



UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Psicologia Generale

Corso di Laurea Magistrale in Neuroscienze e Riabilitazione Neuropsicologica

Tesi di Laurea Magistrale

**Think or Try (ToT): presentazione e proprietà psicometriche
di un nuovo strumento di screening per la valutazione della
stima cognitiva**

**Think or Try (ToT): presentation and psychometric properties of a new
screening tool for cognitive estimation assessment**

Relatrice:

Prof.ssa Sara Mondini

Correlatrici esterne:

Dott.ssa Alice Riccardi

Dott.ssa Veronica Pucci

Laureando: Matteo Lupini

Matricola: 2021200

Anno Accademico 2021/2022

Indice

1. Introduzione	3
2. La capacità di stima cognitiva	5
2.1 Il rapporto con il funzionamento esecutivo	9
2.2 Studi condotti sulle popolazioni cliniche.....	15
2.3 Test neuropsicologici per la valutazione	21
3. Nuovo test di stima cognitiva: Think or Try (ToT)..	35
3.1 La ricerca: studio dell'effetto test-retest	38
3.2 Materiali e Metodi	39
3.2.1 Partecipanti	40
3.2.2 La procedura	41
3.2.3 Analisi dei dati e risultati	43
3.3 Discussione	53
3.3.1 Conclusioni	59
4. Bibliografia	63
5. Appendice: Think or Try (ToT)	71

1. Introduzione

In questo lavoro viene trattata la capacità di stima cognitiva, misurata per la prima volta tramite test da Shallice ed Evans (1978) che riportano lo studio di pazienti neurologici con lesioni ai lobi frontali capaci di compromettere le facoltà di ragionamento, di *problem solving*, di attenzione, di memoria semantica e di lavoro del soggetto (Wagner, MacPherson, Parente, e Trentini, 2011). Shallice ed Evans sono stati i primi a costruire un test composto da diverse tipologie di domande utili a misurare la facoltà di stima cognitiva: questo primo esempio ha permesso successivamente ad altri autori di presentare nuove versioni del test, a partire dalla revisione di questa prima versione originale. In seguito, viene indagato il rapporto tra la capacità di stima cognitiva e il dominio delle funzioni esecutive, cercando di definire la possibilità di utilizzo di strumenti per la valutazione della stima cognitiva, come il Cognitive Estimation Test (CET) nelle sue varie versioni, per rilevare la presenza di deficit a livello esecutivo nelle diverse componenti del dominio. Poi si considera l'analisi degli studi condotti sulle popolazioni cliniche di pazienti affetti da disturbi sia frontali che di altra eziologia, ai quali sono proposti strumenti di valutazione della capacità di stima cognitiva per verificare l'utilità e la capacità di questo tipo di strumenti nel discriminare tra i diversi disturbi, evidenziando anche problematiche appartenenti ad altri domini cognitivi, come la memoria. Successivamente, viene presentata una rassegna della letteratura scientifica circa i principali test di valutazione della stima cognitiva e le loro versioni aggiornate nel tempo; i diversi autori hanno apportato modifiche, a partire dalla versione di Shallice ed Evans del (1978), alla concezione della struttura del test, alla composizione degli item, all'aggiornamento della taratura del campione, e alle correlazioni emerse tra le diverse variabili dal confronto tra popolazioni cliniche e controlli sani. L'analisi della letteratura scientifica di riferimento

consente di introdurre un nuovo strumento di rapida somministrazione per la valutazione della stima cognitiva, il “Think or Try” (ToT, in appendice): esso è composto da item privi di una risposta facilmente evocabile tramite nozioni apprese, come suggerisce il nome del test, le risposte necessitano di un ragionamento logico e di valutazioni astratte (“Think”) al termine delle quali, anche se non si riesce a giungere ad una soluzione precisa o ad un’ipotesi plausibile, è necessario fornire una risposta consapevole che una risposta esatta a queste domande non esista (“Try”). Il ToT viene presentato nelle sue caratteristiche principali andando poi, attraverso il test-retest, a verificare l’affidabilità test-retest dello strumento, la presenza di effetto pratica e le differenze tra le due modalità di somministrazione (autosomministrata e con esaminatore in videochiamata). Dopo un’illustrazione dei materiali e metodi utilizzati che considerano la procedura di somministrazione e i partecipanti coinvolti, il lavoro si concentrerà sulla parte di analisi dei dati e dei risultati ottenuti; essi verranno discussi alla luce delle osservazioni quantitative e qualitative in merito alle somministrazioni, traendo le opportune conclusioni ed evidenziando limiti e prospettive del lavoro.

