e, se possibile, supportate da nuove tecniche innovative automatizzate e indipendenti dal controllo dell'esaminatore.

2.5 Tecniche innovative per la detezione della simulazione a livello del singolo item

In questo paragrafo verranno illustrate due tecniche innovative applicate in ambito forense che consentono la detezione della simulazione a livello del singolo *item*. Infatti, come precedentemente anticipato, le tecniche fino a questo momento utilizzate all'interno dei questionari come le scale di controllo presenti nel MMPPI-2 o nel MCMI-III sono in grado di identificare un atteggiamento simulatorio o dissimulatorio generale, ma non penetrano nella specificità del singolo *item,* fallendo nel riconoscimento delle risposte che sono state volontariamente alterate, comportando spesso l'invalidazione dell'intero test e quindi la perdita anche delle informazioni veritiere.

I modelli di seguito esposti hanno la particolarità di poter essere applicati a livello del singolo *item* e sono le tecniche che sono state implementate nel presente studio.

2.5.1 TF-IDF

Il *Term Frequency – Inverse Document Frequency* (TF-IDF) è uno degli indici più usati nel *computer science* e nel recupero delle informazioni (*information retrieval).* Tale indice è utilizzato per misurare l'importanza di un termine all'interno di un documento o di un sito *web*, che viene quantificata attraverso un peso ponderato dato a ciascuna parola (Zahng et al., 2011). Questo peso rappresenta la misura dell'importanza statistica del termine (Z Yun-tao et al., 2005). Tale funzione aumenta proporzionalmente al numero di volte che il termine è contenuto nel documento, ma cresce in maniera inversamente proporzionale con la frequenza del termine in altri documenti. La logica sottostante è quella per cui, all'interno di un documento (*d*), l'importanza di un termine ha una relazione proporzionale alla sua *Term Frequency* (TF), ovvero a quante volte il termine in questione compare nel documento. Vi è poi una relazione inversamente proporzionale tra lo stesso termine e l'intera collezione (*C*) di documenti (*IDF*), cioè quante volte quel termine compare in altri documenti. L'idea alla base di questo comportamento è quella di dare più importanza ai termini che compaiono nel documento, ma che in generale sono poco frequenti.

Per esempio, un termine che compare spesso nel documento (*d*) ma che raramente compare in altri documenti, è meno specifico rispetto ad un altro termine che non compare frequentemente nel documento (*d*) oppure compare molto spesso anche in altri documenti della collezione (*C*) (Z Yun-Tao et al., 2005). Le parole di congiunzione (come "e", "quindi", "ma"..) risultano essere molto

frequenti in un documento, ma poco specifiche, poiché sono parole molto comuni anche in altri documenti. Il TF-IDF si propone di conferire al termine in esame un'importanza ponderata che tenga conto anche della sua frequenza in altri documenti.

I motori di ricerca attualmente utilizzano il TF-IDF per recuperare le pagine *web* più pertinenti in base alle parole chiave utilizzate da un utente e per classificare questi risultati in base alla loro rilevanza. Negli ultimi anni, l'indice è stato implementato per studiare fenomeni sociali come il *cyberbullismo* (Cheng et al, 2019) e le *fake news* sui *social media* (Ahmed et al., 2017), o per implementare un modello in grado di fare un'analisi psicologica del linguaggio (Li et al., 2019). Inoltre, è stato utilizzato in neuropsicologia cognitiva per fornire un nuovo modello di memoria semantica e dei disturbi ad essa associati (Sartori & Lombardi, 2004), così come per modellare le risposte neuronali nel recupero dei concetti in base alla loro rilevanza semantica (Mechelli et al., 2006). Tuttavia, l'applicazione dell'indice TF-IDF all'ambito della detezione della simulazione risulta essere innovativo e potenzialmente efficace, in quanto sarebbe in grado sia di combinare la posizione relativa di una risposta rispetto al campione di validazione sia lo stile di risposta di uno specifico soggetto (Zhang et al., 2011).

Il TF-IDF è stato ideato nel 1972 da K. S. Jones, il quale definì una prima formula IDF (*Inverse Document Frequency)*. Successivamente, l'algoritmo venne perfezionato da G. Salton, nel 1975, con l'aggiunta della componente TF (*Term Frequency).* L'indice TF-IDF, infatti, è il prodotto di due statistiche stimate con metodi diversi:

1) *Term Frequency* (TF):

$$TF = \frac{n_{td}}{n_d}$$

Dove per n_{td} si intende il numero di volte che il termine *t* compare nel documento *d,* mentre con n_d si fa riferimento al numero di parole totali nel documento *d.*

2) *Inverse Document Frequency* (IDF):

$$IDF(t) = log\frac{N}{n}$$

Dove *N* sta per il numero di documenti nella collezione *C* e n sta per il numero di documenti in cui compare il termine *t.*

La formula del TF-IDF risulta essere la seguente:

data una collezione C di documenti, il TF-IDF per il termine *t* in un documento $d \in C$ è così calcolato:

$$TF - IDF(t, d, C) = TF(d, t) \times IDF(t, C)$$

Ritornando all'ambito forense, come anticipato, il TF-IDF permette di identificare il singolo *item* in cui il soggetto ha alterato la risposta. Applicando tale indice alla detezione della simulazione, esso si calcola per ciascuna risposta data al questionario (*q*), considerando come valore di TF il numero di volte in cui lo stesso partecipante risponde un determinato valore. Ad esempio, se un certo partecipante risponde con il valore di "3" solamente ad un *item* su 10 del questionario, il TF(3) sarà 1. L'IDF, invece, si calcola attraverso il logaritmo delle volte in cui quello stesso valore di risposta, in relazione all'*item* specifico preso in considerazione, compare anche nelle risposte degli altri partecipanti, per quanto riguarda uno stesso *item*: Log(N/*n*), dove N è il numero totale dei partecipanti e *n* è il numero di volte in cui essi hanno risposto nello stesso modo del valore considerato, per quell'*item* specifico. Supponiamo di voler calcolare l'IDF del valore "3" all'*item* 1, in un campione di 100 partecipanti. Troviamo che 70 partecipanti rispondono a quell'*item* con il valore "3", quindi IDF(3) all'*item* 1 sarà: Log(100/70). Il TF-IDF rappresenta Il punteggio finale associato a ciascun *item* del questionario e corrisponde al prodotto di TF per IDF. Il valore che emerge verrà poi interpretato come un indice di rarità e quindi di probabile menzogna.

Quindi, per la presente ricerca, Il TF-IDF è stato calcolato per ogni *item* del questionario nel seguente modo:

a) TF(*x,i,q*) viene calcolato come il numero di volte in cui un certo valore *x* selezionato per uno specifico *item* (*i*) si ripete in altri *item* del questionario (*q*) per lo stesso soggetto.
b) IDF(*x,i,q*) viene calcolato come Log(N/*n*) dove N è il numero di tutti i partecipanti che rispondono al questionario (*q*) e *n* rappresenta il numero delle volte che il valore *x* è ripetuto tra i partecipanti allo stesso specifico *item* in questione.
c) Il punteggio TF-IDF per il valore *x* nell'*item i* è il prodotto tra i punteggi TF(*x,i,q*) e IDF(*x,i,q*).

Il TF-IDF presenta alcune proprietà principali. Una prima proprietà è che tale indice è diverso per soggetti diversi che attribuiscono lo stesso valore ad uno stesso *item*. Per esempio, due partecipanti forniscono il valore "5" allo stesso *item*, ma possono ottenere un TF-IDF diverso perché riportano TF diversi per quel valore. Quindi, l'indice permette di individuare lo stile di risposta di un soggetto

e la detezione della volontà simulatoria. Una seconda proprietà è che otterremo alti valori di TF-IDF quando:

a) TF è alto: cioè, il valore *x* nell'*item i* viene ripetuto molte volte nel questionario.
b) IDF è alto: cioè, il valore *x* nell'*item i* è estremamente infrequente fra i partecipanti.
c) Sia TF che IDF sono alti: cioè, il soggetto in questione ha usato lo stesso valore *x* molte volte nel proprio questionario *q,* ma quello stesso valore è raro tra gli altri partecipanti.

Tenuto conto di queste proprietà, assumiamo che sia TF che IDF nei simulatori dovrebbero assumere valori alti. Infatti, i partecipanti che mentono nel questionario tendono a falsificare molte risposte per fornire una descrizione positiva di sé stessi (o negativa, a seconda del contesto) e ottenere un risultato finale di falsificazione dell'intero questionario, spiegando così alti TF (Levashina et al, 2014; Mazza et al., 2019). Inoltre, le risposte simulate sono di solito poco frequenti nelle distribuzioni delle risposte oneste (Sofroniou, 2014). Quindi, ci aspettiamo che coloro che mentono al questionario abbiano alti valori di IDF quando i loro punteggi grezzi sono confrontati con la distribuzione dei punteggi grezzi dei partecipanti che rispondono in maniera onesta. Quindi, questo adattamento di TF-IDF fornirà valori elevati per le risposte alle singole domande che sono altamente atipiche. Infatti, un'alta misura di TF-IDF rappresenta un valore di risposta altamente atipico, mentre una sua bassa misura è indice di frequenza. In tal modo, l'indice si comporta come un "rilevatore di risposte atipiche" a livello del singolo *item* (non a livello del soggetto). In quest'ottica, ci aspettiamo che le risposte nella condizione simulata presentino un valore di TF-IDF molto più elevato rispetto alla distribuzione dell'indice nelle risposte oneste (*honest condition*).

Calcolato il TF e l'IDF per ciascuna risposta data dal soggetto nel questionario, se il loro prodotto è più elevato di una determinata soglia, si categorizza come alterata la risposta in esame. I valori IDF sono calcolati basandosi sulle risposte date nella condizione onesta. Il valore di soglia per la detezione della simulazione viene scelto come il percentile nella distribuzione di TF-IDF (per ogni *item* del questionario) nella condizione onesta, che massimizza la precisione per la detezione della simulazione su un gruppo di controllo separato. In pratica, inizialmente viene considerato solo il sottogruppo di risposte oneste; in seguito, si calcolano i valori di IDF associati ad ogni domanda del questionario. Si prosegue considerando un piccolo gruppo di risposte simulate (come gruppo di validazione) e si stabilisce la soglia di detezione della simulazione per ogni *item* del questionario, basandosi sulla distribuzione di questi punteggi. Tale valore di soglia non è stimato sui punteggi grezzi TF-IDF, bensì sul percentile corrispondente alla distribuzione delle risposte. Qui sotto

riassumiamo la procedura che abbiamo utilizzato in questo studio per calcolare e confrontare i punteggi di TF-IDF:

1) Abbiamo calcolato il punteggio TF-IDF per le risposte oneste, come illustrato sopra.
2) Abbiamo calcolato il punteggio TF-IDF per le risposte simulate usando l'IDF delle risposte oneste e abbiamo calcolato il TF utilizzando i punteggi grezzi delle risposte simulate. Come risultato, abbiamo ottenuto due nuovi *dataset* (TF-IDF-*honest* e TF-IDF-*dishonest*) in cui ogni partecipante aveva tanti punteggi TF-IDF quanti sono gli *item* del questionario.
3) Poi, abbiamo stimato la distribuzione percentile dei punteggi TF-IDF delle risposte oneste. L'ipotesi è che il punteggio TF-IDF dei simulatori a confronto con la distribuzione percentile dei punteggi TF-IDF degli onesti, sarà posizionato sulla coda destra della distribuzione, poiché un valore così alto è meno frequente nella distribuzione percentile degli onesti.
4) Infine, abbiamo impostato una soglia, in termini di punteggio percentile, che massimizza la precisione nel distinguere tra risposte oneste e simulate su un *set* di risposte di convalida separato.

Forniamo ora un esempio pratico per comprendere meglio l'applicazione del TF-IDF.

Supponiamo di aver somministrato un questionario di 12 *item* sulla *Dark Triad* con l'obiettivo di individuare le risposte simulate. Calcoliamo il TF e lo combiniamo con il rispettivo IDF per il primo *item*, raccolto da un precedente campione di rispondenti onesti. Immaginiamo di aver ottenuto un punteggio TF-IDF associato all'80esimo percentile rispetto alla distribuzione percentile dei punteggi TF-IDF degli onesti. Se questo punteggio percentile è superiore alla soglia che abbiamo impostato in precedenza (ad esempio, il 75esimo percentile), consideriamo la risposta a quell'*item* come alterata. Procediamo allo stesso modo per i rimanenti *item* del questionario. Di conseguenza, otteniamo una panoramica di quali *item* possiamo identificare come falsi.

In conclusione, il TF-IDF può essere considerato un indice rilevatore di anomalie in grado di rilevare risposte atipiche a livello del singolo *item* e lo stile di risposta del soggetto. Questa particolare applicazione del TF-IDF è la prima che mira ad individuare la simulazione a livello del singolo *item*, senza la necessità di inserire scale di controllo all'interno del questionario. Riporteremo i risultati finali nel capitolo quarto e la discussione nel capitolo quinto.

2.5.2 Modello di Machine Learning e Transformer: Self-Attention Based Autoencoders (SABA)

Negli ultimi anni, nel campo della psichiatria forense, si sta assistendo ad una progressiva fusione tra la conoscenza delle neuroscienze e gli strumenti forniti dalle scienze informatiche e dalla bioingegneria. Una delle aree più promettenti legata all'intelligenza artificiale è quella del *Machine Learning* (Mitchell, 1997). Il *Machine Learning* (ML) è una disciplina relata alla statistica computazionale che utilizza algoritmi matematici per categorizzare elementi all'interno di categorie differenti. Nel dettaglio, questa metodologia si basa sull'osservazione di dati reali per predire nuovi dati o produrre nuova conoscenza. Si tratta quindi di algoritmi in grado di predire degli *outcome* dopo essere stati allenati ad alcuni *input.* Sintetizzando, l'oggetto del ML è quello di trasformare i dati in conoscenza. Nel caso della detezione della simulazione, si utilizzano gli indici oggettivi della simulazione per analizzare e valutare le relazioni che intercorrono tra le variabili osservate e per stabilire le regole necessarie per una classificazione automatica dei dati.

Lo sviluppo di algoritmi di *Machine Learning* ha ricevuto sempre più attenzione diffondendosi in diversi ambiti, quali la statistica, la filosofia, la psicologia ed altri ancora (Mitchell, 1997). Negli anni, numerosi approcci basati su questi algoritmi sono stati utilizzati per cercare di individuare i questionari in cui i partecipanti alteravano le loro risposte (Mazza et al., 2019; Orrù et al., 2020). I campi di applicazione del ML si stanno continuamente evolvendo e spaziano dalle campagne di mercato, dove i clienti vengono profilati in base alle loro caratteristiche, all'ambito forense dove il *Machine Learning* è stato utilizzato per esempio nel contesto delle false identità o per smascherare un disturbo depressivo simulato. Secondo Orrù et al., (2020) le analisi di ML avrebbero il vantaggio di massimizzare l'accuratezza e minimizzare i problemi di replicabilità.

Inoltre, il *Machine Learning* può essere suddiviso in due tipologie di funzionamento: (a) *Supervised Machine Learning Models* e (b) *Unsupervised ML Models.* I modelli di apprendimento supervisionato (a) presentano la necessità di essere esposti ad una grande mole di dati affinché l'algoritmo possa essere implementato. Inoltre, nessun algoritmo è stato ancora utilizzato per identificare le specifiche domande a cui il partecipante ha fornito una risposta alterata.

Per questo motivo, nel presente lavoro, oltre alle analisi condotte sui dati grezzi e sull'indice TF-IDF, è stata implementata e validata una rete neurale chiamata *Self-Attention Autoencoder* (SABA) al fine di individuare con precisione gli *item* del questionario cui sono state fornite risposte alterate.

Il *Self-Attention Based Autoencoder* (SABA), si basa su un modello *Transformer* (Marcowitz, 2021) che consiste in un'architettura neurale artificiale in grado di mettere in relazione diverse caratteristiche di *input* basandosi solo su un meccanismo di attenzione. Questo modello ha ottenuto grande risonanza all'interno della comunità di *Natural Language Processing* (NLP) per operazioni come la traduzione del linguaggio o la risposta alle domande, in quanto permette di fare affidamento su una stessa architettura per eseguire diverse operazioni. Senza entrare troppo nel dettaglio, un *Autoencoder* è una rete neurale artificiale non supervisionata (*unsupervised* – (b) basata su un'architettura *encoder-decoder*, in grado di rappresentare in modo efficiente dati in *input* chiamati *codings*, riducendo la loro "dimensionalità" (ad esempio, ignorando il rumore del segnale) e cercando poi di ricostruire l'*input* originale basato su queste rappresentazioni ridotte.

Per questo lavoro è stata implementata una versione ottimizzata di questo modello che si basa sulla *Self-Attention,* una tecnica che cerca di emulare l'attenzione umana e che restituisce solo i dati di *input* che sono più salienti sulla base del contesto. Il grande vantaggio di questo modello, come si vedrà, è che non necessita di un corposo dataset che racchiude diversi *pattern* di falsificazione, ma si basa solo sui dati del gruppo degli onesti. Di seguito viene proposta una rappresentazione grafica dell'architettura dell'*Autoencoder* basato sulla tecnica *Self-Attention* (Figura 1).

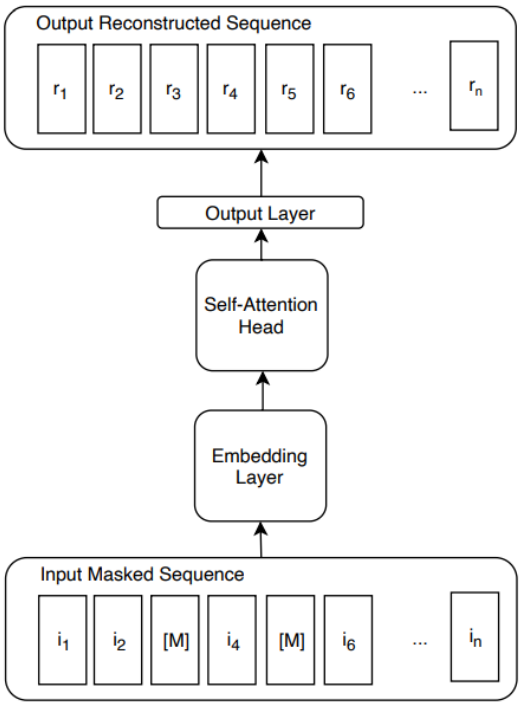


Figura 1. *Rappresentazione grafica della rete neurale artificiale (modello Autoencoder con Self-Attention)*

Come prima cosa abbiamo fornito come *input* al modello la sequenza di risposte oneste raccolte nelle precedenti fasi raggruppate per *item* (i_1,i_2…i_n di Figura 1). Durante questo primo passaggio abbiamo applicato un'operazione di mascheramento di alcuni valori simile al paradigma *Masked Language Model* (MLM) (Devlin et al., 2018), che consiste nel "nascondere" i valori di alcuni *item* dati in *input* con una maschera definita "*special token*" ([M] in Figura 1). Successivamente il modello è stato allenato a predire i valori delle risposte mascherate nella sequenza di *input* (12 valori, uno per ogni *item* del questionario) minimizzando l'errore assoluto medio (MAE) tra i valori predetti e quelli originali. A tal fine, il primo *layer* del nostro modello, chiamato *Embedding Layer*, associa a ciascun *item* nella sequenza un vettore di dimensioni *f*, creando una matrice che contiene i vettori associati ad ogni *item*. Alimentiamo poi queste sequenze al *Self-Attention layer* (SA), seguito da un *feed-forward output layer*[4] (per maggiori dettagli sul funzionamento del *Self-Attention layer* si consiglia la lettura di Vaswaniat al., (2017). Questa procedura ci permette di trattare le risposte come se fossero parole in una frase e di applicare il meccanismo standard del *Transformer Self-Attention* (SA).

L'obiettivo di questa fase di *training* è quello di allenare l'algoritmo a ricostruire i valori delle risposte oneste di ogni *item* sulla base delle risposte non mascherate date dai partecipanti, in modo tale che il modello impari lo stile con cui i partecipanti hanno risposto in modo onesto. La logica di base prevede che i partecipanti rispondano agli *item* seguendo uno stile coerente di risposta: una persona che ha fornito un punteggio elevato ad una specifica domanda fornirà anche punteggi elevati ad altre domande simili. L'*output* sarà quindi la sequenza data come *input* in cui i valori onesti mascherati sono stati ricostruiti come appena descritto.

A questo punto, per identificare le risposte disoneste, viene dato in *input* una sequenza corrispondente agli *item* disonesti mascherando uno alla volta gli elementi, e iterando questo processo per ciascuna domanda del questionario (in questo caso 12 volte), così come per le risposte oneste. Alla fine, si confronterà il valore predetto dal nostro modello con quello fornito dal partecipante al questionario. Se la differenza tra questi valori (previsto-reale) è maggiore di una determinata soglia, allora segnaliamo questa domanda come falsa. Il modello ricostruisce quindi 12 varianti della sequenza ciascuna contenente la ricostruzione dell'*item* che era stato mascherato in fase di *input*. Al termine del processo otterremmo quindi una ricostruzione globale delle risposte *faked* al questionario.

[4] Una rete neurale feed-forward ("rete neurale con flusso in avanti") consiste in una rete neurale artificiale dove, a differenza delle reti neurali ricorrenti, le connessioni tra le unità non formano cicli. In breve, le informazioni al suo interno si muovono solo in avanti, rispetto ai nodi d'ingresso e a quelli nascosti, fino ad arrivare ai nodi d'uscita Vaswaniat al., (2017).

L’applicazione di tale modello in contesti reali come quello forense è particolarmente vantaggiosa in quanto, basandosi su un *set* di risposte oneste, non sarà necessario raccogliere un campione significativo e valido di risposte *dishonest*, le quali sono state infatti utilizzate solo per validare la *performance* dell’approccio proposto. Inoltre, tale modello sfrutta le informazioni contestuali disponibili per individuare le risposte false. A differenza delle altre tecniche psicometriche, che stimano le risposte *faked item* per *item* senza dipendere dalle altre risposte che il partecipante fornisce, il modello proposto individua l’anomalia di una risposta basandosi sull'intero contesto delle risposte fornite dal singolo partecipante.