2. La capacità di stima cognitiva

La stima cognitiva fa riferimento alla capacità e possibilità di generare e fornire una risposta, in genere una misura o comunque un valore quantitativo, ad una domanda di cui teoricamente non si dovrebbe conoscere la soluzione (Brand, Kalbe, Fujiwara, Huber, e Markowitsch, 2003). Il soggetto risponde pertanto a quesiti fornendo una stima di approssimazione che non dovrebbe poter attingere da conoscenze ben acquisite e consolidate (Wagner et al., 2011) ma esser stimolata da opportune strategie di *problem solving* che portano all'attivazione e al coinvolgimento, a livello anatomico, dei lobi frontali (Shallice ed Evans, 1978). Le domande di stima non hanno per definizione una risposta univoca, così come indicato anche nelle istruzioni di somministrazione dello strumento presentato ToT, ma degli intervalli di riferimento che comprendono la media di tutte le risposte plausibili in relazione al riscontro con la realtà. L'individuo che risponde deve fornire un giudizio risultato di un ragionamento che richiede capacità di astrazione, accesso alla memoria sia di lavoro che semantica, attenzione e pianificazione per vagliare le diverse ipotesi e strategie disponibili (Wagner et al., 2011) al fine di giungere ad una risposta alla specifica domanda posta. L'individuo comprende la domanda, opera un ragionamento che possa produrre un'opportuna risposta, vaglia le possibili ipotesi selezionando quelle ritenute più corrette e plausibili, ed una volta che la risposta viene giudicata non bizzarra (fuori luogo), viene espressa (Shallice ed Evans, 1978). Le domande devono essere inusuali e con un livello di difficoltà che deve essere adeguato: non devono essere troppo difficili, tali da richiedere particolari e complesse conoscenze che possono portare il partecipante a rispondere tirando ad indovinare e senza ragionare (Shallice ed Evans, 1978); ma nemmeno essere talmente facili da rendere la risposta del soggetto immediata date conoscenze consolidate e facilmente accessibili dalle

nozioni presenti in memoria. Esempi di quesiti che vanno a stimolare l'attivazione dei diversi processi cognitivi di stima possono essere riscontrati in alcuni test neuropsicologici nei rispettivi item, in cui viene richiesto, per esempio, di fare una stima di quanto possa pesare un determinato oggetto ("Quanto pesano delle scarpe col tacco?" – CET, Della Sala, MacPherson, Phillips, Sacco, e Spinnler, 2003), misurare la lunghezza o la distanza tra due punti ("Quanto è lungo un cucchiaino?" – BCET, Bullard et al., 2004) o ancora misurare il tempo per svolgere un'attività ("Quanto tempo ci mette... A farsi la doccia?" – Stime dei Tempi e dei Pesi, STEP, Nichelli et al., 2002). Le considerazioni di Shallice ed Evans (1978) sulla sede anatomica dei lobi frontali dei processi di stima riprendono l'idea di Luria (1966) che riporta come essi avrebbero un fondamentale ruolo nella capacità di pianificazione dell'individuo per i programmi cognitivi ad alto livello; così importante da controllare anche i programmi cognitivi a più basso livello, appartenenti anatomicamente alle regioni posteriori. Il CET nella sua prima versione proposta da Shallice ed Evans (1978), rappresenta il primo esempio di strumento testistico per la valutazione della stima cognitiva; si propone l'obiettivo di stimolare questi sistemi cognitivi ad alto livello in riferimento alle regioni frontali, a differenza di test per misurare l'intelligenza che sfruttano livelli di difficoltà crescenti che richiedono conoscenze dei sistemi cognitivi a basso livello, oppure abilità cognitive di routine come l'esecuzione di calcoli matematici acquisiti (Shallice ed Evans, 1978). L'analisi del problema e la sua risoluzione passano anche da una serie di valutazioni che si basano su riferimenti di cui si conoscono molte informazioni sensoriali oppure oggetti di cui si hanno ben presente misure come la lunghezza o il peso; essi vengono sfruttati sia a livello fisico che di rappresentazione, a partire dalle esperienze dirette pregresse. Un esempio in tal senso, ricavato dalle osservazioni durante le somministrazioni del ToT con la presenza

dell'esaminatore (test presentato in seguito, presente in appendice), può essere quello dell'utilizzo della misura della propria mano, risultata spesso utile come riferimento ed unità di misura nel verificare la *distanza tra il sedile di guida e il volante di un'automobile*, in riferimento all'item "*Quanto dista il sedile dal volante di un'automobile?*". Proponendo questo item, è spesso capitato di osservare come i partecipanti coinvolti tendano ad allungare le proprie braccia di fronte al petto, simulando la posizione assunta al volante ed immaginandosi con le mani poste sul volante nel tentativo di replicare a livello fisico e propriocettivo la situazione di guida. Una volta assunta la posizione, i partecipanti misurano le proprie braccia tese utilizzando la distanza tra pollice e mignolo, unità di misura ricavata dall'esperienza, a favorire la risposta di stima da fornire circa la distanza tra i due punti.

Un esempio di modello teorico volto a spiegare meglio la capacità di stima cognitiva può essere quello proposto da Brand, Kalbe e Kessler (2002a) supportato dai risultati ottenuti dagli studi su pazienti con morbo di Alzheimer e sindrome di Korsakoff (Brand et al., 2003, presentati in seguito). Esso definisce le componenti cognitive che sarebbero interessate nello svolgimento dei compiti di stima: la prima ad essere attivata dall'input della domanda ricevuta è la memoria di lavoro che ha la funzione di elaborare le informazioni relative al compito richiesto. Successivamente si attivano la memoria dichiarativa a lungo termine, per raccogliere le informazioni potenzialmente utili a trovare una risposta considerando le rappresentazioni dell'oggetto da stimare, le cifre da utilizzare, le unità di misura e le strategie da poter applicare; e la componente del "*central processing control*" che serve a selezionare le informazioni raccolte nell'ottica di risolvere lo specifico quesito posto. A questo punto entra nuovamente in gioco la memoria di lavoro che formula la possibile soluzione applicando le strategie

cognitive basate sulle informazioni raccolte; una volta ottenuta la soluzione essa viene valutata in termini di accuratezza e plausibilità secondo un “*feedback loop*” (Brand et al., 2002a) reiterato tra memoria di lavoro e “*central processing control*”. La soluzione finale viene definita dal “*central processing control*” nel momento in cui non vengono rilevati errori, in modo da definire così la miglior risposta a disposizione; ultimo *step* sarà la produzione della risposta in output passando per un’ultima volta attraverso la memoria di lavoro.

Difficoltà nella capacità di stima cognitiva comportano ripercussioni importanti nella qualità della vita quotidiana del paziente: per esempio, saper stimare intervalli di tempo, sia che essi siano determinati dall’azione diretta del soggetto o meno, diventa un requisito fondamentale nel pianificare la propria settimana; o ancora saper stimare in senso astratto la distanza tra oggetti può essere un requisito fondamentale per compiti all’ordine di tutti i giorni, come può essere mettersi alla guida che richiede proprio abilità di *problem solving*, attenzione, memoria, pianificazione e processamento di informazioni (Freund, Gravenstein, Ferris, Burke, e Shaheen, 2005).