Dopo questa panoramica sulla simulazione e sulle tecniche per la sua detezione, nel prossimo capitolo entreremo nel vivo del metodo dello studio e dell’applicazione dell’indice TF-IDF e del *Machine Learning* alla *Dark Triad*, con l’obiettivo di sfruttare le caratteristiche di queste tecniche innovative per evidenziare l’alterazione delle risposte a livello del singolo *item*. Risulta interessante applicare tali metodi proprio al questionario *Dirty Dozen* al fine di smascherare un comportamento dissimulatorio, poiché in accordo con diversi autori è noto come in letteratura ci siano più tecniche per la detezione del *fake bad* piuttosto che per la detezione del *fake good* (Mazza et al., 2020).

Nel prossimo capitolo verrà discusso del tema specifico del contesto legale della valutazione dell’idoneità genitoriale e della sua relazione con la dissimulazione e con la *dark triad.*

CAPITOLO III

3 LA RICERCA SPERIMENTALE

3.1 Descrizione del progetto

Il presente studio è stato condotto dal Dipartimento di Psicologia Generale (D.P.G.) dell'Università degli Studi di Padova. La ricerca, alla luce del quadro teorico precedentemente esposto, ha avuto come obiettivo principale quello di validare due metodi per rilevare la dissimulazione (*fake good*) a livello del singolo *item* utilizzando:

(a) l'indice TF-IDF (T*erm Frequency – Inverse Document Frequency*), mutuato dall'ambito informatico, definibile come un indice di detezione delle anomalie;

(b) il SABA, un adattamento del modello di *Machine Learning* noto come *Transformer* (Vasawani et al., 2017).

In particolare, nella presente ricerca, entrambi i metodi sono stati applicati al questionario *Dirty Dozen* (DD) di Jonason & Webster (2010) nella sua versione italiana validata da Schimmenti et al., (2017), il quale indaga le tre componenti della triade oscura della personalità (Narcisismo, Machiavellismo, Psicopatia), attraverso l'analisi dei punteggi dei singoli *item* del questionario, somministrato ad un gruppo di individui italiani.

Alla luce dell'attuale emergenza sanitaria dovuta al virus Covid-19, la somministrazione del questionario è avvenuta completamente in forma *online*, attraverso un'apposita piattaforma (Jotform ®) che ha reso possibile la compilazione del questionario in forma digitale.

Nel seguente capitolo presenteremo i dettagli relativi agli obiettivi, allo strumento, ai partecipanti e alle ipotesi della ricerca.

3.1.1 Obiettivi della ricerca

L'obiettivo principale della ricerca è quello di individuare un nuovo metodo di detezione della simulazione in un questionario, attraverso l'analisi delle risposte di ciascun soggetto ai singoli *item*. In particolare, per raggiungere questo obiettivo abbiamo verificato l'efficacia dell'utilizzo dell'indice TF-IDF e l'efficacia dell'utilizzo di algoritmi di *Machine Learning* per la detezione della

simulazione, applicati alla triade oscura della personalità. La verifica dell'efficacia di suddette metodologie innovative è stata valutata tramite il confronto con le analisi dei soli dati grezzi, ipotizzando una maggiore efficacia dell'indice TF-IDF e dell'utilizzo di algoritmi di *Machine Learning*.

Infatti, il TF-IDF, a differenza dei metodi tradizionali finora presenti in letteratura (per i dettagli si rimanda al capitolo 2, par. *2.5.1*), consente di stimare la probabilità di simulazione di un soggetto in singoli *item* di strumenti *self-report* e non solo la tendenza generale alla simulazione, in quanto è in grado sia di combinare la posizione relativa di una risposta rispetto al campione di validazione, sia lo stile di risposta di uno specifico soggetto.

3.2 Materiali e metodo

Di seguito verranno riportati il metodo, il materiale e le procedure utilizzate per la conduzione della ricerca.

3.2.1 Partecipanti

Data l'emergenza sanitaria, i partecipanti allo studio sono stati reclutati *online* tramite Facebook, Instagram o Whatsapp da contatti personali ed esterni, tramite condivisione del *link* del questionario creato su Jotform.

I partecipanti sono stati pienamente informati sullo scopo di questa ricerca e sono stati rassicurati circa l'anonimato delle loro identità. Infatti, i dati raccolti sono anonimi e sono stati utilizzati in forma aggregata solo per questo studio. I partecipanti hanno accettato volontariamente di unirsi a questa ricerca, per la quale non è stato previsto alcun compenso.

Hanno partecipato all'esperimento un totale di 600 soggetti, dei quali 107 sono stati esclusi in quanto avevano risposto in modo errato alle domande di verifica della comprensione delle istruzioni poste prima dell'inizio del questionario, sia per la prima compilazione (*dishonest*) sia per la seconda compilazione (*honest*), sia al termine del questionario stesso.

Prima di iniziare con il questionario sono stati raccolti i dati demografici circa sesso, età, educazione, stato relazionale, numero di figli ed età dei figli.

Stato relazionale e numero di figli sono stati raccolti unicamente a scopo di indagine qualitativa, in quanto lo scenario presentato prima di compilare il questionario richiedeva di immaginarsi davanti ad un Giudice in una condizione di divorzio dal proprio *partner* e di richiesta di affido del/dei minore/i. È ragionevole pensare che coloro che hanno una relazione sentimentale e dei figli possano immedesimarsi maggiormente in questa situazione ipotetica e rispondere in modo più genuino al questionario, secondo le istruzioni date. In ogni caso, qualora i partecipanti all'esperimento non avessero avuto figli, era chiesto loro di scrivere "nessuno" nell'apposito spazio e di procedere ugualmente alla compilazione del questionario secondo le indicazioni fornite.

Riguardo l'età, i partecipanti dovevano scegliere fra diverse classi di età: 18-29; 30-39; 40-49; 50-59; 60-69; 70-79; 80 o più. Per quanto concerne l'educazione, i partecipanti dovevano indicare gli anni di scolarizzazione, potendo scegliere fra diverse opzioni: 5 (scuola elementare), 8 (scuola media), 13 (scuola superiore), 16 (laurea triennale), 18 (laurea magistrale), 19 o più (master, dottorato, ecc.). Ai fini delle analisi statistiche, sono stati considerati gli anni di scolarizzazione (e.g., 5, 8, 13 etc.) e l'etichetta "19 o più" è stata sostituita con il valore "19". Per quanto riguarda il numero di figli, abbiamo chiesto ai partecipanti di scrivere quanti figli avessero, potendo scegliere fra: nessuno; 1; 2; 3; 3 o più.

Il campione finale, quindi, comprendeva 493 partecipanti di cui 369 femmine (75%) e 124 maschi (25%). L'età dei partecipanti variava dai 18 ai 69 anni (M= 32.87; DS= 11.54) e il livello di educazione dagli 8 ai 19 o più anni di scolarizzazione (M= 16.68; DS= 2.52). La maggior parte dei partecipanti non aveva figli (75%), mentre fra i partecipanti che avevano figli (25%) la maggior parte ne aveva 1 o 2 (rispettivamente 37% e 52%), mentre i partecipanti rimanenti (11%) hanno dichiarato di avere 3 figli o più di 3 figli.

3.2.2 Struttura dell'esperimento

Il questionario finale era composto da 12 *item* randomizzati, creati con JotForm (https://www.jotform.com). La prova è stata somministrata completamente *online* e la durata per il suo completamento è stata mediamente di 5-10 minuti. I partecipanti potevano usare qualsiasi dispositivo per partecipare all'esperimento, come cellulare, tablet o computer; l'unico requisito era avere la connessione ad Internet e trovarsi in un ambiente tranquillo e privo di distrazioni.

Il questionario somministrato era il *Dirty Dozen* di Jonason & Webster (2010) nella sua versione italiana validata da Schimmenti et al., (2017), interamente riportato in Appendice A. Come descritto

nel paragrafo *3.2.1*, al questionario sono state aggiunte alcune domande preliminari di tipo demografico e di verifica di comprensione delle istruzioni, manipolate per la somministrazione dell'esperimento.

Prima della prova tutti i soggetti hanno preso visione e autorizzato il consenso informato all'esperimento, garantendo la protezione dei dati personali e l'assoluto anonimato.

L'esperimento è stato suddiviso in 2 fasi:

1. Nella prima parte è stato chiesto ai partecipanti di rispondere una prima volta al questionario mentendo, cioè simulando le risposte date con l'obiettivo di apparire migliori in una ipotetica situazione nella quale era chiesto ai soggetti di immedesimarsi (condizione disonesta – *dishonest condition; fake good*).
2. Nella seconda parte ai partecipanti è stato chiesto di compilare nuovamente lo stesso questionario rispondendo in maniera onesta (condizione onesta – *honest condition*).

Nella condizione *dishonest,* i partecipanti erano istruiti ad immaginarsi come genitori in una situazione di divorzio dal proprio *partner* nel contesto di una consulenza richiesta dal Giudice per decidere sulle migliori condizioni di affido del/i minore/i. In particolare, nello scenario presentato, al soggetto veniva chiesto di compilare un questionario somministrato da uno psicologo che avrebbe utilizzato per valutare le caratteristiche dei partecipanti come genitori: l'obiettivo è ottenere la custodia dei figli. Di seguito le istruzioni fornite:

*"**Immagina che tu e tua/o moglie/marito stiate divorziando e stiate litigando per l'affido dei figli** (rispondi anche se non hai figli). Siete nel contesto di una consulenza richiesta dal Giudice che deve decidere sulle migliori condizioni di affido dei figli. Lo psicologo incaricato vi chiede di compilare questa prova che sarà usata per valutare le vostre caratteristiche di genitore.*

***Rispondi alle domande in modo da fare bella figura, nascondendo comportamenti o pensieri generalmente considerati negativi. Il tuo obiettivo è ottenere l'affido dei figli e risultare migliore di tua/o moglie/marito agli occhi del Giudice. Cerca di dare un'immagine positiva, anche se questo vuol dire mentire**."*

Le parole chiave delle istruzioni erano in grassetto e/o sottolineate per catturare l'attenzione dei partecipanti e assicurarci che rispondessero secondo le indicazioni date.

Successivamente, nella seconda parte, sono stati presentati gli stessi 22 *item* del questionario, nello stesso ordine proposto nella prima parte. Di seguito le istruzioni fornite:

"*Ti verranno ora presentate le stesse 22 affermazioni di prima, riguardanti dei modi di pensare o di comportarsi e devi scegliere una tra le 5 alternative che ti verranno proposte.*

Questa volta rispondi in modo totalmente sincero, anche se questo vuol dire ammettere aspetti negativi di te stesso."

La ragione per cui è stato chiesto prima ai partecipanti di mentire e solo successivamente di rispondere in modo onesto, è stata quella di rendere la simulazione più ecologica. Infatti, proponendo ai partecipanti di rispondere fin da subito mentendo ad un questionario con il quale non avevano familiarità, ci siamo assicurati che la simulazione fosse più genuina e più simile a quella che sarebbe stata nella realtà, cioè quando un soggetto vede per la prima volta delle domande senza alcun *bias* dato dal fatto di conoscere già le domande stesse.

Quindi, ciascun partecipante ha svolto lo stesso questionario due volte: la prima nella condizione *dishonest* e la seconda nella condizione *honest.* La logica era proprio quella di raccogliere dati nella condizione onesta e dati nella condizione disonesta dallo stesso partecipante per valutare il livello di accuratezza dell'indice TF-IDF e del ML nell'individuare le risposte simulate.

Ai fini delle analisi statistiche sono state considerate prima le risposte date in modo onesto (seconda parte dell'esperimento) e poi le risposte date in modo disonesto (prima parte dell'esperimento).

Infine, le risposte erano a scelta multipla con una sola possibile opzione di risposta per ogni *item* che i partecipanti potevano cliccare con il dito o con il *mouse* a seconda del dispositivo utilizzato. Per ogni *item* i soggetti potevano scegliere il proprio grado di accordo tra cinque opzioni, disposte verticalmente, una sotto l'altra:

- Fortemente d'accordo
- D'accordo
- Né in accordo né in disaccordo
- In disaccordo
- Fortemente in disaccordo

Ai fini delle analisi statistiche le seguenti opzioni di risposta sono state trasformate in valori da 1 (fortemente in disaccordo) a 5 (fortemente d'accordo).

Se un partecipante non rispondeva ad un *item*, volontariamente o involontariamente, il sistema era programmato in modo da non permettere di procedere con l'*item* successivo. L'obbligo di risposta era segnalato con un asterisco rosso. La ragione di questa scelta è quella di evitare valori mancanti, che sarebbero risultati problematici nelle analisi statistiche.

I partecipanti, inoltre, non potevano tornare indietro all'*item* precedente una volta che avevano risposto, perché, sempre nell'ottica di una maggior ecologicità, volevamo raccogliere la prima impressione che veniva in mente ai soggetti, evitando che questa potesse essere influenzata da un ripensamento successivo.

3.2.3 Strumento utilizzato

Come precedentemente riportato, per questo studio è stato utilizzato il questionario *Dirty Dozen* di Jonason & Webster (2010) nella sua versione italiana validata da Schimmenti et al., (2017) per indagare la triade oscura della personalità. La DD consta di 12 *item* totali, 4 per ciascuna subscala della triade oscura (Machiavellismo, Narcisismo e Psicopatia). Per un approfondimento sullo strumento si rimanda al capitolo I.

Per verificare che i partecipanti avessero letto e compreso correttamente le istruzioni prima di compilare il questionario, sono state aggiunte tre domande immediatamente dopo le istruzioni circa la parte in cui veniva loro chiesto di mentire, una domanda immediatamente dopo le istruzioni della seconda parte in cui veniva chiesto loro di rispondere sinceramente e una domanda conclusiva alla fine del questionario. Le domande riguardavano le istruzioni fornite ai partecipanti circa il modo in cui dovevano rispondere al questionario nelle due condizioni. Infatti, dalla letteratura emerge come l'utilizzo e la manipolazione di domande di verifica della comprensione dei partecipanti possa garantire una compilazione del questionario più accurata. In uno studio di Crighton et al., (2017), che ha utilizzato delle domande di controllo dopo le istruzioni fornite per la compilazione del questionario, emerge come i partecipanti che hanno rispettato tali istruzioni rispondendo correttamente alle domande di controllo, hanno ottenuto risultati migliori per la parte *dishonest* rispetto ai partecipanti che hanno ricevuto istruzioni *standard* senza dover successivamente rispondere a domande di verifica della comprensione.

Nel nostro campione sono stati inclusi solo i partecipanti che hanno risposto correttamente a tutte e 5 le domande di controllo. Quindi, sono stati esclusi 107 soggetti dalle analisi statistiche in quanto non hanno superato questa fase di verifica della comprensione.

Qui sotto riportiamo le domande di controllo utilizzate:

Domande poste immediatamente dopo le istruzioni della prima parte (*dishonest*):

"Prima di procedere, 3 brevi domande per verificare che le istruzioni siano chiare:

In base alle istruzioni lette ora, in che situazione dovrai immaginare di trovarti per rispondere alle successive domande?

- *Io e mia/o moglie/marito stiamo divorziando e stiamo litigando per l'affido dei figli*
- *Ho commesso un reato e mio figlio verrà dato a mia/o moglie/marito*
- *Io e mia/o moglie/marito vogliamo adottare un bambino*

In base alle istruzioni lette ora, qual è l'obiettivo che devi tenere a mente mentre rispondi?

- *Devo fare bella figura davanti al Giudice per ottenere l'affido dei figli*
- *Devo risultare il/la più onesto/a possibile per ottenere l'affido dei figli*

In base alle istruzioni lette ora, in che modo devi rispondere alle domande?

- *Nasconderò pensieri e comportamenti per fare bella figura anche se questo vuol dire mentire*
- *Sarò onesto/a anche se questo vuol dire mostrare lati negativi di me*

Domanda posta immediatamente dopo le istruzioni della seconda parte (*honest*):

Una domanda per verificare che tu abbia compreso le istruzioni

In base alle istruzioni lette ora, in che modo devi rispondere alle domande?

- *Sarò onesto/a anche se questo vuol dire mostrare lati negativi di me*
- *Nasconderò pensieri e comportamenti per fare bella figura anche se questo vuol dire mentire*

Domanda posta alla fine del questionario, dopo i ringraziamenti per la partecipazione:

Prima di lasciare la pagina, un'ultima domanda per verificare che le istruzioni fossero chiare

Quali istruzioni hai avuto nella compilazione del questionario?

- *Ho dovuto sempre dire la verità*
- *Non ho ricevuto istruzioni particolari*
- *Prima ho dovuto dire la verità e poi ho dovuto mentire per ottenere l'affido dei figli*
- *Ho dovuto sempre mentire per fare bella figura*
- *Prima ho dovuto mentire per ottenere l'affido dei figli e poi ho dovuto dire la verità*

Il testo completo del questionario nella sua versione in italiano è incluso nell'Appendice A.

3.2.4 Ipotesi di ricerca

Le ipotesi alla base del seguente studio sono:

1. (H1) I punteggi grezzi raccolti nelle condizioni *dishonest* e *honest* sono diversi tra loro in modo statisticamente significativo.
2. (H2) La detezione della dissimulazione risulta migliore mediante l'utilizzo della tecnica TF-IDF o mediante l'utilizzo di algoritmi di *Machine Learning* rispetto alle analisi dei soli dati grezzi.
3. (H3) I valori TF-IDF *dishonest* si collocano prevalentemente al di sopra di un determinato valore di soglia e, sopra tale valore, il rapporto tra risposte *faked* e risposte oneste è maggiore di 1.

Nel seguente capitolo verranno presentati e analizzati i dati raccolti.

CAPITOLO IV

4 ANALISI DEI DATI E RISULTATI

In questo capitolo proponiamo una rappresentazione visiva dei dati raccolti e successivamente verranno presentate le analisi statistiche effettuate sui dati grezzi e sull'indice TF-IDF e infine l'applicazione del modello di *Machine Learning,* proponendo anche un confronto tra suddette tecniche. Si specifica che per le analisi statistiche sono stati utilizzati i programmi Excel e JASP, mentre per la parte relativa al *Machine Learning* (SABA) è stata effettuata l'implementazione di algoritmi tramite il linguaggio di programmazione Python.

4.1 Rappresentazione visiva dei dati

In primo luogo, sono state calcolate le medie dei dati raccolti tramite la somministrazione della DD. Di seguito verrà presentata una rappresentazione grafica dei punteggi medi grezzi e delle deviazioni standard nelle condizioni *dishonest e honest* (Figura 2) e i valori del punteggio medio a livello di ogni *item* (Tabella a).

Come è possibile osservare dalla Figura 2, l'andamento generale dei punteggi medi grezzi per ciascun *item* mostra una differenza evidente tra le due diverse condizioni: disonesta e onesta. I punteggi medi nella condizione *dishonest* risultano essere inferiori rispetto a quelli nella condizione *honest.*

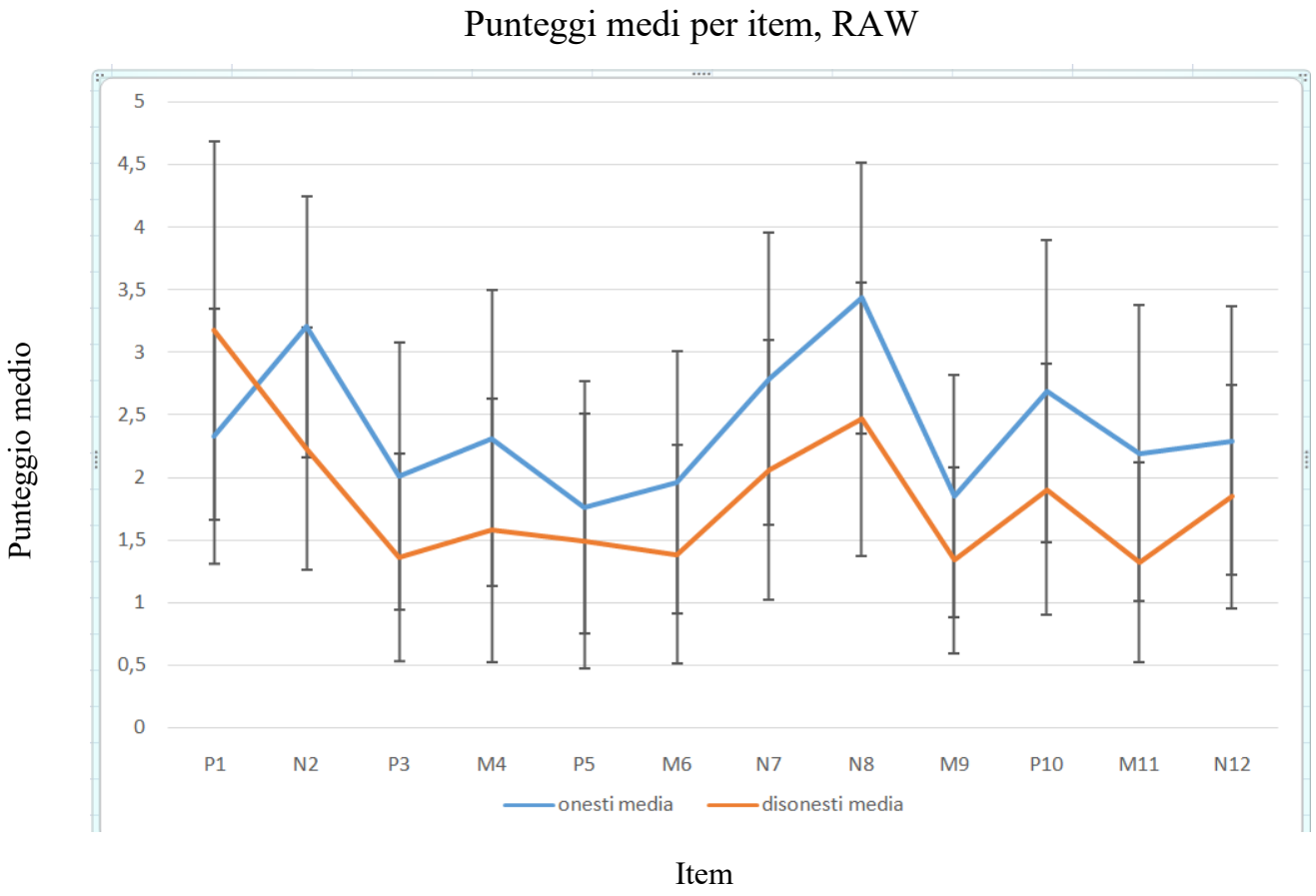


Figura 2. *Grafico delle distribuzioni dei punteggi grezzi medi e delle rispettive deviazioni standard nelle condizioni onesta/honest (blu) e disonesta/dishonest (rosso). Sull'asse delle ascisse sono raffigurati gli item, mentre sull'asse delle ordinate il punteggio medio ottenuto per il relativo item, all'interno di ciascuna condizione.*

La Tabella a) riporta quanto sopra rappresentato. In particolare, possiamo osservare come nella condizione onesta, la distribuzione delle risposte presenta una maggiore variabilità: M (H) = 2,40, SD (H) = 1,09; M (D) = 1,85, SD (D) = 0,99, dimostrando che le risposte date nella condizione onesta si differenziano maggiormente tra di loro. Questo potrebbe riflettere una tendenza generale nella condizione disonesta a rispondere con punteggi estremi nella maggior parte degli *item*, indipendentemente dallo specifico contenuto semantico.

Tabella a). *Tabella dei valori dei punteggi medi e delle deviazioni standard di ciascun item per le condizioni honest e dishonest.*

MISURE	**P1**	**N2**	**P3**	**M4**	**P5**	**M6**
Media (H)	2,33	3,21	2,01	2,31	1,76	1,96
SD (H)	1,01	1,04	1,07	1,18	1,01	1,05
Media (D)	3,17	2,23	1,36	1,58	1,49	1,38
SD (D)	1,51	0,97	0,83	1,05	1,02	0,87
MISURE	**N7**	**N8**	**M9**	**P10**	**M11**	**N12**
Media (H)	2,79	3,43	1,85	2,68	2,19	2,29
SD (H)	1,17	1,08	0,97	1,21	1,18	1,07
Media (D)	2,06	2,46	1,34	1,91	1,32	1,84
SD (D)	1,04	1,09	0,74	1,01	0,8	0,89

Note. *In Tabella a) sono rappresentati i valori del punteggio medio e della deviazione standard per ogni item. Nelle colonne sono indicati gli item, mentre nelle righe sono indicate le misure a cui fanno riferimento. La condizione onesta viene abbreviata con H, la condizione disonesta con D.*

Un'analisi esplorativa dei dati appena esposti permette di affermare che solo l'*item* P1 ("*Tendo a non provare rimorso"*) mostra una media più alta nella condizione *dishonest* rispetto alla media dello stesso *item* nella condizione *honest.*

Successivamente, è stata calcolata la differenza tra le risposte grezze dei partecipanti nella condizione disonesta rispetto alla condizione onesta, al fine di indagare eventuali modifiche attuate dal soggetto nel rispondere al questionario ed una possibile strategia adottata dai mentitori. Nella Tabella b), sono riportate le percentuali delle risposte che dalla condizione *honest* alla condizione *dishonest* hanno subito un incremento (per esempio da 1 a 3) o un decremento (per esempio da 3 a 1), oppure che non hanno subito alcun cambiamento, rimanendo quindi invariate (per esempio da 3 a 3). È possibile notare che la percentuale maggiore delle risposte (51%) ha subito un decremento di punteggio. Il questionario, infatti, chiedeva al soggetto di dissimulare un determinato atteggiamento ed in linea con il concetto di *faking good* (si veda il capitolo 2 per una spiegazione più dettagliata) il soggetto nella condizione *dishonest* risponde attribuendo un punteggio più basso agli *item*, abbassando di conseguenza il punteggio (più alto è il punteggio più vi è la presenza del tratto malevolo di personalità). Le risposte che non subiscono alcuna modifica presentano una percentuale del 30%. La percentuale di risposte che, contrariamente alla direzione attesa, dalla condizione

onesta alla condizione disonesta va incontro ad un incremento del punteggio (19%) può essere interpretata in accordo all'effetto *test-retest*. Generalmente, somministrando lo stesso questionario allo stesso soggetto in due momenti diversi (per esempio a distanza di un'ora), le risposte subiscono delle piccole modifiche. Tale risultato potrebbe anche essere spiegato dalla presenza di differenti strategie di dissimulazione nella condizione disonesta; infatti, dal momento che i partecipanti sono stati istruiti a fingere in modo credibile, essi potrebbero aver attuato un approccio per rispondere alle domande scegliendo alcuni *item* specifici a cui non mentire. In Figura 2.1 è rappresentata visivamente la percentuale di *item* che ha subito una modifica o che è rimasta invariata nella condizione disonesta, rispetto alla condizione onesta.

Tabella b). *Numero di risposte che nella condizione dishonest subiscono modifiche rispetto alla condizione honest*

Totale risposte	Decremento risposte	Risposte invariate	Incremento risposte
5916	2995 (51%)	1770 (30%)	1139 (19%)

Note. *Nella prima colonna della Tabella b) è indicato il numero totale di risposte fornite nel complesso al questionario. Nelle colonne successive è indicato il numero di risposte che subiscono un decremento, che rimangono invariate o che subiscono un incremento nella condizione dishonest rispetto alla condizione honest. Infine, tra parentesi è indicato il rispettivo valore in percentuale.*

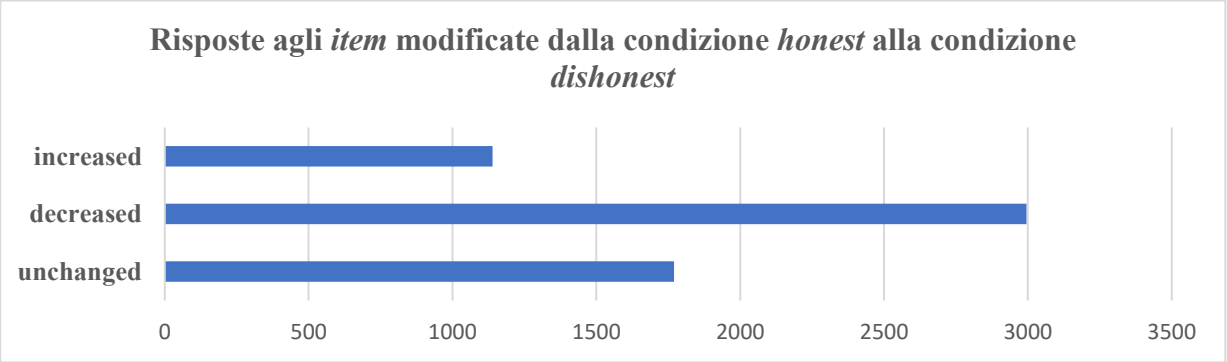


Figura 2.1. *Rappresentazione grafica delle modifiche subite dalle risposte dalla condizione dishonest rispetto alla condizione honest. In ascissa abbiamo il numero delle risposte; in ordinata abbiamo le risposte che dalla condizione onesta sono stata incrementate (increased), decrementate (decreased) o che sono rimaste invariate (unchanged).*

4.2 Analisi statistica dei dati grezzi

Al fine di verificare se sia presente una differenza statisticamente significativa tra la condizione onesta (H) e disonesta (D), emersa anche nelle analisi preliminari dei grafici delle medie, è stata condotta un'analisi statistica più approfondita.

Prima di proseguire con le analisi, è stata valutata la normalità della distribuzione dei punteggi tramite il test di Shapiro-Wilk[5]. Questo test assume come ipotesi nulla che i dati provengano da una distribuzione normale. Se emerge un *p-value*> 0,05, allora l'ipotesi nulla viene confermata. Al contrario, nel caso in cui il *p-value* sia inferiore a 0,05 è molto probabile che la distribuzione dei dati non segua un andamento normale.

Nel presente caso, i valori del *p-value* sono inferiori a 0,001 per tutti gli *item*, ad eccezione del primo (P1), per questo si può rifiutare l'ipotesi nulla e ritenere che i dati in analisi non provengano da una distribuzione normale. La non normalità della distribuzione ci permette di scegliere i test da utilizzare nelle prossime analisi. Poiché i dati non sono distribuiti normalmente (per i risultati del test di Shapiro-Wilk), è stato utilizzato il test di Wilcoxon, equivalente del t-test di Student per dati non parametrici. Nella Tabella c) sono riportati i valori. In questo caso, un *p-value*> 0,05 confermerebbe l'ipotesi nulla, secondo cui le due distribuzioni non presentano differenze statisticamente significative. I *p-value* nella Tabella c), inferiori a 0,001, supportano dunque l'ipotesi alternativa e quindi una differenza significativa tra le due distribuzioni, fatta eccezione per l'*item* P1, per cui non risulta una differenza statisticamente significativa tra condizione onesta e disonesta.

Per quanto riguarda la dimensione dell'effetto (riferito al test di Wilcoxon, con esclusione dell'*item* n.1) il *range* dell'*effect size* è compreso tra 0,54 e 0,73. In accordo con la classificazione di Cohen, si tratta di un effetto con magnitudo *medium,* ad indicare una differenza tra i valori medi delle due distribuzioni mediamente elevata (in controtendenza con l'*item* P1 che presenta un *effect size* negativo, a testimoniare una prossimità nei valori medi delle condizioni H e D – Tabella c). Un valore *d* di Cohen superiore a 0,5 significa che i punteggi medi della condizione *dishonest* si accentuano di almeno mezza deviazione standard rispetto a quelli della condizione *honest*. Tale conclusione è valida per tutti gli *item* ad esclusione di P5 e N12, i cui *effect size* risultano più bassi (0,33 e 0,42; *small*).

[5]Pubblicato nel 1965 da Samuel Sanford Shapiro e Martin Wilk, è uno dei test statistici più potenti per valutare la normalità della distribuzione (https://en.wikipedia.org/wiki/Shapiro-Wilk_test).

Tabella c). *Valore statistico e p-value per ciascun item in riferimento al test di Wilcoxon.*

Paired Samples T-Test

Measure 1		Measure 2	W	df	p	Rank-Biserial Correlation
P1_H	-	P1_D	18850.500		1.000	-0.530
N2_H	-	N2_D	64814.000		< .001	0.727
P3_H	-	P3_D	40946.500		< .001	0.615
M4_H	-	M4_D	47092.500		< .001	0.578
P5_H	-	P5_D	26328.500		< .001	0.329
M6_H	-	M6_D	36487.500		< .001	0.584
N7_H	-	N7_D	54399.500		< .001	0.535
N8_H	-	N8_D	64790.500		< .001	0.691
M9_H	-	M9_D	34160.500		< .001	0.575
P10_H	-	P10_D	58779.000		< .001	0.599
M11_H	-	M11_D	44344.000		< .001	0.738
N12_H	-	N12_D	43461.500		< .001	0.423

***Note**. In tabella si presenta il confronto tra lo stesso item nelle due diverse condizioni (H/honest e D/dishonest). Nella colonna W è indicato il valore statistico di Wilcoxon, mentre nella colonna p è rappresentato il p-value. Poiché questo risulta essere sempre <001, ad eccezione del primo item (P1), si può rifiutare l'ipotesi nulla e affermare che vi è la presenza di una differenza significativa tra le due condizioni in tutti gli item, escluso P1. In particolare, l'ipotesi alternativa specifica che Measure 1 è più grande di Measure 2 (per esempio, N2_H è più grande di N2_D). Infine, nell'ultima colonna è indicato il valore di effect size.*

Matrici di correlazione

In figura 2.2 vengono riportate le matrici di correlazione lineare realizzate tramite *r* di Pearson sui punteggi grezzi rispettivamente delle condizioni *honest-honest, dishonest-dishonest* e *honest-dishonest.* È evidente come la correlazione nelle condizioni *honest-honest* e *dishonest-dishonest* (Figura 2.2 A e B) appaia moderata, con valori che si aggirano attorno allo 0,5. Nella Figura 3.2 C, ovvero nella condizione di paragone tra *honest* e *dishonest,* tale correlazione si avvicina allo zero. Un valore di zero è indice di un'assenza di correlazione lineare; quindi, le risposte oneste correlano in misura debole con quelle disoneste, per cui non vi è una procedura immediata per predire le risposte disoneste. Questo dimostra quanto la predizione dell'esatta risposta dissimulata all'interno di un questionario partendo dai soli dati grezzi sia un'operazione tutt'altro che banale.

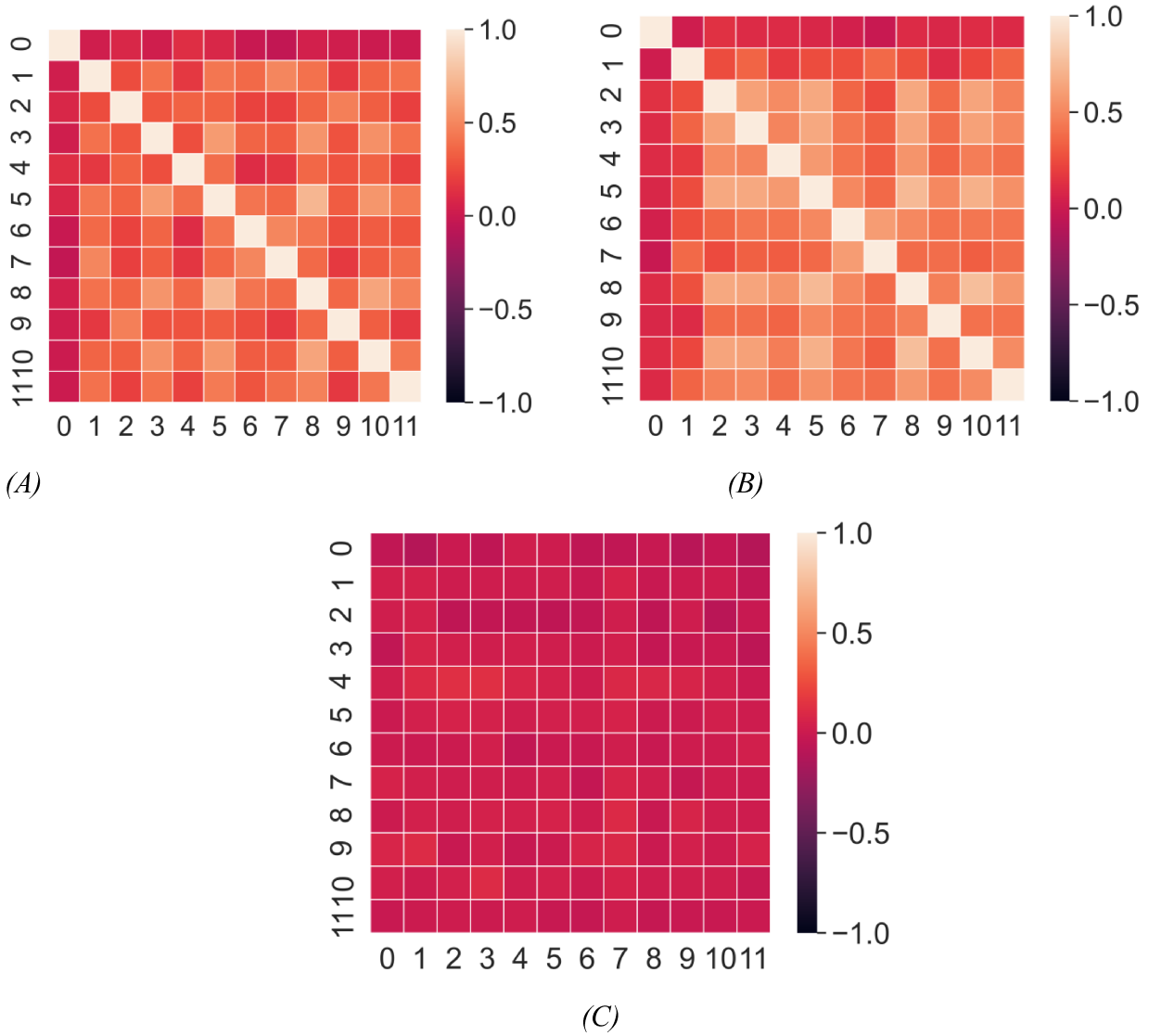


Figura 2.2. *Matrici di correlazione dei confronti tra le condizioni honest e dishonest. Nello specifico: A) confronto honest/honest; B) confronto dishonest/dishonest ed infine C) matrice di correlazione honest/dishonest. Si può notare come nella condizione honest/dishonest, a differenza delle altre due matrici, la correlazione si avvicini allo 0.*

Test Kruskal-Wallis (ANOVA)

Allo scopo di verificare l'interazione della condizione sperimentale (*honest/dishonest*) con le altre variabili qualitative indagate, cioè genere, età, anni di scolarizzazione e presenza o assenza di figli, è stato effettuato un test sulla varianza. A seguito dei risultati riportati nel test di Shapiro-Wilk, si assume che la distribuzione dei valori campionari non segua una distribuzione normale. Di conseguenza, è stato utilizzato il test Kruskal-Wallis, il corrispettivo non parametrico dell'ANOVA. Il test Kruskal-Wallis è stato condotto su tutti gli *item* del questionario all'interno delle due condizioni sperimentali.

Si specifica che un valore *p-value*< 0,05 indica la presenza di una differenza significativa tra le distribuzioni *honest* e *dishonest,* dipendente dalla variabile qualitativa considerata.

GENERE. Sono state condotte le analisi per ciascun *item* del questionario, con il risultato che tale variabile ha influenzato la risposta dei soggetti solo in determinati *item* e specificamente nella condizione *honest.* In particolare, il genere "maschio" presenta punteggi più elevati rispetto al genere "femmina" in tutti gli *item* significativi. Di seguito sono riportati i nomi degli *item* e il corrispondente *p-value:* P1_H (*p-vaue* = 0,025); P3_H (*p-value* = 0,027); P5_H (*p-value* = 0,001); P10_H (*p-value* = 0,002); M6_H (*p-value* = 0,006); N7_H (*p-value* = 0,026).

In Tabella d) sono riportati i valori della media e della deviazione standard dei punteggi di ciascun *item* elencato.

Tabella d). *Valori di media e di deviazione standard dei punteggi dei soggetti, suddivisi secondo il genere.*

ITEM	**MEDIA**		**DEVIAZIONE STANDARD**	
	M	**F**	**M**	**F**
P1_H	2,524	2,260	1,100	0,982
P3_H	2.202	1,949	1,133	1,037
P5_H	2,065	1,659	1,167	0,925
P10_H	2,984	2,583	1,236	1,182
M6_H	2,169	1,889	1,080	1,030
N7_H	3,00	2,715	1,249	1,129

Note. *In tabella sono presentati i valori della media e della deviazione standard per quanto riguarda gli item significativi a seguito del test Kruskall Wallis. Tali item sono indicati nelle righe (H, honest). Sia la colonna della media che la colonna della deviazione standard sono a loro volta suddivise secondo il genere: M, maschio e F, femmina.*

ETÁ. Dalle analisi condotte sono emersi alcuni *item* significativi nella condizione *honest* e un *item* significativo nella condizione *dishonest.*

Nella condizione onesta sono risultati significativi i seguenti *item*: P3_H (*p-value*< 0,001); P10_H (*p-value*< 0,001); M11_H (*p-value*< 0,001); N2_H (*p-value* = 0,027); N8_H (*p-value* = 0,004); N12_H (*p-value* = 0,033).

Per quanto riguarda la distribuzione disonesta l'unico *item* significativo è risultato essere P3_D (*p-value* = 0,044).

È stato poi condotto un test *Post Hoc* non parametrico sugli *item* sopracitati, ovvero il test di Dunn, allo scopo di verificare quali classi di età in particolare differissero significativamente tra di loro. Per una consultazione precisa dei valori si rimanda alle tabelle riportate in Appendice – B. In generale, è emersa una tendenza dei soggetti più giovani (appartenenti alla classe 18-29 anni) a totalizzare punteggi maggiori rispetto alle altre categorie d'età, in entrambe le condizioni.

ANNI DI SCOLARIZZAZIONE. Anche questa variabile è risultata avere un'influenza sulle risposte dei soggetti in entrambe le condizioni (*honest* e *dishonest*).

In particolare, per quanto riguarda la condizione onesta gli *item* significativi sono i seguenti: P3_H (*p-value* = 0,004); M11_H (*p-value* = 0,015); N2_H (*p-value* = 0,001); N7_H (*p-value* = 0,002); N8_H (*p-value* = 0,001).

Nella distribuzione disonesta invece è emersa una significatività per quanto riguarda gli *item*: P3_D (*p-value* = 0,006); M9_D (*p-value* = 0,040).

Sempre in riferimento agli *item* significativi è stato nuovamente condotto un test *Post Hoc* (test di Dunn, vedi tabelle in Appendice – C), per verificare quale gruppo si differenziasse maggiormente dagli altri. In generale, nella condizione *honest* è emersa una tendenza a riportare punteggi superiori nei soggetti con alta istruzione (dalla scuola superiore fino alla laurea magistrale). Invece, nella condizione *dishonest* i punteggi più alti sono stati prevalentemente ottenuti con scolarità di 13 anni (scuola superiore).

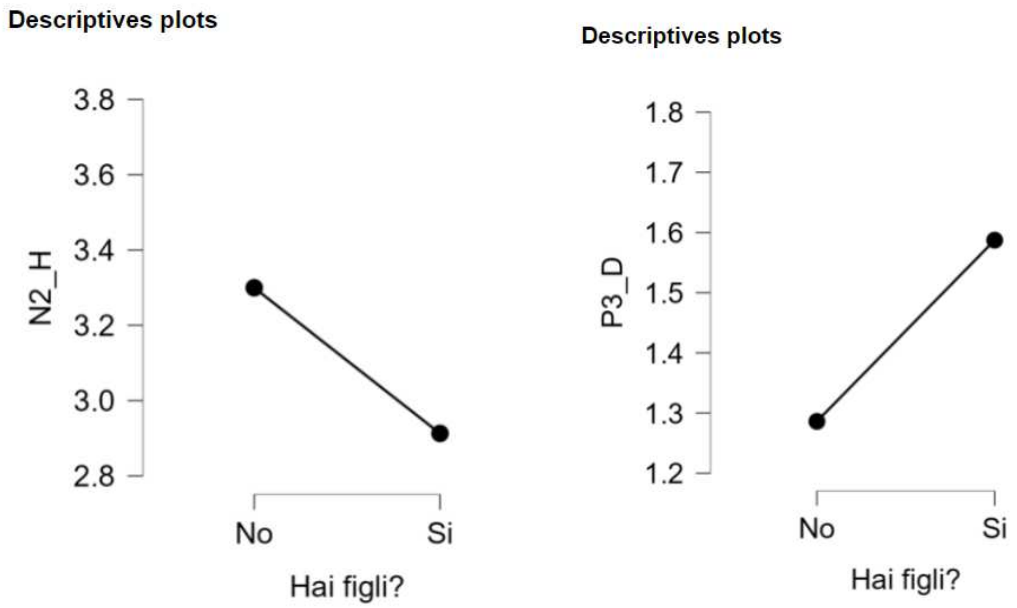
FIGLI. Anche la variabile riguardante la presenza o l'assenza di figli è risultata essere significativa sia per la condizione onesta che per la condizione disonesta.

In particolare, per la condizione *honest* sono risultati significativi i seguenti *item*: P3_H (*p-value*< 0,001); P10_H (*p-value*< 0,001); M11_H (*p-value*< 0,001); N2_H (*p-value*< 0,001); N7_H (*p-value* = 0,021); N8_H (*p-value*< 0,001).

Nella condizione *dishonest* gli *item* significativi sono risultati i seguenti: P3_D (*p-value*< 0,001); P5_D (*p-value* = 0,044); M6_D (*p-value* = 0,033).

A titolo esemplificativo, in Figura 2.3 riportiamo i *descriptives plots* per gli *item* N2_H (Figura 2.3 A) e P3_D (Figura 2.3 B). In generale, nella condizione *honest,* i soggetti che non avevano figli

hanno ottenuto punteggi più bassi; al contrario, nella condizione *dishonest,* i soggetti che avevano figli hanno ottenuto punteggi più alti.



(A) *(B)*

Figura 1.3. *In figura sono rappresentati i descriptives plots degli item N2_H (A) e P3_D (B) risultati significativi per la variabile Figli. In ascissa abbiamo la condizione "No" (figli) e "Sì" (figli); in ordinata abbiamo i punteggi medi ottenuti agli item presi in considerazione.*

4.3 TF-IDF

Concluse le analisi sui dati grezzi, i dati sono stati trasformati in valori TF-IDF, seguendo le formule riportate nel capitolo II, al fine di valutare l'ipotesi (H3), per cui l'indice TF-IDF permetterebbe una discriminazione più accurata tra la distribuzione dei dati onesti e disonesti rispetto all'utilizzo dei soli dati grezzi, attraverso un confronto dei rispettivi risultati.

Si specifica che nel calcolo del TF-IDF dei *dishonest*, è stato utilizzato il valore IDF relativo alle risposte degli *honest*, in modo tale da valutare la maniera in cui i *dishonest* si collocavano all'interno della distribuzione onesta, per cui la risposta dissimulata avrà un valore TF-IDF tanto alto quanto la risposta del partecipante nella condizione dissimulata sarà diversa dalle risposte allo stesso *item* degli altri partecipanti nella condizione onesta.

Si evidenzia come le due distribuzioni dei valori TF-IDF assumano un orientamento opposto rispetto alle distribuzioni dei dati grezzi. Infatti, la distribuzione dei valori dissimulati ottiene punteggi maggiori rispetto alla distribuzione dei punteggi onesti. Invece, con l'utilizzo dei dati grezzi avviene il contrario. Questa inversione è dovuta al fatto che l'indice TF-IDF, essendo un indice di anomalia, assume punteggi tanto maggiori quanto più le risposte a cui si riferisce sono anomale – o dissimulate, come in questo caso.

KL-Divergence (KLD)

In primo luogo, è stato indagato, il *KL-divergence* (KLD), una misura che si basa sull'entropia e che permette un paragone tra due diverse distribuzioni di probabilità. Il suo *range* varia da 0 a infinito, dove 0 è indice di una completa sovrapposizione delle due distribuzioni considerate (in questo caso *honest* e *dishonest*), mentre valori crescenti di KLD indicano un livello maggiore di separazione tra le distribuzioni esaminate, quindi un maggiore distanziamento tra di esse. Di seguito, riportiamo l'indice KLD calcolato per ciascun *item* sui dati grezzi e sull'indice TF-IDF (Tabella e):

Tabella e). *Valori KLD per le distribuzioni Honest e Dishonest calcolato sui punteggi grezzi e sui valori TF-IDF .*

MISURA	**P1**	**N2**	**P3**	**M4**	**P5**	**M6**
DATI GREZZI	0,35	0,62	0.42	0,41	0,15	0,37
TF-IDF	0,82	4,30	2,19	2,39	2,05	1,04
MISURA	**N7**	**N8**	**M9**	**P10**	**M11**	**N12**
DATI GREZZI	0,24	0,41	0,26	0,30	0,53	0,15
TF-IDF	1,39	4,95	1,77	1,91	1,06	1,33

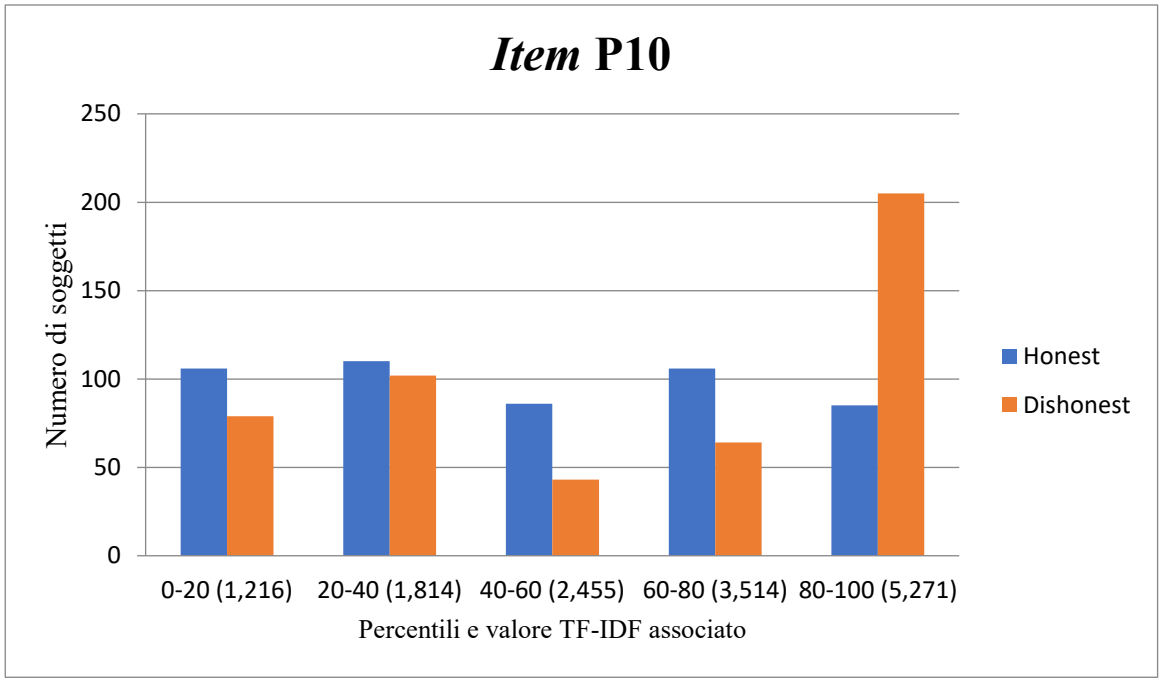
Note. *In tabella sono indicati i valori KLD di ciascun item, sia per quanta riguarda i dati grezzi che il TF-IDF. Maggiore è il valore KLD, tanto più le distribuzioni sono discriminabili tra loro. Si noti come i valori KLD corrispondenti al TF-IDF, siano più elevati rispetto a quelli dei dati grezzi, ad indicare che le distribuzioni sono più discriminabili nella prima condizione.*

Dall'osservazione della Tabella e) si evince che l'indice KLD mostra valori più elevati tra le distribuzioni delle risposte *honest* e *dishonest* quando esso viene calcolato sull'indice TF-IDF rispetto a quando viene calcolato sui punteggi grezzi. Tale esito avvalora l'ipotesi inizialmente formulata, per cui la discriminazione tra risposte oneste e dissimulate appare più accurata mediante l'utilizzo del TF-IDF.

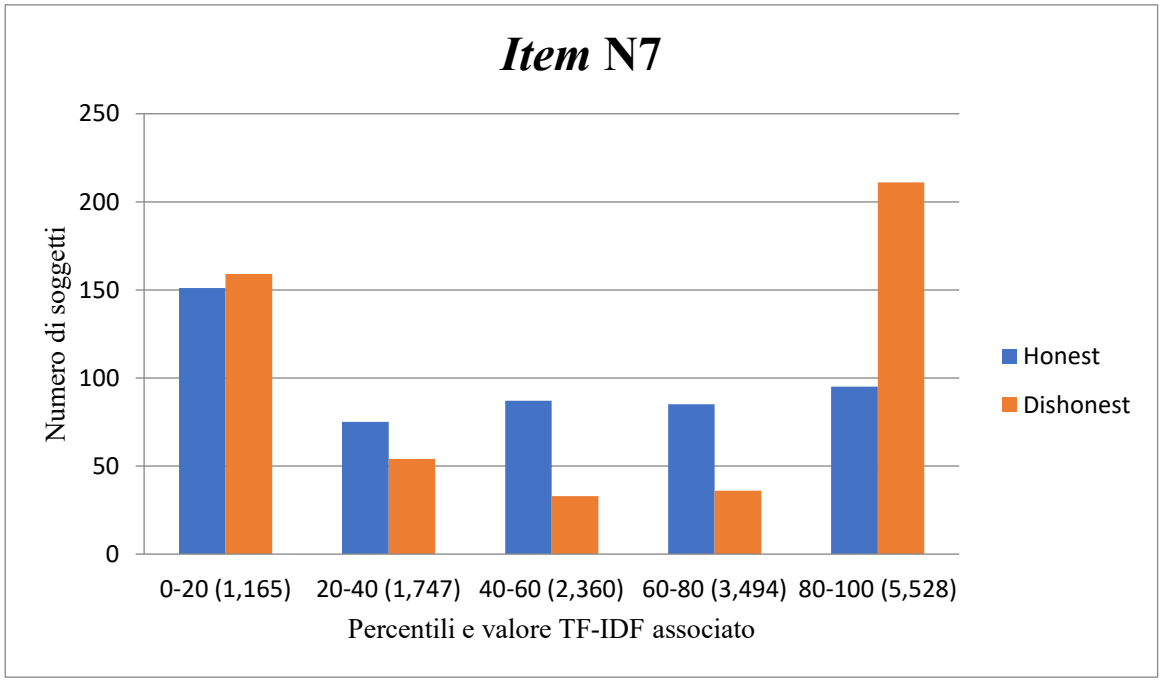
Distribuzione percentile di valori TF-IDF

Ad ulteriore validazione dell'ipotesi sopra descritta, relativa al fatto che l'indice TF-IDF sia in grado di discriminare meglio le risposte *honest* da quelle *dishonest* rispetto all'analisi dei soli dati grezzi, si valutano le distribuzioni dell'indice nelle due condizioni indagate.

Inizialmente, è stata suddivisa la distribuzione in cinque classi percentili: 0-20; 20-40; 40-60; 60-80 e 80-100 e per ciascun *item* sono stati identificati i valori TF-IDF ad essi associati (visibili tra parentesi nella Figura 3). Si precisa che i valori dei percentili che suddividono la distribuzione disonesta sono i medesimi della distribuzione onesta. Questo perché lo scopo è quello di identificare quanti soggetti *dishonest* ricadono all'interno degli intervalli della distribuzione *honest*. Successivamente, è stata calcolata la frequenza relativa di ciascun percentile, ovvero il numero di soggetti che ricadono in ciascun intervallo. Appaiando infine le frequenze degli *honest* a quelle dei *dishonest*, sono stati costruiti 12 istogrammi, corrispondenti al numero di *item* del questionario. A titolo esemplificativo, si riportano i risultati conseguiti per gli *item* P10 (Figura 3 A) e N7 (Figura 3 B).



(A)



(B)

Figura 3. *Istogrammi raffiguranti le distribuzioni dei valori TF-IDF nelle condizioni honest (colonne blu) e dishonest (colonne rosse). In entrambi gli istogrammi sull'asse delle ordinate è rappresentato il numero dei soggetti, mentre sull'asse delle ascisse i percentili. Tra parentesi è indicato il valore TF-IDF corrispondente a ciascun percentile. A) istogramma delle distribuzione dei valori TF-IDF nelle per l'item P10 e B) istogramma delle distribuzione dei valori TF-IDF per l'item N7.*

In entrambe le figure, si può notare la presenza di una notevole differenza tra l'andamento delle distribuzioni. Nella condizione *honest* (colonne blu in Figura 3), i soggetti risultano essere distribuiti uniformemente all'interno di tutte e cinque le classi in cui è suddivisa la distribuzione. Al contrario, nella condizione *dishonest* (colonne rosse), la maggior parte dei soggetti si colloca sopra l'ottantesimo percentile, ovvero nell'ultima classe. Questo dimostra che è possibile identificare un valore *cut-off* che permetta di distinguere i soggetti che hanno risposto in maniera sincera da quelli che invece hanno risposto dissimulando.

Quanto illustrato a titolo esemplificativo per gli *item* P10 e N7 (Figura 3) emerge a grandi linee anche in tutti gli altri *item*, ad eccezione dell'*item* P1 che viene raffigurato di seguito in figura 3.1:

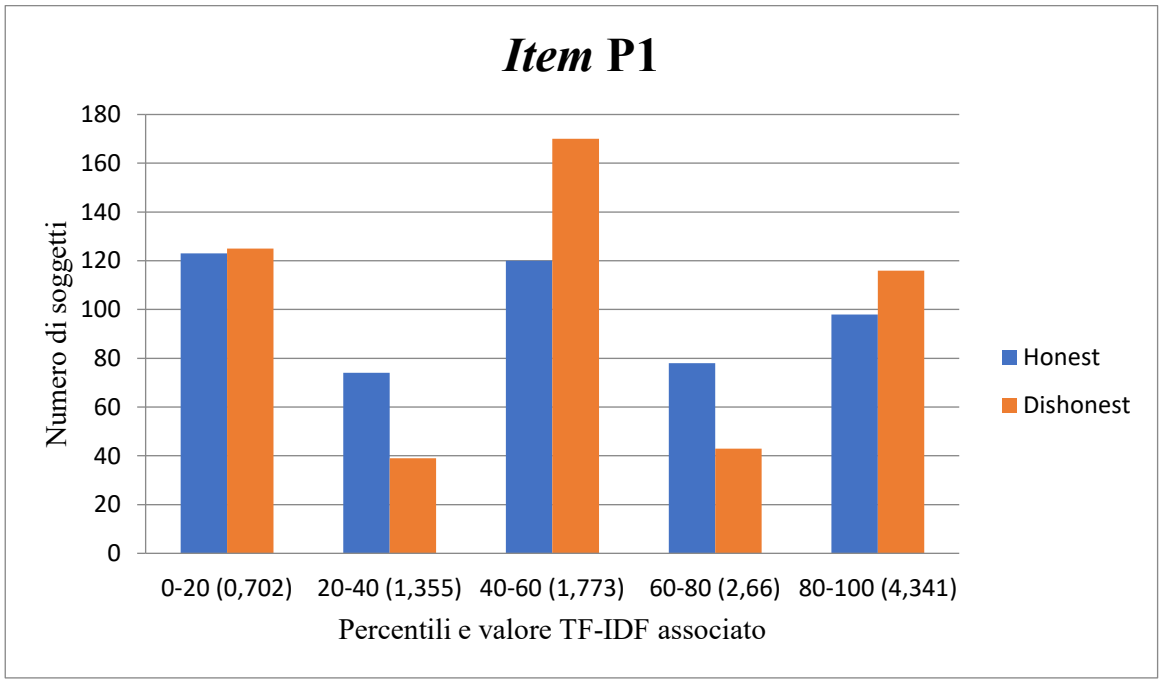


Figura 3.1. *Istogramma raffigurante le distribuzioni dei valori TF-IDF nelle condizioni honest (colonne blu) e dishonest (colonne rosse). Sull'asse delle ordinate è rappresentato il numero dei soggetti, mentre sull'asse delle ascisse i percentili. Tra parentesi è indicato il valore TF-IDF corrispondente a ciascun percentile.*

Per *l'item* P1 la maggior parte degli individui nella condizione *dishonest* si colloca nel primo percentile (0-20) e, sempre in questa condizione, i soggetti risultano maggiormente distribuiti all'interno delle cinque classi percentili; a differenza della condizione *honest* in cui la maggior parte dei soggetti si colloca sopra l'ottantesimo percentile. Questo risultato sembra in linea con la parziale sovrapposizione delle due distribuzioni (onesta e disonesta) emersa anche nelle analisi precedenti.

A titolo informativo, nella Tabella f) è riportato il numero di soggetti che ricade all'interno di ciascun percentile, per quanto riguarda sia la condizione *honest* che la condizione *dishonest*.

Tabella f). *Numero di soggetti che ricadono all'interno di ciascun percentile a livello di ogni item.*

ITEM	Perc. 0-20		Perc. 20-40		Perc. 40-60		Perc. 60-80		Perc. 80-100	
	H	D	H	D	H	D	H	D	H	D
P1	123	125	74	39	120	170	78	43	98	116
N2	112	93	116	96	86	16	84	104	95	184
P3	104	34	124	78	78	122	93	142	94	117
M4	103	64	119	56	74	53	115	152	82	168
P5	132	43	66	44	107	130	91	197	97	79
M6	129	70	87	79	87	76	96	146	94	122
N7	151	159	75	54	87	33	85	36	95	211
N8	101	84	114	61	106	77	75	61	97	210
M9	111	46	118	95	70	80	106	160	88	112
P10	106	79	110	102	86	43	106	64	85	205
M11	113	63	119	92	79	81	86	132	96	125
N12	103	102	106	66	91	68	102	70	91	187

***Note**. In tabella è rappresentato il numero di soggetti, per ogni item (indicati nelle righe), che ricadono all'interno di ciascun percentile (abbreviato con perc.), sia per quanto riguarda la condizione onesta (H), che la condizione disonesta (D). I percentili sono espressi in colonna.*

Odds di probabilità

Un ulteriore calcolo che è stato realizzato, prendendo in considerazione le classi percentili dei valori TF-IDF ottenuti nelle due condizioni sperimentali, è quello degli *odds* di probabilità. In particolare, l'*odds* di probabilità è stato calcolato per il percentile 80-100, riportato in Tabella g), poiché si è riscontrato che in tale intervallo ricade la maggior parte dei punteggi dei *dishonest*. L'*odds* di probabilità rappresenta il rapporto tra il numero dei disonesti ed il numero di onesti per quanto riguarda il percentile in esame. In altre parole, quanti soggetti disonesti sono presenti rispetto ai soggetti sinceri all'interno di una classe percentile, per ciascun *item* (ad esempio, un *odds* pari a 6, per il percentile 80-100 di uno specifico *item*, indica che ci sono 6 soggetti disonesti per ogni soggetto onesto). Si può notare come in tutti gli *item,* ad eccezione dell'*item* P5, ci sia un rapporto uguale o maggiore di 1, fino ad arrivare ad un rapporto di 2:1. Possiamo quindi concludere che i disonesti nel percentile 80-100 sono sempre in numero maggiore rispetto agli onesti nella classe percentile 80-100 in quasi tutti gli *item.* Nel caso dell'*item* n.5, l'*odds* è di poco inferiore ad 1

(0,814) ed il picco di soggetti si colloca nel percentile 60-80. Mentre, gli *item* che hanno ottenuto punteggi *odds* maggiori sono i seguenti: M4, N7, N8, P10 e N12 (Tabella g).

Tabella g). *Odds di probabilità onesti - disonesti, in riferimento al percentile 80-100.*

ITEM	**P1**	**N2**	**P3**	**M4**	**P5**	**M6**
Odds 80-100	1,184	1,937	1,245	2,049	0,814	1,298
ITEM	**N7**	**N8**	**M9**	**P10**	**M11**	**N12**
Odds 80-100	2,221	2,165	1,273	2,412	1,302	2,055

***Note**. Nelle righe è rappresentato il valore dell'odds di probabilità per ciascun item (indicato nelle colonne).*

Valutazione della performance

In seguito, è stata condotta un'ulteriore verifica sulla capacità di discriminazione dell'indice TF-IDF per le distribuzioni *honest* e *dishonest*. A tal fine, è stata effettuata una ricerca del valore soglia più accurato (*cut-off*) per ciascun *item*, attraverso la *k-fold cross validation*[6] (*k* =10). Suddetto *cut-off*, il medesimo per tutti gli *item*, ha lo scopo di categorizzare le risposte quando esse sono anomale. In altre parole, attraverso questa procedura di comparazione, se i valori TF-IDF dovessero superare tale valore, allora le risposte verrebbero categorizzate come anomale e quindi come dissimulate; al contrario un punteggio più basso di tale *range* sarà indice di sincerità.

Nella presente ricerca, l'intero campione è stato suddiviso in 10 sezioni. Si è proceduto poi, all'esclusione di una di queste parti e al *training* sulle restanti 9, allo scopo di individuare il valore soglia migliore, valido per tutti gli *item*, che permettesse l'identificazione delle risposte anomale. Il gruppo escluso era quello su cui veniva testato il *cut-off* emerso in fase di *test*. Tale procedura è stata successivamente iterata escludendo un secondo gruppo ed effettuando il *training* sui restanti. In questo modo si è potuto stimare e scegliere il miglior percentile usando il 90% dei dati e allo stesso tempo iterare il processo per 10 volte senza mai mescolare *training set* e *test set*.

A conclusione dell'intero procedimento, sono stati ottenuti 10 valori percentili di *cut-off,* ovvero quelli che per ciascuna iterazione avevano dimostrato di possedere una precisione maggiore. Di questi 10 percentili è stato infine scelto il più frequente, ovvero quello che emergeva nel maggior numero di iterazioni e che quindi massimizzava la precisione nella discriminazione delle due

[6] La *k-fold cross validation*, o cosiddetta anche convalida incrociata, consiste nella suddivisione dell'intero *dataset* in k parti di uguale numerosità e, ad ogni passo, una di queste k parti viene ad essere quella di convalida, mentre le restanti costituiscono l'insieme di addestramento. In altre parole, si suddivide il campione in gruppi di numerosità uguale, si esclude poi iterativamente un gruppo alla volta e, con i gruppi non esclusi si cerca di predirlo, in maniera tale da verificare la bontà del modello di predizione utilizzato. (https://it.wikipedia.org/wiki/Convalida_incrociata)

distribuzioni (*honest* e *disonest*). Nel caso del presente studio il *cut-off* è rappresentato dal 55esimo percentile.

Per verificare l'efficacia della tecnica TF-IDF per la detezione della simulazione e per effettuare un paragone con le altre misure presentate nell'elaborato, utilizzate per discriminare i due gruppi (*honest* e *dishonest*), la stessa procedura è stata effettuata anche per i dati grezzi (*Distribution-based Model* - DM). Il DM, applicato ai dati grezzi e utilizzato come *baseline* di confronto, utilizza come criterio per classificare una risposta il confronto delle prestazioni del singolo soggetto con la distribuzione delle prestazioni del campione. Anche per suddetta metodologia sono stati evidenziati dei valori *cut-off* che permettessero la distinzione tra risposte sincere e simulate. In questo caso, il valore *cut-off* è rappresentato dal 95esimo percentile. Il confronto, quindi, è avvenuto tra la risposta del partecipante al singolo item e la distribuzione percentile della condizione onesta per quell'item. Se la risposta all'item in questione corrisponde ad un punteggio che supera il percentile precedentemente stabilito come soglia (95esimo percentile), il modello identifica quella risposta come dissimulata.

Per quanto riguarda il *Machine Learning*, invece è stato utilizzato il modello *Self-Attention Based Autoencoders*, anche noto come "SABA", precedentemente discusso nel capitolo II.

Infine, è stato implementato un ulteriore modello (*Multi-Label Classification task*) per valutare la prestazione degli approcci proposti nella detezione della dissimulazione in termini di *Precision*, *Recall*, *F1-score* e *Accuracy*. La Tabella h) riassume gli indici di *performance* dei tre metodi, permettendo un confronto immediato tra gli stessi.

Tabella h). *Valore degli Indici di performance indagati, per ciascuna metodologia (TF-IDF, dati grezzi e SABA).*

	Precision	***Recall***	***F1-score***	***Accuracy***
TF-IDF	0,6923	0,4116	0,4552	0,4498
Dati grezzi	0,6922	0,9460	0,7791	0,6644
SABA	0,6930	0,8240	0,7319	0,6261

Note. *In tabella sono indicati i valori degli indici di prestazione per ciascuna metodologia. Tali indici sono presentati nelle colonne, mentre le tecniche di riferimento sono presentate nelle righe (TF-IDF, dati grezzi e SABA).*

Gli indici presentati in Tabella g) sono così calcolati:

$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

Dove:

- TP = *true positive* (veri positivi), cioè numero di risposte simulate categorizzate correttamente dal modello come simulate;
- TN = *true negative* (veri negativi), cioè numero di risposte non simulate categorizzate correttamente come non simulate;
- FP = *false positive* (falsi positivi), cioè numero di risposte categorizzate come simulate, ma che in realtà sono sincere;
- FN = *false negative* (falsi negativi), cioè numero di risposte categorizzate come sincere, ma che in realtà sono simulate.

Quindi, la *Precision* è una misura che dipende dal numero di veri positivi e falsi positivi; il *Recall* è una misura che si basa sul numero di veri positivi e falsi negativi; l'*F1 score* è la media armonica delle misure precedenti; l'*Accuracy* è definita come il rapporto tra il numero di previsioni corrette (sia oneste che dissimulate) e il numero di tutte le previsioni corrette e sbagliate, quindi misura se l'approccio proposto sia in grado di classificare correttamente una risposta come onesta o dissimulata utilizzando un'attività di classificazione *Multi-Label.*

Dai risultati riportati nella Tabella h), si evince che a livello di *Precision* i tre indici sono allineati. Inoltre, dati grezzi e SABA presentano valori più elevati rispetto al TF-IDF nei tre indici di prestazione: *Recall, F1 Score* e *Accuracy*. Questo significa che per questo tipo di analisi, l'utilizzo della distribuzione dei dati grezzi si dimostra una metodologia utile ed informativa per la discriminazione delle due distribuzioni (*honest* e *dishonest*). Il TF-IDF non deve essere utilizzato come indice sostitutivo alle altre analisi, ma come valida aggiunta metodologica.

Di seguito (Figura 4) si propongono gli indici di performance del TF-IDF in base al numero crescente di risposte dissimulate, mentre in rosso è riportato l'andamento della Funzione di Densità di Probabilità (PDF), riferita al numero effettivo delle risposte dissimulate dai partecipanti durante la compilazione:

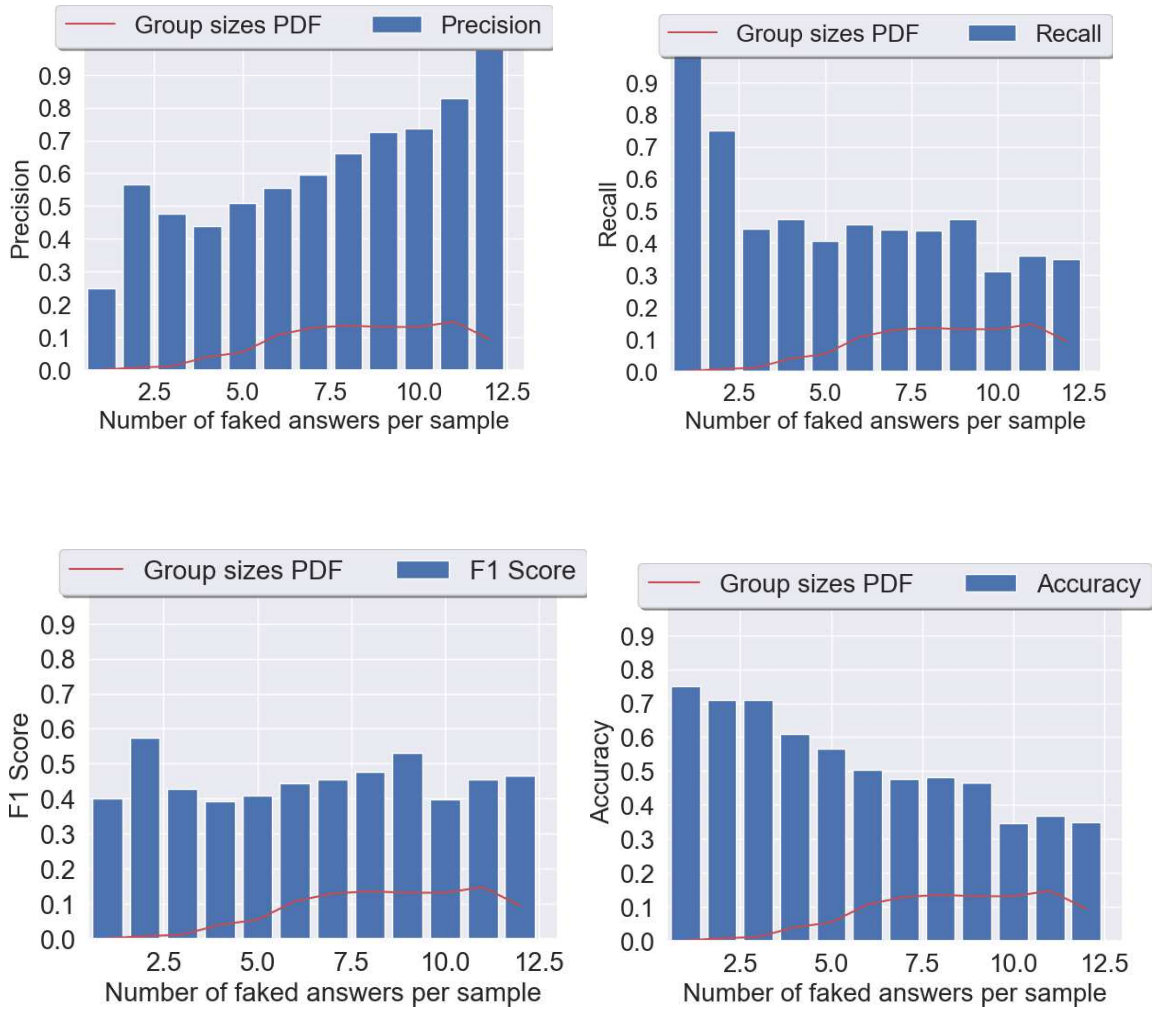


Figura 4. *Indici di performance riferiti al metodo TF-IDF. Per tutti i grafici sull'asse delle ascisse è rappresentato il numero di domande simulate, mentre sull'asse delle ordinate i valori da 0 a 1 per l'indice di prestazione a cui si riferisce il grafico.*

Esaminando l'andamento della funzione di densità di probabilità si evidenzia come pochissimi individui del campione mentano ad un numero di risposte compreso tra 0 e 3, ma anche tra 11 e 12; la maggior parte dei partecipanti ha la tendenza a dissimulare in un numero di *item* compreso tra 4 e 11.

L'accuratezza generale del metodo varia da 0,4 a 0,8 e presenta un andamento di crescita lineare solo per l'indice della *Precision.*

La *Precision* è in generale compresa tra 0,6 e 0,8 e tende a crescere con l'aumentare del numer delle risposte dissimulate, con un picco quando il numero di *item* simulati supera i 10. Questo indica che più elevato è il numero di risposte dissimulate, più alta è la *Precision.* Ad eccezione degli indici di *Accuracy* e *Recall* che sembrano avere un andamento opposto, si nota che l'indice di *Precision* assume valori minori con un numero ridotto di risposte dissimulate.

Di seguito (Figura 4.1) proponiamo anche i grafici degli indici di *performance* relativi al SABA, che complessivamente è risultato un metodo più informativo rispetto al TF-IDF:

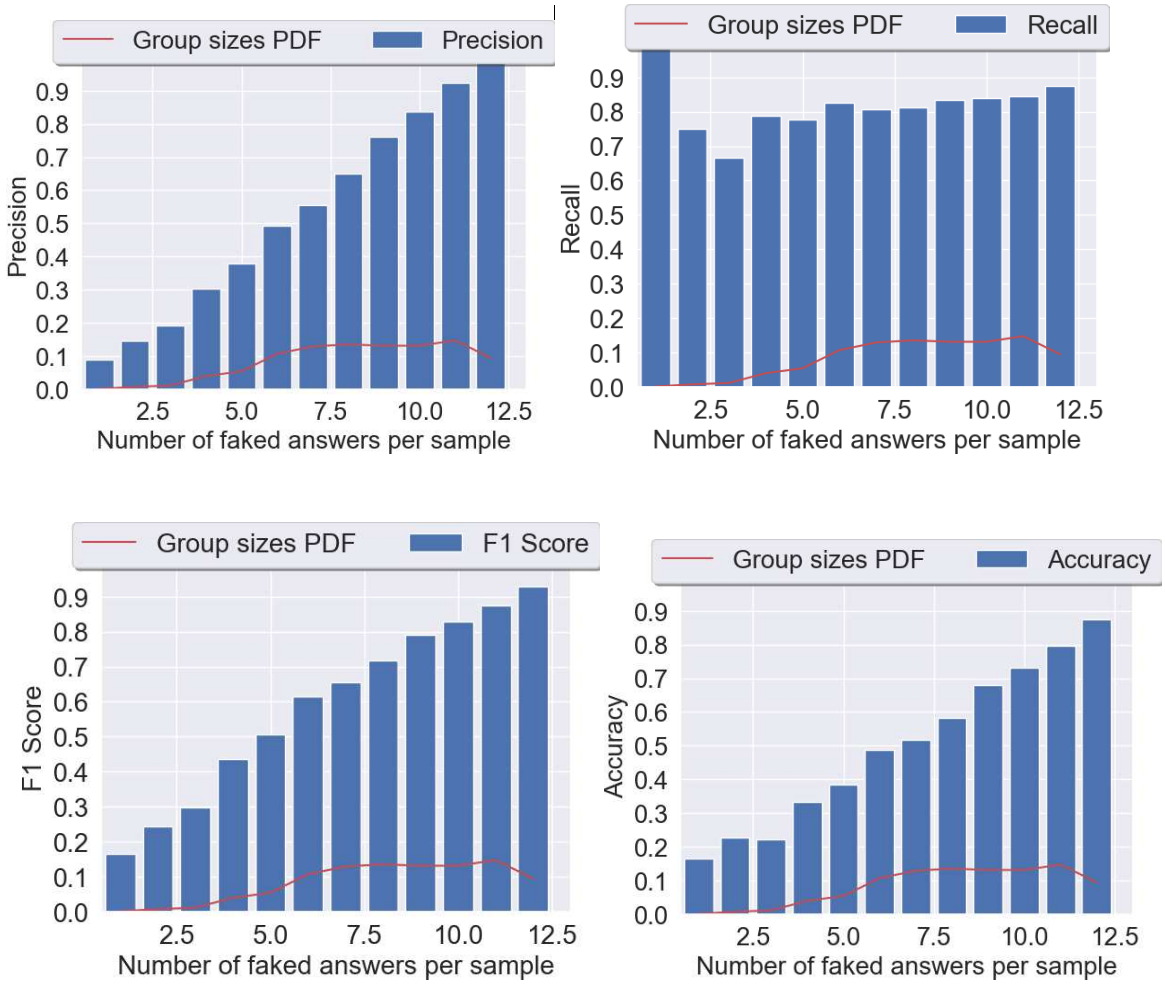


Figura 4.1.. *Indici di performance riferiti al metodo SABA. Per tutti i grafici sull'asse delle ascisse è rappresentato il numero di domande simulate, mentre sull'asse delle ordinate i valori da 0 a 1 per l'indice di prestazione a cui si riferisce il grafico.*

I grafici degli indici di prestazione del modello SABA ottenute utilizzando la strategia *Masked Language Model* (MLM-SA) sono rappresentati in Figura 4.1. Possiamo osservare come la funzione di densità di probabilità possieda un andamento crescente. Inoltre, anche in questo caso, nessun soggetto altera la propria risposta a meno di 3 domande. L'accuratezza generale del metodo varia da 0,5 a 0,9 ed essa tende ad aumentare al crescere delle domande in cui la risposta è alterata (ad eccezione dell'indice di *Recall* che non presenta questo andamento di linearità crescente), raggiungendo un picco di 0,9 in tutti e quattro gli indici di prestazione. Probabilmente l'elevato numero di dati a disposizione ha permesso al metodo di *Machine Learning* (SABA) di sfruttare al meglio le proprie potenzialità, riuscendo a compiere un *training* adeguatamente ampio.

Boxplot

Al fine di comprendere ulteriormente il funzionamento del metodo TF-IDF si propone in Figura 3.3 un grafico *boxplot* relativo al metodo TF-IDF. Nel grafico è illustrata la particolarità di tale metodo, ovvero la capacità di individuare la dissimulazione a livello del singolo *item* e non solo come comportamento generale.

Si propongono i risultati relativi al soggetto n. 9 (Fig. 5)

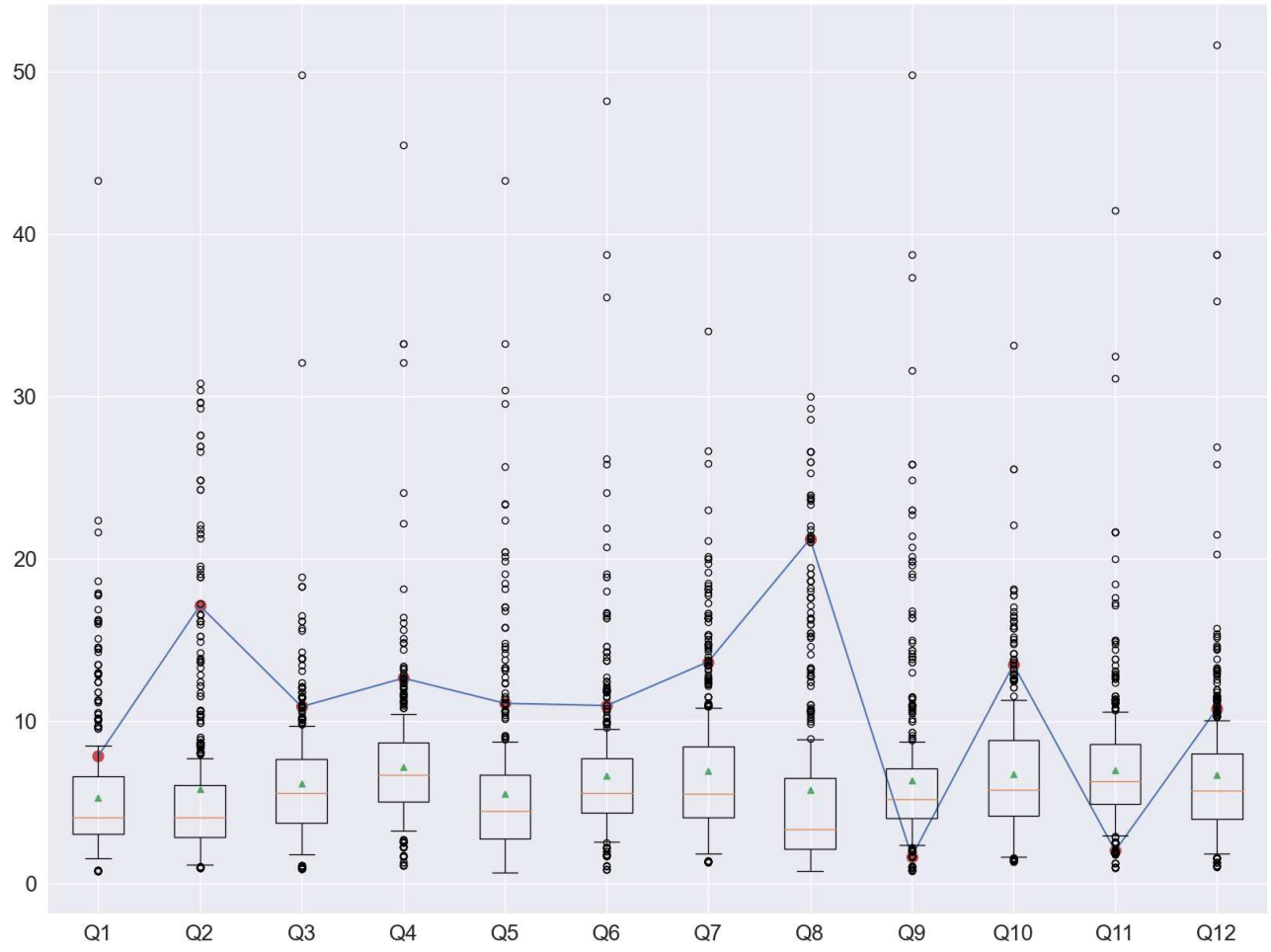


Figura 5. *Grafico boxplot raffigurante l'andamento dei valori TF-IDF nella condizione disonesta (linea blu nella parte alta del grafico) comparato con la distribuzione del TF-IDF nella condizione onesta (box nella parte inferiore). Sull'asse delle ascisse sono rappresentati i 12 item del questionario, mentre sull'asse delle ordinate i valori TF-IDF. Il pallino rosso indica che la risposta alla domanda cui fa riferimento è stata alterata rispetto alla condizione honest.*

Nel grafico in Figura 5 viene comparata la distribuzione *dishonest* dei punteggi del partecipante n. 9, con la distribuzione dei punteggi *honest*, riferiti all'intero campione di soggetti.

Sull'asse delle ascisse sono rappresentati i 12 *item* del questionario, mentre sull'asse delle ordinate sono raffigurati i valori TF-IDF. La distribuzione onesta si colloca nella parte inferiore del grafico

ed è raffigurata attraverso i *box*. La linea arancione all'interno di ciascuno di essi, indica il punteggio medio dello specifico *item*, mentre il primo e il terzo quartile (25esimo e 75esimo) sono identificati dagli estremi dello stesso *box*. Infine, il punteggio minimo e quello massimo coincidono con gli *whiskers* ("baffi") superiori e inferiori di ciascun *box*. La distribuzione dei punteggi disonesti ottenuti dal soggetto n. 9 invece, è raffigurata dalla linea blu, sulla quale sono presenti dei punti rossi. Questi puntini sono indice di alterazione della risposta rispetto alla condizione onesta, per lo specifico *item*. Questo permette di identificare quali sono le risposte che hanno presentato un'anomalia se paragonate alla condizione onesta, ovvero le risposte simulate.

Nella Figura 5, si può osservare che in tutti gli *item* vi è la presenza del pallino rosso, ciò significa che l'alterazione riguarda la totalità delle risposte del questionario, rispetto alla condizione onesta. Inoltre, delle 12 risposte simulate, 11 di queste sono riconosciute giustamente dal modello come alterate, infatti tra queste nessun pallino rosso ricade nel *box* onesto; mentre 1 risposta simulata non è stata correttamente individuata, collocandosi sotto il 75esimo percentile del *box* della distribuzione onesta (Q1). L'accuratezza risulta essere di 11/12: una risposta su 12 è risultata alterata rispetto alla condizione onesta, ma non correttamente individuata dal metodo.

CAPITOLO V

5 DISCUSSIONE DEI RISULTATI E CONCLUSIONI

In questo capitolo riportiamo un riassunto dello scopo della seguente ricerca, così da sintetizzarne brevemente le principali caratteristiche e circostanziare l'ambito del lavoro. Successivamente, procederemo nella verifica delle ipotesi alla base di tale elaborato, come descritte nel capitolo III, mediante l'interpretazione dei risultati emersi dalle analisi condotte. Infine, saranno esposti i limiti dello studio e i possibili sviluppi futuri della ricerca.

5.1 Sintesi degli approcci metodologici proposti

Lo scopo di questo studio è quello di testare l'efficacia di due tecniche di *lie detection* innovative, ovvero il TF-IDF (*Term Frequency-Inverse Document Frequency*) e il modello di *Machine Learning* noto come SABA, applicate al questionario *Dirty Dozen* (per la cui trattazione si rimanda al capitolo I) in un contesto simulato di valutazione per l'affido di minori. A differenza delle metodologie generalmente utilizzate in ambito forense, che riescono a cogliere la tendenza generale dell'individuo a simulare avvalendosi di scale di controllo appositamente inserite all'interno di test psicometrici (ad esempio MMPI-2) oppure di test psicologici costruiti *ad hoc* per cogliere la simulazione di disturbi specifici (ad esempio *Rey 15-Item Memory Test*, per la simulazione dei disturbi della memoria o il SIMS), il TF-IDF e il SABA possono essere applicati a qualsiasi tipo di test e hanno la peculiarità di riuscire ad identificare l'alterazione della singola risposta. In altre parole, permettono di valutare la simulazione e la dissimulazione a livello del singolo *item* e non solamente come tendenza generale di un soggetto.

Inoltre, un vantaggio specifico del TF-IDF è che due risposte uguali alla stessa domanda, date da due partecipanti diversi, possono dar luogo a valutazioni diverse a seconda sia delle risposte fornite da altri partecipanti allo stesso *item* sia delle risposte di ciascun soggetto alle altre domande del questionario; invece, nelle analisi tradizionali, risposte identiche di due soggetti allo stesso *item* vengono valutate in ugual modo. In altre parole, il TF-IDF implica la sostituzione dei punteggi grezzi aggregando in un'unica misura sia la distribuzione delle risposte di un gruppo di partecipanti ad uno specifico *item* sia lo stile di risposta del soggetto in questione in relazione alle altre risposte del questionario fornite dallo stesso (ad esempio, con quale frequenza il soggetto in esame seleziona una determinata risposta anche in altri *item*), fornendo un unico valore "carico di informazioni"

circa la veridicità delle risposte. In questo modo, anche se due soggetti rispondono nella medesima maniera alla stessa domanda, il TF-IDF risulterà molto probabilmente diverso, permettendo così un'analisi più informativa. Tale valore è poi interpretato come indice di anomalia. Questo significa che più il TF-IDF è elevato, più alta è la probabilità che il soggetto abbia mentito all'*item* cui il valore fa riferimento. Si ricava infine un *cut-off* per ogni singolo *item*, sopra il quale la risposta può essere categorizzata come menzognera. Tale valore soglia è ottenuto confrontando le risposte agli stessi *item* fornite da un gruppo di partecipanti che rispondono in maniera onesta.

Invece, per quanto riguarda l'applicazione del SABA possiamo innanzitutto affermare che l'alto numero di dati a disposizione ha influito positivamente sui risultati. Infatti, una più ampia disponibilità di dati permette una maggior quantità di dati *input* e di conseguenza un *training* più preciso. In contesti ecologici come quello forense, tale algoritmo è particolarmente vantaggioso in quanto, basandosi su un *set* di risposte oneste, non è necessario raccogliere un campione significativo e valido di risposte *faked*. Nella presente ricerca il campione disonesto è stato utilizzato solamente allo scopo di validare la *performance* dell'approccio proposto. Inoltre, similmente al TF-IDF, questa metodologia sfrutta anche le informazioni contestuali a disposizione e non solo la singola risposta. In altre parole, il SABA riesce ad individuare l'anomalia non solo basandosi sul singolo *item*, bensì prendendo in considerazione anche le altre risposte fornite da quel soggetto.

Le tecniche sopra descritte sono state applicate al questionario *Dirty Dozen* (Jonason & Webster, 2010) nella sua versione italiana da Schimmenti et al., (2017), costituito da 12 *item* che i partecipanti alla ricerca sono stati chiamati a compilare in una duplice versione: dapprima disonesta (D) immaginandosi in un contesto civile di affido di minori e in seguito onesta (H). Il questionario non presenta un *cut-off*, poiché misura semplicemente la presenza o assenza di tre tratti della personalità (Machiavellismo, Narcisismo e Psicopatia) e la loro intensità. Di conseguenza, i partecipanti allo studio sono stati scartati sulla base della correttezza o meno alle risposte di controllo poste prima dell'inizio del questionario e alla fine dello stesso.

Il campione su cui sono state svolte le analisi è formato da 493 soggetti.

5.2 Discussione dei risultati

Di seguito proponiamo un'interpretazione dei risultati ottenuti.

5.2.1 Dati grezzi

L'analisi grafica della figura 2 nel capitolo IV ("*grafico delle distribuzioni dei punteggi medi e delle deviazioni standard per singolo item degli onesti e dei disonesti*"), sottolinea innanzitutto un'evidente differenza tra la distribuzione dei punteggi onesti e quella dei punteggi disonesti. Infatti, questi ultimi si collocano in posizione inferiore rispetto ai primi, quindi il punteggio medio per singolo *item* è risultato sempre più basso del suo corrispettivo nella condizione *honest,* con scostamenti significativi, ad eccezione *dell'item* P1 ("*Tendo a non provare rimorso"*), dove ad occhio si può notare come le medie nelle due condizioni indagate non differiscano in modo significativo. Una possibile spiegazione per il risultato in controtendenza di questo *item* potrebbe essere una sua difficile interpretazione da parte dei rispodenti; in particolare, il contenuto dell'*item* risulterebbe particolarmente controverso. Oppure, un'altra spiegazione potrebbe risiedere nell'applicazione di una diversa strategia di dissumunazione; infatti, i partecipanti sono stati istruiti a fingere in modo credibile; quindi, potrebbero aver attuato un approccio dissimulatorio selettivo, scegliendo solamente alcune domande specifiche a cui mentire, rispondendo alle altre domande in maniera onesta.

In generale però, la differenza tra la distribuzione dei punteggi onesti e quella dei punteggi disonesti che si collocano in posizione inferiore rispetto ai primi, significa che in media i partecipanti tendono a dare delle risposte con un punteggio minore nel momento in cui sono chiamati a svolgere il questionario dissimulando, come gli era chiesto di fare dalle istruzioni, ponendosi nella condizione di dover mentire per risultare migliori agli occhi del Giudice per ottenere l'affido dei figli e di conseguenza nascondendo o non riportando alcune caratteristiche rispetto alla condizione onesta. Infatti, in un contesto di *faking good* i soggetti mostrano un più basso livello di *dark triad* come dimostrato dalla letteratura sul punteggio ottenuto ai questionari di personalità che può andare incontro ad un incremento o a un decremento nella media delle scale a seconda di cosa viene misurato da queste e dagli obiettivi dei rispondenti (Birkelon et al., 2006; Salgado et al., 2016).

Per quanto riguarda la variabilità delle risposte, si evidenzia che le deviazioni standard dei singoli *item* risultano tendenzialmente più elevate per la distribuzione onesta, dimostrando una maggior dispersione dalla media dei punteggi in tale condizione. Questo potrebbe dipendere da una strategia di dissimulazione che comporta la scelta di punteggi estremi, indipendentemente dal contenuto *semantico* dell'*item.* Prediligendo quindi le risposte estreme nella condizione *dishonest* si crea di conseguenza una minor dispersione del punteggio. Infatti, la letteratura nel contesto del *fake good* applicato ai questionari di personalità mostra una maggiore omogeneità delle risposte con una conseguente riduzione della deviazione standard comparata con la condizione onesta (Birkelon et al., 2006; Salgado et al., 2016).

Quanto affermato fin qui rappresenta una conferma, al momento parziale, della seconda ipotesi della ricerca relativa all'andamento dei dati grezzi (*i punteggi grezzi e i valori TF-IDF nella condizione honest sono superiori rispetto a quelli nella condizione dishonest*).

È stato successivamente condotto un test per dati non parametrici (test di Wilcoxon) - in quanto dal test di Shapiro-Wilk la distribuzione di dati è risultata non normale, con un *p-value*< 0,001 per tutti gli *item* ad eccezione dell'*item* P1 - per verificare se la differenza presentata tra le distribuzioni *honest* e *dishonest* risultasse significativa a livello statistico. Il *p-value*< 0,05 emerso in tutti gli *item* (Tabella c), ad eccezione del primo *item* (P1) risulta essere a supporto dell'ipotesi alternativa per cui le due distribuzioni non sono sovrapponibili. In questa maniera, si conferma la prima ipotesi (*i punteggi grezzi raccolti nelle condizioni dishonest e honest sono diversi tra loro in modo statisticamente significativo*) per cui nella condizione *honest* e *dishonest* i punteggi grezzi presentano una differenza statisticamente significativa tra loro. Tale ipotesi è vera per tutti gli *item* fatta eccezione per *l'item* P1 "*Tendo a non provare rimorso*" (*p-value* = 1,00) per cui le medie nelle due condizioni indagate non differiscono in modo statisticamente significativo.

Proseguendo nell'interpretazione, dalla lettura delle matrici di correlazione si nota che la correlazione lineare per quanto riguarda la combinazione dei punteggi grezzi *honest-dishonest* (figura 2.2 C*)*, si aggira attorno allo zero. Ciò significa che vi è una tendenza molto bassa delle due condizioni a covariare, ovvero a variare insieme. Si può quindi inferire che con il solo utilizzo dei punteggi grezzi sia difficile dedurre le risposte dissimulate partendo da quelle oneste. Questo dimostra che non esistono delle tecniche per comprendere quali e quanti *item* siano stati dissimulati da un soggetto partendo dall'analisi dei soli dati grezzi; il processo di finzione appare molto complesso e ogni individuo può assumere scelte e strategie diverse difficilmente prevedibili.

A questo punto sono state condotte alcune analisi sulla varianza attraverso il test di Kruskal-Wallis, equivalente dell'ANOVA, ma utilizzato su dati provenienti da una distribuzione non normale. È stato analizzato in che modo le variabili indagate nel questionario (genere, età, anni di scolarizzazione e presenza o assenza di figli) influissero sulle condizioni *honest* e *dishonest*. Per gli *item* risultati significativi è stato poi condotto un test *Post Hoc* (test di Dunn) per valutare quale tra i gruppi differisse significativamente dagli altri.

Per il genere è emersa una significatività in alcuni *item* solo nella condizione onesta, dimostrando come il genere "maschio" presenti punteggi più elevati rispetto al genere "femmina" in tutti gli *item* significativi. Questo dato è confermato dalla letteratura secondo cui il genere maschile otterrebbe punteggi maggiori in tutte e tre le componenti della *dark triad* (Paulhus & Williams, 2002). Questi risultati sono stati replicati anche in altre ricerche sull'argomento ed in particolare nello studio di Jonason (2013) è emerso come i maschi ottenessero punteggi più alti delle donne in tutte e tre le componenti della triade oscura a prescindere dalla loro origine (il campione infatti era formato da partecipanti provenienti da nazioni diverse). Secondo l'autore, questo dato potrebbe essere legato al fatto che i maschi generalmente ottengono più benefici e pagano meno costi rispetto alle donne, mettendo in atto strategie di vita legate alla *dark triad.* Tale evidenza viene confermata nel nostro studio anche se il campione di partecipanti è composto da un numero molto minore di maschi (25%) rispetto alle femmine (75%).

Anche per l'età sono emersi alcuni *item* significativi (vedi tabelle in Appendice – B), questa volta sia per la condizione onesta e nel caso di un solo *item* (P3) anche nella condizione disonesta. Per quanto non sia possibile individuare una vera e propria tendenza riscontrabile in tutti gli *item* significativi, si può affermare che in entrambe le condizioni i più giovani ("18-29 anni") possiedano un punteggio generalmente più alto rispetto alle altre categorie nella maggior parte degli *item.*

Anche la variabile "anni di scolarità", in alcuni *item* significativi (p-*value*< 0.05), è risultata avere un'influenza sulle risposte dei soggetti in entrambe le condizioni indagate. Si rimanda all'Appendice – C per le tabelle dei risultati del test Dunn. In generale, è emersa una tendenza a riportare punteggi superiori nei soggetti con alta istruzione (dalla scuola superiore fino alla laurea magistrale) per entrambe le condizioni.

La variabile "presenza o assenza di figli" è risultata essere significativa per entrambe le condizioni indagate. In particolare, la presenza di figli ha portato i soggetti ad ottenere punteggi più bassi nella condizione *honest;* al contrario l'assenza di figli ha portato i soggetti a riportare punteggi più bassi nella condizione d*ishonest*. Questo dato è in controtendenza rispetto a quello che ci saremmo

aspettati, ovvero una migliore dissimulazione per i rispondenti genitori, rispetto ai rispondenti non-genitori dovuta ad una prevedibile migliore capacità di immedesimarsi nella condizione descritta nelle istruzioni. È possibile che gli *item* influenzati significativamente da questa variabile abbiano risentito di una diversa strategia di dissimulazione o non siano stati correttamente interpretati dai rispondenti.

Infine, per valutare ulteriormente la discriminazione tra le distribuzioni *honest* e *dishonest* con i dati grezzi, è stato calcolato l'indice *KL-Divergence* (Tabella e). Si ricorda che il KLD possiede un *range* da 0 a infinito: più il numero è alto, maggiore sarà la differenza tra le distribuzioni confrontate. Dai risultati emersi in questa analisi, si può notare come i valori KLD siano generalmente contenuti tra 0 e 0,6 mostrando quindi che i dati grezzi possono essere intesi come indici mediocri per la discriminazione delle distribuzioni, non riuscendo a discriminare facilmente fra distribuzione di risposte oneste e distribuzione di risposte dissimulate. Nel paragrafo che segue si confronteranno tali risultati con quelli emersi per i valori TF-IDF.

5.2.2 TF-IDF

Una volta trasformati i punteggi grezzi in valori TF-IDF, si è in primo luogo svolto un confronto dei loro valori KLD con quelli appartenenti ai dati grezzi. In questo caso, il *range* è contenuto tra 2 e 4,6. Questo significa che l'indice KLD risulta tendenzialmente più elevato rispetto a quello dei dati grezzi, a dimostrazione che il TF-IDF presenta potenzialità più alte e permette una discriminazione più accurata tra le distribuzioni *honest* e *dishonest*.

Quest'analisi conferma la prima parte dell'ipotesi H3 (*la detezione della dissimulazione risulta migliore mediante l'utilizzo della tecnica TF-IDF o mediante l'utilizzo di algoritmi di Machine Learning rispetto alle analisi dei soli dati grezzi*).

In secondo luogo, si è svolta una valutazione della distribuzione dei valori assunti dall'indice TF-IDF calcolati per ciascun *item*. È emerso che la maggior parte delle risposte disoneste si colloca sopra un determinato valore percentile della corrispondente distribuzione onesta, ovvero quello che identifica l'ottantesimo percentile. In particolare, la distribuzione era stata suddivisa in 5 classi ed era emerso come per tutti gli *item* i soggetti si distribuissero in maniera differente a seconda della condizione a cui appartenevano. Infatti, mentre gli onesti si distribuivano in maniera uniforme in tutte e 5 le categorie, i disonesti si collocavano prevalentemente sopra l'ottantesimo percentile, dimostrando che è possibile identificare un valore *cut-off* che permetta di distinguere i soggetti che hanno risposto in maniera sincera da quelli che invece hanno risposto dissimulando. In particolare,

l'*item* P1, come già emerso dalle altre analisi, presenta un andamento opposto collocando la maggior parte dei partecipanti nel primo percentile (0-20).

Successivamente, attraverso il calcolo dell'*odds* di probabilità si valuta il rapporto tra i soggetti disonesti e onesti per la classe 80-100. Si è visto come esso sia sempre maggiore di 1, ad eccezione dell'*item* P5 (*"Tendo a non interessarmi della moralità delle mie azioni"*), dove infatti il maggior numero di soggetti si colloca nel percentile 60-80, ma comunque *l'odds* di probabilità si avvicina al valore di 1 (0,81). In generale, questo significa che per il percentile 80 il numero dei soggetti disonesti è sempre maggiore dei soggetti onesti in tutti gli *item* - ad eccezione dell'item n.5 che si colloca nel percentile 60-80 - con un picco del rapporto di 2:1 (2 soggetti disonesti ogni soggetto onesto). Questo andamento generale conferma l'ultima ipotesi (*i valori TF-IDF dishonest si collocano prevalentemente al di sopra di un determinato valore di soglia e, sopra tale valore, il rapporto tra risposte faked e risposte oneste è maggiore di* 1). Ciò va interpretato considerando come viene calcolato l'indice TF-IDF dei disonesti, per ciascun *item,* che è dato dal prodotto tra TF (numero di volte che compare una data risposta nella serie delle 12 risposte del partecipante) e IDF (valore connesso alla rarità della data risposta al preciso *item* nella totalità dei soggetti onesti). In sintesi, valori alti di TF-IDF delle risposte dissimulate indicano che il valore grezzo utilizzato per rispondere è raro nella condizione onesta e al contempo frequente nelle risposte disoneste del soggetto. Perciò, il fatto che la maggior parte dei valori TF-IDF delle risposte disoneste sia al di sopra di un dato percentile rispetto a quelle oneste, indica che in generale i valori TF-IDF degli onesti sono minori di quelli dei disonesti, confermando nuovamente l'ipotesi H3.

Per quanto riguarda l'ipotesi H2 (*la detezione della dissimulazione risulta migliore mediante l'utilizzo della tecnica TF-IDF o mediante l'utilizzo di algoritmi di Machine Learning rispetto alle analisi dei soli dati grezzi*), essa è parzialmente confermata. Infatti, la capacità di discriminare le due distribuzioni – e quindi identificare la dissimulazione - tramite le tre diverse metodologie (dati grezzi, TF-IDF e SABA) è stata valutata attraverso l'analisi della *performance* con *k-fold cross validation*, i risultati sono riportati in Tabella g). È stato quindi implementato un ulteriore modello (*Multi-Label Classification task*) che ha come *output* gli stessi indici di *performance* (*Precision*, *F1 score*, *Accuracy* e *Recall*). Considerando solo i risultati del *Multi-Label Classification Task*, essi potrebbero indurre a pensare che l'utilizzo dei dati grezzi conduca a risultati migliori rispetto alle metodologie innovative. Infatti, si nota che sia per i dati grezzi che per il TF-IDF che per il SABA, tutti e 4 gli indici di prestazione (*Precision*, *F1 score*, *Accuracy* e *Recall*) presentano valori relativamente alti. Tuttavia, tale modello utilizzato è solo l'ultimo di una serie di analisi volte alla valutazione dell'efficacia dei metodi presentati, analisi che in precedenza hanno ampiamente

dimostrato che l'uso del TF-IDF è decisamente più efficace nella detezione della simulazione. Inoltre, è importante sottolineare che l'indice TF-IDF non deve essere utilizzato come indice sostitutivo alle altre analisi, ma come valida aggiunta metodologica. Nel dettaglio, per il TF-IDF si può notare come la funzione di densità di probabilità per la *Precision* (rappresentata nella figura 3.1) cresca all'aumentare del numero di *item* a cui il soggetto mente: più alto è tale numero, maggiore è l'accuratezza del modello. Infatti, la bontà della tecnica TF-IDF per il valore di *Precision* cresce con il numero di risposte dissimulate (picchi più elevati delle colonne degli istogrammi verso destra). Inoltre, analizzando sempre l'andamento della funzione di densità di probabilità riferito al numero effettivo delle risposte dissimulate dai partecipanti, è emerso come pochi individui del campione mentissero ad un numero di risposte compreso tra 0 e 3, ma anche tra 11 e 12; la maggior parte dei soggetti ha la tendenza a dissimulare in un numero di *item* compreso tra 4 e 11. Infatti, i partecipanti sono stati istruiti a fingere in modo credibile, quindi potrebbero aver attuato un approccio dissimulatorio selettivo, scegliendo solo alcune domande specifiche a cui mentire e rispondendo alle altre in modo onesto e questo risultato dimostra come vi sia una scelta strategica da parte dei partecipanti in base al proprio stile di *faking*.

Infine, sono stati costruiti per ogni partecipante i *box-plot* della distribuzione del TF-IDF. Essi dimostrano come tale indice sia estremamente utile nell'identificazione della singola risposta dissimulata, informazione che non può essere ottenuta attraverso i dati grezzi. L'accuratezza risulta essere alta (11/12 nel grafico in figura 3.3). I *box-plot* mostrano in maniera graficamente chiara la distribuzione delle risposte disoneste del soggetto. La presenza di un pallino rosso a livello della risposta indica che essa è stata alterata. Se poi tale puntino ricade oltre il limite superiore del *box* allora si può concludere che la detezione è avvenuta correttamente, cioè la risposta è dissimulata e come tale è stata identificata dal modello.

5.2.3 Self-Attention Based Autoencoders, modello di Machine Learning (SABA)

La ricerca condotta nel presente studio ha evidenziato come il *Self-Attention Based Autoencoders* (SABA) abbia mostrato la sua validità di utilizzo nell'ambito della detezione della dissimulazione, come dimostrato dal *Multi-label Classification Task* (Tabella g) con indici di *performance* (*Precision*, *F1 score*, *Accuracy* e *Recall*) leggermente superiori rispetto a quelli ottenuti dalla tecnica TF-IDF e dall'utilizzo dei dati grezzi. L'ampiezza del campione (493 soggetti) ha permesso al modello di sfruttare le sue potenzialità, permettendogli un *training* più ampio e preciso. Questo conferma definitivamente la seconda ipotesi (*la detezione della dissimulazione risulta migliore*

mediante l'utilizzo della tecnica TF-IDF o mediante l'utilizzo di algoritmi di Machine Learning rispetto alle analisi dei soli dati grezzi).

Alla luce di quanto sopra esposto, si evince che l'indice TF-IDF ed il modello SABA di *Machine Learning* possono essere considerati come tecniche di detezione della dissimulazione basate sui singoli *item* efficaci ed innovative. Andranno eseguite ulteriori ricerche per confermare ed ampliare i risultati ottenuti in questo studio, il quale non è privo di limiti che possono aver interferito con una corretta valutazione della *performance* delle due metodologie.

5.3 Limiti dello studio e sviluppi futuri

Il primo limite dello studio è relativo alle verifiche connesse al corretto funzionamento delle tecniche di detezione della dissimulazione del TF-IDF e del modello di *Machine Learning* SABA. Nella presente ricerca tali verifiche sono state applicate al questionario *Dirty Dozen* (DD), il quale sicuramente non è privo di limiti. Il limite più significativo, esposto nel capitolo I, è l'eccessiva brevità del questionario composto da soli 12 *item*, che può aver portato alla rimozione di contenuti essenziali per la rappresentazione dei tratti indagati e quindi ad un minor livello di validità di costrutto, come dimostrato dalle correlazioni modeste fra le scale della DD e Narcisismo (.46) e Psicopatia (.42); e basse con Machiavellismo (.34). Tali limiti dello strumento sicuramente hanno avuto un loro peso nel determinare i risultati del presente studio. La DD, tuttavia, è solo uno dei questionari attualmente disponibili per la valutazione della *dark triad*, motivo per cui sarebbe interessante condurre un secondo studio analogo somministrando un differente questionario e comparare i risultati ottenuti in termini di *performance* dei modelli con quelli di questa ricerca.

Sebbene il nostro campione, di 493 partecipanti, costituisca un numero abbastanza elevato e abbia influito positivamente sui risultati ottenuti dal modello SABA, tale numerosità andrebbe comunque ampliata, poiché la metodologia di *Machine Learning* per essere correttamente implementata necessita di un maggior numero di dati *input* (ad esempio sarebbero ottimai 1000 partecipanti).

Un ulteriore limite potrebbe essere costituito dall'alto livello di scolarità dei partecipanti (valori in termini di anni di studi con M = 16,68; DS = 2,52): la maggior parte di essi, infatti, ha conseguito la Laurea Triennale. La scolarità poteva essere molto più variabile tra gli individui al fine di verificare il funzionamento dei modelli di analisi impiegati anche in relazione ad individui non appartenenti al contesto universitario e, probabilmente, meno istruiti sul tema della presente ricerca. Un'ipotesi che non si esclude, infatti, è che i partecipanti maggiormente istruiti, soprattutto coloro coinvolti in un

percorso di studi riguardante la psicologia, possano mettere in atto strategie di dissimulazione più efficienti o difficili da identificare e, soprattutto, che non creino delle differenze troppo evidenti rispetto alle *performance* oneste. Da un'altra prospettiva invece, la bravura dei partecipanti nella dissimulazione dei tratti della triade, potrebbe aver fornito un'ottima situazione di *training* per i modelli proposti, in quanto è stato possibile "allenarli" ad individuare anche le *performance* dissimulatorie più studiate ed elaborate (simili a quelle prodotte tramite *coaching*).

Ancora, un limite da tenere in considerazione riguarda il fatto che, alla luce delle restrizioni dovute al Covid-19 durante il periodo di somministrazione del questionario, la modalità di somministrazione e compilazione sono state totalmente da remoto, senza possibilità né di spiegare dal vivo le istruzioni del test né di poter chiedere ulteriori chiarimenti da parte dei partecipanti. Al termine della situazione pandemica sarebbe interessante somministrare il questionario a seguito di spiegazioni dal vivo circa le istruzioni preliminari alla compilazione. Tuttavia, le domande di controllo che sono state inserite per verificare la corretta comprensione da parte dei partecipanti, portando all'eliminazione di 107 soggetti che non avevano risposto correttamente a tali domande, ci ha in un certo senso tutelati dal rischio di incomprensione. Sempre relativamente alla modalità da remoto con cui si è svolta la partecipazione allo studio, questa può aver influito sull'età media (M = 32,87; DS = 11,54) ed il livello di istruzione dei nostri partecipanti, portando solo i soggetti più "tecnologici" a partecipare allo studio ed escludendo partecipanti con caratteristiche diverse.

Infine, va pur sempre considerato il fatto che i partecipanti sono stati "forzati" a fornire delle risposte dissimulate, immaginando di trovarsi in un contesto di valutazione per l'affido dei figli e di poter trarre un beneficio da tale dissimulazione. In futuro, potrebbe essere interessante applicare l'uso del TF-IDF e del SABA a contesti forensi reali, nei quali gli individui sono potenzialmente dei reali dissimulatori che possono trarre reali benefici dalle loro risposte. I risultati di queste ricerche più ecologiche potrebbero essere un banco di prova reale per l'impiego di tali tecniche, i quali andranno poi confrontati con quelli derivanti da studi su popolazioni diverse da quella forense ma anche con i risultati derivanti dall'impiego delle classiche metodologie di detezione della dissimulazione.

BIBLIOGRAFIA

Ackerman, M. J., & Ackerman, M. C. (1997). Custody evaluation practices: A survey of experienced professionals (revisited). *Professional Psychology: Research and Practice, 28*, 137-145.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders (3rd ed.).* Washington DC: Author.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, DSM-5.* Arlington, VA.

Amernic, J. H., and Craig, R. J. (2010). Accounting as a facilitator of extreme narcissism. *Journal of Business Ethics*, 96, 79-93.

Ames, R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality, 40,* 440–450.

Aziz, A. (2004). Machiavellianism scores and self-rated performance of automobile salespersons. *Psychol Rep.* 2004*; 94(2)*, 464-466.

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval* (Vol. 463). New York: ACM press.

Bagby, R. M., Nicholson, R. A., Buis, T., Radovanovic, H., & Fidler, B. J. (1999). Defensive responding on the MMPI-2 in family custody and access evaluations. *Psychological Assessment, 11(1),* 24.

Barrick, M. R., Mount, M. K., and Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9(1–2), 9-30.

Bass, C., & Halligan, P. W. (2007). Illness related deception: social or psychiatric problem? *Journal of the Royal Society of Medicine*, *100*(2), 81–84.

Bathurst, K., Gottfried, A. W., & Gottfried, A. E. (1997). Normative data for the MMPI2 in child custody litigation. *Psychological Assessment, 7*, 419-423.

Baumrind, D. (1966). Effects of authoritative parental control on child behavior. *Child Development, 37,* 887.

Bensch, D., Maaß, U., Greiff, S., Horstmann, K. T., & Ziegler, M. (2019). The nature of faking: A homogeneous and predictable construct?. *Psychological Assessment, 31*(4), 532.

Bianchini, K. J., Mathias, C. W., & Greve, K. W. (2001). Symptom validity testing: A critical review. *The Clinical Neuropsychologist, 15*(1), 19-45.

Binder, L., & Pankratz, L. (1987). Neuropsychological evidence of a factitious memory complaint. *Journal of Clinical and Experimental Neuropsychology (9),* 167-171.

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., and Smith, M. A. (2006). A Meta-Analytic Investigation of Job Applicant Faking on Personality Measures. *International Journal of Selection and Assessment*, *14(4),* 317-335.

Bobbio, A., and Manganelli, A. M. (2011). Measuring social desirability responding. A short version of Paulhus' BIDR 6. *TPM-Testing, Psychometrics, Methodology in Applied Psychology, 18(2),* 117-135.

Boddy, C. R. (2010). Corporate Psychopaths and Productivity. *Management Services*, 26- 30. Spring.

Book, A.S., Holden, R.R., Starzyk, K.B., Wasylkiw, L., and Edwards, M.J. (2006). Psychopathic traits and experimentally induced deception in self-report assessment. *Personality and Individual Differences, Volume 4, Issue 4, September 2006*, 601-608.

Boone, K. B., Salazar, X., Lu, P., Warner-Chacon, K., &Razani, J. (2002). The Rey 15-item recognition trial: A technique to enhance sensitivity of the Rey 15-item memorization test. *Journal of clinical and Experimental Neuropsychology*, *24*(5), 561-573.

Bricklin, B., & Elliot, G. (1995). ACCESS: *A comprehensive custody evaluation standard system.* Furlong, PA: Village.

Butcher, J. N. (2010). Minnesota multiphasic personality inventory. *The Corsini Encyclopedia of Psychology*, 1-3.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., and Kaemmer, B. (2001). *MMPI-2 (Minnesota Multiphasic Personality Inventory-2). Manual for administration, scoring, and interpretation*. Revised Edition. Minneapolis, MN: University of Minnesota Press.

Camerini G. B., (2006). Aspetti legislativi e psichiatrico-forensi nei procedimenti riguardanti i minori, in VOLTERRA V. (a cura di), *Psichiatria forense, criminologia ed etica psichiatrica (Trattato Italiano di Psichiatria, TIP)*. Masson, Milano, pp. 710-767.

Campbell, J., Schermer, J. A., Villani, V. C., Nguyen, B., Vickers, L., & Vernon, P. A. (2009). A behavioral genetic study of the Dark Triad of personality and moral development. *Twin Research and Human Genetics, 12*, 132-136.

Charter, R. A., & Lopez, M. N. (2003). MMPI-2: Confidence intervals for random responding to the F, F Back, and VRIN scales. *Journal of clinical psychology, 59*(9), 985-990.

Christie, R., and Geis, F.L. (1970). *Studies in Machiavellianism*. Academic Press.

Choca, J. P., & Grossman, S. D. (2015). Evolution of the Millon Clinical Multiaxial Inventory. *Journal of personality assessment, 97*(6), 541-549.

Clemente, M., Padilla-Racero, D., & Espinosa, P. (2020). The dark triad and the detection of parental judicial manipulators. Development of a judicial manipulation scale. *International journal of environmental research and public health, 17(8),* 2843.

Clemente, M., & Espinosa, P. (2021). Revenge in Couple Relationships and Their Relation to the Dark Triad. *International Journal of Environmental Research and Public Health, 18(14),* 7653.

Conroy, M. A., &Kwartner, P. P. (2006). The definition of malingering. *Applied Psychology in Criminal Justice, 2*(3), 30-51.

Corral, S., & Calvete, E. (2000). Machiavellianism: Dimensionality of the Mach IV and its relation to self-monitoring in a Spanish sample. *The Spanish journal of psychology, 3*, 3-13.

Costa, P. T., and McCrae, R. R. (1985). *The NEO personality inventory manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., and McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders, 6(4),* 343- 359.

Craig, R. I. (2014). The Millon Clinical Multiaxial Inventory-III.

Crego, C., and Widiger, T. A. (2014). Psychopathy and the DSM. *Journal of Personality, Volume 83, Issue 6.*

Crighton, A. H., Marek, R. J., Dragon, W. R., & Ben-porath, Y. S. (2017). Utility of the MMPI-2-RF validity scales in detection of simulated underreporting: implications of incorporating a manipulation check. *Asessment 24(7),* 853-64.

De Beni, R., Carretti, B., Moè, A., and Pazzaglia, F. (2008). *Psicologia della personalità e delle differenze individuali*. Il Mulino, Bologna.

Edens, J. F., Hart, S. D., Johnson, D. W., Johnson, J. K., &Olver, M. E. (2000). Use of the Personality Assessment Inventory to assess psychopathy in offender populations. *Psychological assessment, 12*(2), 132.

El Naqa, I., & Murphy, M. J. (2015). What is machine learning?. In *machine learning in radiation oncology* (pp. 3-11). Springer, Cham.

Faust, D., Hart, K. J., Guilmette, T. J., &Arkes, H. R. (1988). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice*, *19*(5), 508.

Feldman, R. S., Forrest, J. A., &Happ, B. R. (2002). Self-presentation and verbal deception: Do self-presenters lie more?. *Basic and applied social psychology, 24*(2), 163-170.

Fernandes, M., & Randall, D. (1991). The social desirability response bias in ethics research. *Journal of Business Ethics, 10* (11), 805-807.

Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The Dark Triad of personality: A 10 year review. *Social and personality psychology compass, 7(3),* 199-216.

Gardner, A., (1998). Recommendation for dealing with parents who induce a parental alien-ation syndrome in their children. *Journal of Divorce & Remarriage, 28*, 3/3, pp. 1-23.

Geis, F. L., and Moon, T. H. (1981). Machiavellianism and deception. *Journal of Personality and Social Psychology, 41(4)*, 766-775.

Goffman, E. (1959). *The presentation of self in everyday life*. New York: Doubleday Anchor.

Goldberg, L. R. (1990). An alternative ‘‘description of personality”: The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*, 1216–1229.

Griffith, R. L., and Converse, P. D. (2012). The rules of evidence and the prevalence of applicant faking. In M. Ziegler, C. MacCann, R. D. Roberts, *New Perspectives on Faking in Personality Assessment.* Oxford University Press.

Gulotta, G. (2011). *Compendio di psicologia giuridico-forense, criminale e investigativa* (Vol. 53). Giuffrè Editore.

Haines, M. E., & Norris, M. P. (1995). Detecting the malingering of cognitive deficits: An update. *Neuropsychology review, 5*(2), 125-148.

Hare, R. D. (1985). Comparison of procedures for the assessment of psychopathy. *Journal of Consulting and Clinical Psychology, 53(1),* 7-16.

Hare, R. D. (1991). *Manual for the revised psychopathy-checklist*. Toronto: Multi-Health Systems.

Hare, R. D. (1996). Psychopathy: A clinical construct whose time has come. *Criminal Justice and Behavior, 23(1),* 25-54.

Heggestad, E. D. (2012). A conceptual representation of faking: putting the horse back in front of the cart. In M. Ziegler, C. MacCann, and R. D. Roberts, *New Perspectives on Faking in Personality Assessment*. Oxford University Press.

Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin, 30(2),* 161-172.

Iverson, G. L., &Barton, E. (1999). Interscorer reliability of the MMPI-2: Should TRIN and VRIN be computer scored?. *Journal of clinical psychology, 55*(1), 65-69.

Jonason, P. K., ti, N. P., Webster, G. D., & Schmitt, D. P. (2009). The Dark Triad: Facilitating a short-term mating strategy in men. *European Journal of Personality, 23*, 5- 18.

Jonason, P. K., & Tost, J. (2010). I just cannot control myself: The Dark Triad and selfcontrol. *Personality and Individual Differences, 49,* 611-61.

Jonason, P. K., and Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment, 22(2),* 420-432.

Jones, D. N., & Paulhus, D. L. (2010). Different provocations trigger aggression in narcissists and psychopaths. *Social Psychological and Personality Science, 1(1),* 12-18.

Jones, D. N., & Paulhus, D. L. (2010). *Mating Strategies among the Dark Triad: Retention, infidelity, and short- vs.long-tenn relationship focus*. Manuscript submitted for publication.

Jones, D. N., and Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A brief measure of dark personality traits. *Assessment, 21,* 2841.

Kaufman, S. B., Yaden, D. B., Hyde, E., and Tsukayama, E. (2019). The light vs. dark triad of personality: contrasting two very different profiles of human nature. *Front. Psychol. 10:467.*

Kernberg, O. (1975). *Borderline conditions and pathological narcissism*. New York: Jason Aronson.

Kiazad, K., Restubog, S. L. D., Zagenczyk, T. J., Kiewitz, C., and Tang, R. L. (2010). In pursuit of power: The role of authoritarian leadership in the relationship between supervisors' machiavellianism and subordinates' perceptions of abusive supervisory behavior. *Journal of Research in Personality*, *44(4),* 512-519.

Kiehl, K. A., and Hoffman, M. B. (2011). The criminal psychopath: history, neuroscience, treatment and economics. *Jurimetrics, 51*, 355-397.

Kohut, H. (1968). The psychoanalytic treatment of narcissistic personality disorders: Outline of a systematic approach. *The psychoanalytic study of the child, 23(1),* 86-113.

Kowalski R. M. (Ed) (2001), *Behaving badly: Aversive behaviors in interpersonal relationships*. American Psychological Association: Washington, DC.

Lampel, A. (1999). Use of the Millon Clinical Multiaxial Inventory–III in evaluation child custody litigants. *American Journal of Forensic Psychology, 17*, 10–31.

Láng, A. (2020). Machiavellianism scale (Mach-IV). *Encyclopedia of Personality and Individual Differences,* 2718-2720.

Langleben, D. D., Loughead, J. W., Bilker, W. B., Ruparel, K., Childress, A. R., Busch, S. I., & Gur, R. C. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Human brain mapping, 26*(4), 262-272.

Lenny, P., & Dear, G. E. (2009). Faking Good on the MCMI–III: Implications for child custody evaluations. *Journal of personality assessment, 91(6),* 553-559.

Lilienfeld, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. *Journal of Personality Assessment, 66*, 488–524.

Machiavelli, N. (1513). Il Principe. In V. De Caprio and S. Giovanardi (Eds.) (1994), *I testi della Letteratura Italiana*. Milano: Einaudi Scuola.

MacNeil, B. M., and Holden, R. R. (2006). Psychopathy and the detection of faking on self-report inventories of personality. *Personality and Individual Differences, 41(4),* 641-651.

Maples, J. L., Lamkin, J., & Miller, J. D. (2014). A test of two brief measures of the dark triad: The dirty dozen and short dark triad. *Psychological assessment, 26(1),* 326.

Mazza, C., Monaro, M., Orrù, G., Burla, F., Colasanti, M., Ferracuti, S., and Roma, P. (2019). Introducing Machine Learning to Detect Personality Faking-Good in a Male Sample: A New Model Based on Minnesota Multiphasic Personality Inventory-2 Restructured Form Scales and Reaction Times. *Frontiers in Psychiatry, vol. 10, 6 June 2019*.

Mazza, C., Monaro, M., Burla, F., Colasanti, M., Orrù, G., Ferracuti, S., and Roma, P. (2020). Use of mouse-tracking software to detect faking-good behavior on personality questionnaires: an explorative study. *Scientific Reports 10(1):4835.*

McCann, J. T., Flens, J. R., Campagna, V., Collman, P., Lazzaro, T., and Connor, E. (2001). The MCMI-III in child custody evaluations: A normative study. *Journal of Forensic Psychology Practice, 1(2),* 27-44.

McCrae, R. R., and Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51(6),* 882-888.

McHoskey, J. (1995). Narcissism and machiavellianism. *Psychological reports, 77(3),* 755-759.

McHoskey, J. W., Worzel, W., & Szyarto, C. (1998). Machiavellianism and psychopathy. *Journal of personality and social psychology, 74(1),* 192.

Meadow, R. (1982). Munchausensyndrome by proxy. *Archives of disease in childhood*, *57*(2), 92-98.

Mealey, L. (1995). The sociobiology of sociopathy: An integrated evolutionary model. *Behavioral and Brain Sciences, 18*, 523-599.

Miller, J. D., & Lynam, D. R. (2012). An examination of the Psychopathic Personality Inventory's nomological network: A meta-analytic review. *Personality Disorders: Theory, Research, and Treatment, 3,* 305–326.

Miller, J. D., Few, L. R., Seibert, L. A., Watts, A., Zeichner, A., & Lynam, D. R. (2012). An examination of the Dirty Dozen measure: A cautionary tale about the costs of brief measures. *Psychological Assessment, 24*, 1048-1053.

Millon, T., & Davis, R. D. (1997). The MCMI--III: present and future directions. *Journal of personality assessment, 68*(1), 69-85.

Mitchell, T. (1997). *Machine Learning*. Columbus: WCB/McGraw-Hill.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exeggeration. *Journal of clinical and experimental neuropsychology, 24*(8), 1094-1102.

O'Boyle, E. H., Jr., Forsyth, D. R., Banks, G. C., and McDaniel, M. A. (2012). A metaanalysis of the Dark Triad and work behavior: A social exchange perspective. *Journal of Applied*

Psychology, 97(3), 557-579.

Orrù, G., Monaro, M., Conversano, C., Gemignani, A. and Sartori, G. (2020): Machine Learning in Psychometrics and Psychological Research. *Front. Psychol. 10:2970.*

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46(3),* 598-609.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, and L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes, Vol. 1*, 17-59. San Diego, CA: Academic Press.

Paulhus, D. L., Bruce, M., and Trapnell, Paul (1995). Effects of Self-Presentation Strategies on Personality Profiles and their Structure. *Personality and Social Psychology Bulletin. 21,* 100-108.

Paulhus, D. L. (2002). Social desirable responding: The evolution of a construct. In H. I. Braun & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49 – 69). Mahwah, NJ: Erlbaum.

Paulhus, D. L., and Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism and psychopathy. *Journal of Research in Personality, 36(6),* 556-563.

Paulhus, D. L., Harms, P. D., Bruce, M. N., and Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology, 84(4),* 890-904.

Paulhus, D. L. (2014). Toward a taxonomy of dark personalities. *Current Directions in Psychological Science, 23(6)*, 421-426.

Paulhus, D. L., Neumann, C. S., & Hare, R. D. forthcoming. *Manual for the Self-Report Psychopathy (SRP)* Scale. Toronto: Multi-Health Systems.

Petrides, K. V., Pita, R., & Kokkinaki, F. (2007). The location of trait emotional intelligence in personality factor space. *British Journal of Psychology, 98*, 273–289.

Petrides, K. V., Vernon, P. A., Schermer, J. A., & Veselka, L. (2011). Trait emotional intelligence and the Dark Triad of personality. *Twin Research and Human Genetics, 14*, 35–41.

Raskin, R. N., and Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports, 45(2),* 590.

Raskin, R., & Terry, H. (1988). A principal-components analysis of the narcissistic personality inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology, 54(5),* 890–902.

Raschka, S., & Mirjalili, V. (2017). Python Machine Learning: Machine Learning and Deep Learning with Python. *Scikit-Learn, and TensorFlow. Second edition ed.*

Rawling, P. J. (1992). The simulation index: a reliability study. *Brain Injury, 6(4)*, 381- 383.

Rees, L. M., Tombaugh, T. N., Gansler, D. A., &Moczynski, N. P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological assessment*, *10*(1), 10.

Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique. *Archives de Psychologie (23)*, 286-340.

Rey, A. (1958). *L'examen clinique en psychologie*. Parigi: Presse Universitaires de France.

Rogers, R., Kropp, P. R., Bagby, R. M., & Dickens, S. E. (1992). Faking specific disorders: A study of the Structured Interview of Reported Symptoms (SIRS). *Journal of Clinical Psychology, 48*(5), 643-648.

Rogers, R., Harrell, E. H., &Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review, 13*(3), 255-274.

Rogers, R., Sewell, K. W., Martin, M. A., &Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment, 10*(2), 160-177.

Rogers, R., Jackson, R. L., Sewell, K. W., & Salekin, K. L. (2005). Detection strategies for malingering: A confirmatory factor analysis of the SIRS. *Criminal Justice and Behavior, 32*(5), 511-525.

Rogers, R., Gillard, N. D., Wooley, C. N., & Kelsey, K. R. (2013). Cross-validation of the PAI Negative Distortion Scale for feigned mental disorders: A research report. *Assessment, 20*(1), 36-42.

Roma, P., Ricci, F., Kotzalidis, G.D., Abbate, L., Lavadera, A.L., Versace, G., et al. (2014). MMPI-2 in child custody litigation: a comparison between genders. *Eur J Psychol Assess 30(2),*110-6.

Salgado, J. F. (2016). A Theoretical Model of Psychometric Effects of Faking on Assessment Procedures: Empirical findings and implications for personality at work. *International Journal of Selection and Assessment, Volume 24, Number 3, September 2016.*

Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S., and Castiello, U. (2008). How to accurately detect autobiographical events. *Psychological Science, 19(8),* 772-780.

Sartori, G., Orrù, G., and Zangrossi, A. (2016). Detection of malingering in personal injury and damage ascertainment. In S.D. Ferrara, R. Boscolo-Berto, and G. Viel (Eds.), *Personal injury and damage ascertainment under civil law.* Cham, Switzerland: Springer.

Schimmenti, A., Jonason, P. K., Passanisi, A., La Marca, L., Di Dio, N., and Gervasi, A. M. (2017). Exploring the Dark Side of Personality: Emotional Awareness, Empathy, and the Dark Triad Traits in an Italian Sample. *Current Psychology, March 2017*.

Schlenker, B. R. (2003). Self-presentation. In M. R. Leary and J. P. Tangney (Eds.), *Handbook of self and identity*, 492-518. The Guilford Press.

Slick, D. J., & Sherman, E. M. (2012). Differential diagnosis of malingering and related clinical presentations. *Pediatric forensic neuropsychology*, 113-135.

Smith, G. P., and Burger, G. K. (1997). Detection of malingering: validation of the Structured Inventory of Malingered Symptomatology (SIMS). *J. Am. Acad. Psychiatry Law 25,* 183-189.

Spain, S. M., Harms, P. D., and LeBreton, J. M. (2014). The dark side of personality at work. *Journal of Organizational Behavior 35(S1):*S41-S60.

Stracciari, A., Bianchi, A., and Sartori, G. (2010). *Neuropsicologia forense*. Il Mulino, Bologna.

Sweet, J., King, J., Malina, A., Bergman, M., & Simmons, A. (2002). Documenting the prominence of forensic neuropsychology at national meetings and in relevant professional journals from 1990 to 2000. *The Clicinal Neuropsychologist*, 481-494.

Tombaugh, T. N. (1996). *Test of memory malingering: TOMM*. Multy-Health Systems.

Van Impelen, A., Merckelbach, H., Jelicic, M., &Merten, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A systematic review and meta-analysis. *The Clinical Neuropsychologist, 28*(8), 1336-1365.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... &Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008).

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and implications for professional practice*. Wiley.

Vernon, P. A., Villani, V. C., Vickers, L. C., & Harris, J. A. (2008). A behavioral genetic investigation of the Dark Triad and the Big 5. *Personality and Individual Differences, 44*, 445–452.

Veselka, L., Schermer, J. A., & Vernon, P. A. (2012). The Dark Triad and an expanded framework of personality. *Personality and Individual Differences, 53(4),* 417-425.

Watson, P. J., & Morris, R. J. (1991). Narcissism, empathy and social desirability. *Personality and Individual Differences, 12*, 575–579.

Winkelspecht, C., Lewis, P., and Thomas, A. (2006). Potential Effects of Faking on the NEO-PI-R: Willingness and Ability to Fake Changes Who Gets Hired in Simulated Selection Decisions. *Journal of Business and Psychology*, *Vol. 21, No. 2 (Winter, 2006),* 243-259. Springer.

Yun-tao, Z., Ling, G., & Yong-cheng, W. (2005). An improved TF-IDF approach for text classification. *Journal of Zhejiang University-Science A, 6*(1), 49-55.

Zhang, W., Yoshida, T., & Tang, X. (2011). A comparative study of TF*IDF, LSI and multi-words for text classification. *Expert Systems with Applications, 38*(3), 2758-2765.

Ziegler, M., MacCann, C., & Roberts, R. (Eds.). (2011). *New perspectives on faking in personalityassessment*. Oxford University Press.

APPENDICE - A Questionario (Dirty Dozen – Jonason & Webster, 2010 nella sua versione italiana validata da Schimmenti et al., 2017)

P1. Tendo a non provare rimorso

N2. Tendo a volere che gli altri mi diano attenzioni

P3. Tendo ad essere freddo o insensibile

M4. Ho usato l'adulazione per ottenere ciò che volevo

P5. Tendo a non interessarmi della moralità delle mie azioni

M6. Tendo a manipolare gli altri per ottenere ciò che voglio

N7. Tendo a cercare prestigio o potere

N8. Tendo a volere che gli altri mi ammirino

M9. Tendo a sfruttare gli altri per i miei scopi

P10. Tendo ad essere cinico

M11. Ho usato l'inganno o la menzogna per ottenere ciò che volevo

N12. Tendo ad aspettarmi favori speciali dagli altri

Si specifica che *P, N, M* stanno rispettivamente per *Psicopatia, Narcisismo, Machiavellismo* e che l'ordine riportato degli *item* è quello a cui sono stati sottoposti i partecipanti all'esperimento.

APPENDICE – B Tabelle Test Dunn (età)

Tabelle dei risultati del test Dunn (*Post Hoc*) per quanto riguarda l'età in entrambe le condizioni (*honest* e *dishonest*).

Tabella B.1. *Test Post Hoc di Dunn condotto sull'item M11 nella condizione honest*

Comparison	Z	W_i	W_j	p	p_{bonf}	p_{holm}
18 - 29 - 30 - 39	2.018	269.165	232.435	0.022*	0.218	0.175
18 - 29 - 40 - 49	2.880	269.165	211.982	0.002**	0.020*	0.018*
18 - 29 - 50 - 59	3.941	269.165	189.075	< .001***	< .001***	< .001***
18 - 29 - 60 - 69	1.962	269.165	190.333	0.025*	0.249	0.175
30 - 39 - 40 - 49	0.833	232.435	211.982	0.202	1.000	0.810
30 - 39 - 50 - 59	1.739	232.435	189.075	0.041*	0.410	0.246
30 - 39 - 60 - 69	0.986	232.435	190.333	0.162	1.000	0.810
40 - 49 - 50 - 59	0.876	211.982	189.075	0.191	1.000	0.810
40 - 49 - 60 - 69	0.499	211.982	190.333	0.309	1.000	0.810
50 - 59 - 60 - 69	-0.029	189.075	190.333	0.489	1.000	0.810

* p < .05, ** p < .01, *** p < .001

Tabella B.2. *Test Post Hoc di Dunn condotto sull'item N2 nella condizione honest*

Paragone	Z	W_i	W_j	p	p_{bonf}	p_{holm}
18 - 29 - 30 - 39	1.934	262.680	227.529	0.027*	0.265	0.212
18 - 29 - 40 - 49	2.357	262.680	215.982	0.009**	0.092	0.092
18 - 29 - 50 - 59	2.254	262.680	216.953	0.012*	0,22r	0.109
18 - 29 - 60 - 69	0.553	262.680	240.500	0.290	1.000	1.000
30 - 39 - 40 - 49	0.471	227.529	215.982	0.319	1.000	1.000
30 - 39 - 50 - 59	0.425	227.529	216.953	0.335	1.000	1.000
30 - 39 - 60 - 69	-0.304	227.529	240.500	0.380	1.000	1.000
40 - 49 - 50 - 59	-0.037	215.982	216.953	0.485	1.000	1.000
40 - 49 - 60 - 69	-0.566	215.982	240.500	0.286	1.000	1.000
50 - 59 - 60 - 69	-0.541	216.953	240.500	0.294	1.000	1.000

Tabella B.3. *Test Post Hoc di Dunn condotto sull'item N8 nella condizione honest*

Paragone	Z	Wi	Wj ·	p	pbonf	pholm
18 - 29 - 30 - 39	1.955	265.271	230.116	0.025*	0.253	0.202
18 - 29 - 40 - 49	2.554	265.271	215.188	0.005**	0.053	0.048*
18 - 29 - 50 - 59	2.836	265.271	208.340	0.002**	0.023*	0.023*
18 - 29 - 60 - 69	1.595	265.271	201.958	0.055	0.553	0.387
30 - 39 - 40 - 49	0.616	230.116	215.188	0.269	1.000	1.000
30 - 39 - 50 - 59	0.884	230.116	208.340	0.188	1.000	1.000
30 - 39 - 60 - 69	0.668	230.116	201.958	0.252	1.000	1.000
40 - 49 - 50 - 59	0.265	215.188	208.340	0.395	1.000	1.000
40 - 49 - 60 - 69	0.308	215.188	201.958	0.379	1.000	1.000
50 - 59 - 60 - 69	0.148	208.340	201.958	0.441	1.000	1.000

* p < .05, ** p < .01

Tabella B.4. *Test Post Hoc di Dunn condotto sull'item N12 nella condizione honest*

Paragone	Z	Wi	Wj ·	p	pbonf	pholm
18 - 29 - 30 - 39	1.390	260.673	235.275	0.082	0.823	0.659
18 - 29 - 40 - 49	2.960	260.673	201.670	0.002**	0.015*	0.015*
18 - 29 - 50 - 59	1.492	260.673	230.236	0.068	0.679	0.611
18 - 29 - 60 - 69	0.147	260.673	254.750	0.442	1.000	1.000
30 - 39 - 40 - 49	1.364	235.275	201.670	0.086	0.864	0.659
30 - 39 - 50 - 59	0.201	235.275	230.236	0.420	1.000	1.000
30 - 39 - 60 - 69	-0.454	235.275	254.750	0.325	1.000	1.000
40 - 49 - 50 - 59	-1.088	201.670	230.236	0.138	1.000	0.692
40 - 49 - 60 - 69	-1.218	201.670	254.750	0.112	1.000	0.670
50 - 59 - 60 - 69	-0.560	230.236	254.750	0.288	1.000	1.000

* p < .05, ** p < .01

Tabella B.5. *Test Post Hoc di Dunn condotto sull'item P3 nella condizione dihonest*

Comparison	Z	W_i	W_j	p	p_{bonf}	p_{holm}
18 - 29 - 30 - 39	-1.070	237.312	252.225	0.142	1.000	0.755
18 - 29 - 40 - 49	-2.512	237.312	275.482	0.006**	0.060	0.060
18 - 29 - 50 - 59	-2.076	237.312	269.604	0.019*	0.189	0.170
18 - 29 - 60 - 69	0.276	237.312	228.833	0.391	1.000	0.769
30 - 39 - 40 - 49	-1.238	252.225	275.482	0.108	1.000	0.755
30 - 39 - 50 - 59	-0.911	252.225	269.604	0.181	1.000	0.755
30 - 39 - 60 - 69	0.716	252.225	228.833	0.237	1.000	0.755
40 - 49 - 50 - 59	0.294	275.482	269.604	0.385	1.000	0.769
40 - 49 - 60 - 69	1.404	275.482	228.833	0.080	0.802	0.642
50 - 59 - 60 - 69	1.221	269.604	228.833	0.111	1.000	0.755

* p < .05, ** p < .01

Tabella B.6. *Test Post Hoc di Dunn condotto sull'item P3 nella condizione honest*

Comparison	z	W_i	W_j	p	p_{bonf}	p_{holm}
18 - 29 - 30 - 39	2.217	270.682	230.819	0.013*	0.133	0.106
18 - 29 - 40 - 49	4.238	270.682	187.598	< .001***	< .001***	< .001***
18 - 29 - 50 - 59	3.544	270.682	199.566	< .001***	0.002**	0.002**
18 - 29 - 60 - 69	1.056	270.682	228.792	0.145	1.000	0.727
30 - 39 - 40 - 49	1.783	230.819	187.598	0.037*	0.373	0.261
30 - 39 - 50 - 59	1.270	230.819	199.566	0.102	1.000	0.613
30 - 39 - 60 - 69	0.048	230.819	228.792	0.481	1.000	0.746
40 - 49 - 50 - 59	-0.463	187.598	199.566	0.322	1.000	0.746
40 - 49 - 60 - 69	-0.961	187.598	228.792	0.168	1.000	0.727
50 - 59 - 60 - 69	-0.678	199.566	228.792	0.249	1.000	0.746

* p < .05, ** p < .01, *** p < .001

Tabella B.7. *Test Post Hoc di Dunn condotto sull'item P10 nella condizione honest*

Comparison	z	W_i	W_j	p	p_{bonf}	p_{holm}
18 - 29 - 30 - 39	2.821	278.733	226.565	0.002**	0.024*	0.019*
18 - 29 - 40 - 49	4.385	278.733	190.304	< .001***	< .001***	< .001***
18 - 29 - 50 - 59	5.661	278.733	161.896	< .001***	< .001***	< .001***
18 - 29 - 60 - 69	1.839	278.733	203.708	0.033*	0.330	0.198
30 - 39 - 40 - 49	1.454	226.565	190.304	0.073	0.729	0.365
30 - 39 - 50 - 59	2.554	226.565	161.896	0.005**	0.053	0.037*
30 - 39 - 60 - 69	0.527	226.565	203.708	0.299	1.000	0.598
40 - 49 - 50 - 59	1.069	190.304	161.896	0.142	1.000	0.570
40 - 49 - 60 - 69	-0.304	190.304	203.708	0.381	1.000	0.598
50 - 59 - 60 - 69	-0.943	161.896	203.708	0.173	1.000	0.570

* p < .05, ** p < .01, *** p < .001

APPENDICE – C Tabelle Test Dunn (anni di scolarizzazione)

Tabelle dei risultati del test Dunn (*Post Hoc*) per quanto riguarda gli anni di scolarizzazione, in entrambe le condizioni (*honest* e *dishonest*).

Tabella C.1.*Test Post Hoc di Dunn condotto sull'item P3 nella condizione honest*

Comparison	z	W_i	W_j	p	p_{bonf}	p_{holm}
13 (scuola superiore) - 16 (laurea triennale)	-2.515	233.103	272.256	0.006**	0.060	0.054
13 (scuola superiore) - 18 (laurea magistrale)	-1.192	233.103	252.383	0.117	1.000	0.410
13 (scuola superiore) - 8 (scuola media)	1.144	233.103	192.438	0.126	1.000	0.410
13 (scuola superiore) - da 19 in su (es. master, dottorato, ecc.)	1.476	233.103	197.103	0.070	0.699	0.350
16 (laurea triennale) - 18 (laurea magistrale)	1.267	272.256	252.383	0.103	1.000	0.410
16 (laurea triennale) - 8 (scuola media)	2.259	272.256	192.438	0.012*	0.120	0.095
16 (laurea triennale) - da 19 in su (es. master, dottorato, ecc.)	3.123	272.256	197.103	< .001***	0.009**	0.009**
18 (laurea magistrale) - 8 (scuola media)	1.684	252.383	192.438	0.046*	0.461	0.277
18 (laurea magistrale) - da 19 in su (es. master, dottorato, ecc.)	2.260	252.383	197.103	0.012*	0.119	0.095
8 (scuola media) - da 19 in su (es. master, dottorato, ecc.)	-0.117	192.438	197.103	0.454	1.000	0.454

* p < .05, ** p < .01, *** p < .001

Tabella C.2. *Test Post Hoc di Dunn condotto sull'item N8 nella condizione honest*

Comparison	z	W_i	W_j	p	p_{bonf}	p_{holm}
13 (scuola superiore) - 16 (laurea triennale)	-2.137	227.213	260.497	0.016*	0.163	0.081
13 (scuola superiore) - 18 (laurea magistrale)	-2.431	227.213	266.529	0.008**	0.075	0.056
13 (scuola superiore) - 8 (scuola media)	2.448	227.213	140.125	0.007**	0.072	0.056
13 (scuola superiore) - da 19 in su (es. master, dottorato, ecc.)	-0.459	227.213	238.410	0.323	1.000	0.646
16 (laurea triennale) - 18 (laurea magistrale)	-0.384	260.497	266.529	0.350	1.000	0.646
16 (laurea triennale) - 8 (scuola media)	3.405	260.497	140.125	< .001***	0.003**	0.003**
16 (laurea triennale) - da 19 in su (es. master, dottorato, ecc.)	0.917	260.497	238.410	0.179	1.000	0.538
18 (laurea magistrale) - 8 (scuola media)	3.549	266.529	140.125	< .001***	0.002**	0.002**
18 (laurea magistrale) - da 19 in su (es. master, dottorato, ecc.)	1.149	266.529	238.410	0.125	1.000	0.501
8 (scuola media) - da 19 in su (es. master, dottorato, ecc.)	-2.455	140.125	238.410	0.007**	0.070	0.056

* p < .05, ** p < .01, *** p < .001

Tabella C.3. *Test Post Hoc di Dunn condotto sull'item N7 nella condizione honest*

Comparison	z	W_i	W_j	p	p_{bonf}	p_{holm}
13 (scuola superiore) - 16 (laurea triennale)	0.029	256.248	255.787	0.488	1.000	1.000
13 (scuola superiore) - 18 (laurea magistrale)	-0.016	256.248	256.522	0.493	1.000	1.000
13 (scuola superiore) - 8 (scuola media)	2.486	256.248	165.531	0.006**	0.065	0.045*
13 (scuola superiore) - da 19 in su (es. master, dottorato, ecc.)	3.147	256.248	177.487	< .001***	0.008**	0.008**
16 (laurea triennale) - 18 (laurea magistrale)	-0.046	255.787	256.522	0.482	1.000	1.000
16 (laurea triennale) - 8 (scuola media)	2.488	255.787	165.531	0.006**	0.064	0.045*
16 (laurea triennale) - da 19 in su (es. master, dottorato, ecc.)	3.169	255.787	177.487	< .001***	0.008**	0.008**
18 (laurea magistrale) - 8 (scuola media)	2.490	256.522	165.531	0.006**	0.064	0.045*
18 (laurea magistrale) - da 19 in su (es. master, dottorato, ecc.)	3.148	256.522	177.487	< .001***	0.008**	0.008**
8 (scuola media) - da 19 in su (es. master, dottorato, ecc.)	-0.291	165.531	177.487	0.385	1.000	1.000

* p < .05, ** p < .01, *** p < .001

Tabella C.4. *Test Post Hoc di Dunn condotto sull'item N1 nella condizione honest*

Comparison	z	W_i	W_j	p	p_{bonf}	p_{holm}
13 (scuola superiore) - 16 (laurea triennale)	-0.210	247.812	251.116	0.417	1.000	0.594
13 (scuola superiore) - 18 (laurea magistrale)	-0.979	247.812	263.810	0.164	1.000	0.594
13 (scuola superiore) - 8 (scuola media)	3.684	247.812	115.406	< .001***	0.001**	< .001***
13 (scuola superiore) - da 19 in su (es. master, dottorato, ecc.)	1.043	247.812	222.115	0.149	1.000	0.594
16 (laurea triennale) - 18 (laurea magistrale)	-0.801	251.116	263.810	0.212	1.000	0.594
16 (laurea triennale) - 8 (scuola media)	3.799	251.116	115.406	< .001***	< .001***	< .001***
16 (laurea triennale) - da 19 in su (es. master, dottorato, ecc.)	1.192	251.116	222.115	0.117	1.000	0.583
18 (laurea magistrale) - 8 (scuola media)	4.123	263.810	115.406	< .001***	< .001***	< .001***
18 (laurea magistrale) - da 19 in su (es. master, dottorato, ecc.)	1.686	263.810	222.115	0.046*	0.459	0.275
8 (scuola media) - da 19 in su (es. master, dottorato, ecc.)	-2.638	115.406	222.115	0.004**	0.042*	0.029*

* p < .05, ** p < .01, *** p < .001

Tabella C.5.*Test Post Hoc di Dunn condotto sull'item M11 nella condizione honest*

Comparison	z	W_i	W_j	p	p_{bonf}	p_{holm}
13 (scuola superiore) - 16 (laurea triennale)	-0.676	254.199	264.856	0.250	1.000	0.612
13 (scuola superiore) - 18 (laurea magistrale)	0.828	254.199	240.650	0.204	1.000	0.612
13 (scuola superiore) - 8 (scuola media)	1.934	254.199	184.563	0.027*	0.265	0.186
13 (scuola superiore) - da 19 in su (es. master, dottorato, ecc.)	2.371	254.199	195.641	0.009**	0.089	0.080
16 (laurea triennale) - 18 (laurea magistrale)	1.524	264.856	240.650	0.064	0.638	0.300
16 (laurea triennale) - 8 (scuola media)	2.244	264.856	184.563	0.012*	0.124	0.099
16 (laurea triennale) - da 19 in su (es. master, dottorato, ecc.)	2.840	264.856	195.641	0.002**	0.023*	0.023*
18 (laurea magistrale) - 8 (scuola media)	1.555	240.650	184.563	0.060	0.599	0.300
18 (laurea magistrale) - da 19 in su (es. master, dottorato, ecc.)	1.817	240.650	195.641	0.035*	0.346	0.208
8 (scuola media) - da 19 in su (es. master, dottorato, ecc.)	-0.273	184.563	195.641	0.392	1.000	0.612

* p < .05, ** p < .01

Tabella C.6. *Test Post Hoc di Dunn condotto sull'item M9 nella condizione dishonest*

Comparison	z	W_i	W_j	p	p_{bonf}	p_{holm}
13 (scuola superiore) - 16 (laurea triennale)	0.788	263.390	253.912	0.215	1.000	1.000
13 (scuola superiore) - 18 (laurea magistrale)	2.914	263.390	226.982	0.002**	0.018*	0.018*
13 (scuola superiore) - 8 (scuola media)	1.069	263.390	234.031	0.143	1.000	0.998
13 (scuola superiore) - da 19 in su (es. master, dottorato, ecc.)	1.505	263.390	235.026	0.066	0.661	0.529
16 (laurea triennale) - 18 (laurea magistrale)	2.222	253.912	226.982	0.013*	0.132	0.118
16 (laurea triennale) - 8 (scuola media)	0.728	253.912	234.031	0.233	1.000	1.000
16 (laurea triennale) - da 19 in su (es. master, dottorato, ecc.)	1.016	253.912	235.026	0.155	1.000	0.998
18 (laurea magistrale) - 8 (scuola media)	-0.256	226.982	234.031	0.399	1.000	1.000
18 (laurea magistrale) - da 19 in su (es. master, dottorato, ecc.)	-0.426	226.982	235.026	0.335	1.000	1.000
8 (scuola media) - da 19 in su (es. master, dottorato, ecc.)	-0.032	234.031	235.026	0.487	1.000	1.000

* p < .05, ** p < .01

Tabella C.7. *Test Post Hoc di Dunn condotto sull'item P3 nella condizione dishonest*

Comparison	z	W_i	W_j	p	p_{bonf}	p_{holm}
13 (scuola superiore) - 16 (laurea triennale)	2.218	267.337	240.575	0.013*	0.133	0.087
13 (scuola superiore) - 18 (laurea magistrale)	2.484	267.337	236.212	0.007**	0.065	0.052
13 (scuola superiore) - 8 (scuola media)	-1.119	267.337	298.188	0.131	1.000	0.402
13 (scuola superiore) - da 19 in su (es. master, dottorato, ecc.)	2.677	267.337	216.731	0.004**	0.037*	0.037*
16 (laurea triennale) - 18 (laurea magistrale)	0.359	240.575	236.212	0.360	1.000	0.402
16 (laurea triennale) - 8 (scuola media)	-2.103	240.575	298.188	0.018*	0.177	0.089
16 (laurea triennale) - da 19 in su (es. master, dottorato, ecc.)	1.278	240.575	216.731	0.101	1.000	0.402
18 (laurea magistrale) - 8 (scuola media)	-2.245	236.212	298.188	0.012*	0.124	0.087
18 (laurea magistrale) - da 19 in su (es. master, dottorato, ecc.)	1.027	236.212	216.731	0.152	1.000	0.402
8 (scuola media) - da 19 in su (es. master, dottorato, ecc.)	2.626	298.188	216.731	0.004**	0.043*	0.039*

* p < .05, ** p < .01