2.1 Il rapporto con il funzionamento esecutivo

Le capacità esecutive sono di estrema importanza per il funzionamento cognitivo dell'individuo e per tutte le attività che vengono svolte quotidianamente: un costrutto neuropsicologico multidimensionale composto da un ventaglio di abilità cognitive proprie del soggetto che comprendono la volontà, la pianificazione, l'iniziativa a cominciare un'attività, il monitoraggio del corretto svolgimento del compito mentre viene eseguito, la generazione di ipotesi interpretative, la correzione di eventuali errori, la capacità di saper anticipare esiti e conseguenze del proprio comportamento, l'analisi del compito, la ricerca di strategie alternative in funzione dei riferimenti sociali e ambientali, la flessibilità cognitiva, il riuscire a resistere alle interferenze interne e/o esterne ed aspetti della memoria di lavoro e del ragionamento logico-deduttivo (Borkowsky e Burke, 1996; Anderson, Northam, Hendy, e Wrenall, 2001; Hobson e Leeds, 2001; Stuss e Alexander, 2000; Stuss et al., 2002; Jurado e Rosselli, 2007). Le capacità appartenenti a questo dominio cognitivo influenzano anche il corretto funzionamento di altri domini: deficit a questa componente che funge da "esecutivo centrale" (Baddeley e Hitch, 1974), possono, per esempio, compromettere i diversi *step* o la sequenzialità dei processi cognitivi condizionando aspetti dell'attenzione o della memoria. Problemi alle funzioni esecutive possono compromettere in maniera evidente l'agire quotidiano dell'individuo, con difficoltà nel fissare degli obiettivi e studiare il modo di raggiungerli (Jurado e Rosselli, 2007). Le funzioni esecutive anatomicamente vengono spesso identificate a livello dei lobi frontali, tanto che i due termini vengono talvolta utilizzati in maniera interscambiabile (Stuss et al., 2002) senza considerare il più esteso coinvolgimento dell'encefalo nelle diverse componenti del costrutto delle funzioni esecutive ed il contributo di altri domini cognitivi

(Zelazo e Müller, 2002). La stima cognitiva farebbe riferimento anche ad aspetti legati del corretto funzionamento dei lobi frontali, in riferimento a studi di pazienti con lesioni focali considerate, però, in maniera specifica e riduttiva: nel 1978 Shallice ed Evans propongono degli studi sperimentali a partire da osservazioni cliniche condotte su pazienti affetti da lesioni frontali che riportavano risposte eccessivamente esagerate oppure particolarmente limitate alle domande di stima proposte, con un'elevata produzione di risposte bizzarre prive di riferimenti logici rispetto a prestazioni di pazienti con lesioni posteriori e soggetti di controllo, dove ciò non si verificava. La capacità di stima cognitiva stimola una serie di processi appartenenti al dominio esecutivo, tra cui pianificazione e ragionamento astratto, oltre quelli elicitati da altri domini come memoria, sia di lavoro che semantica, ed attenzione (Wagner et al., 2011). In generale, pazienti con lesioni frontali possono riportare prestazioni deficitarie nella somministrazione di alcuni strumenti e prove cliniche utili alla misura del funzionamento esecutivo (Wagner et al., 2011): tra queste ritroviamo la Torre di Londra (Tower of London - ToL, Shallice, 1982), il Wisconsin Card Sorting Test (WCST, Anderson, Damasio, Jones, e Tranel, 1991) e il Stroop Color Word Test (SCWT, Stroop, 1935) che misurano abilità esecutive come l'inibizione, la pianificazione, il *set-shifting*, la *rule detection* e la formazione di concetti (Jurado e Rosselli, 2007). Zelazo, Carter, Reznick, e Frye (1997), a partire dalle considerazioni di Luria (1973), identificano le capacità esecutive come vere e proprie funzioni, e non solo come strutture e meccanismi cognitivi: ciò comporta la definizione di un costrutto determinato dal comportamento dell'individuo e da ciò che viene portato a termine a livello dell'abilità di *problem solving*, il tutto organizzato secondo una struttura gerarchica con varie sottocomponenti finalizzate alla risoluzione del problema posto. Questa visione delle funzioni esecutive può essere paragonata ai processi cognitivi coinvolti

anche nei test di stima: a partire dalla rappresentazione del problema, dalla sua pianificazione, dall'esecuzione guidata secondo intenzione e regole d'uso e dalla valutazione finale di quanto prodotto, con rilevazione e correzione degli errori (Zelazo et al. 1997). Questo schema dei diversi *step* cognitivi compiuti si ritrova anche durante lo svolgimento di test per il funzionamento esecutivo, come il già citato WCST (Grant e Berg, 1948): in questo strumento viene richiesta una rappresentazione del problema, una selezione di un piano cognitivo che possa portare alla soluzione, un mantenimento in elaborazione del piano finché non si giunge ad un pensiero o ad un'azione tale da portare l'individuo ad agire il comportamento richiesto, ed una successiva valutazione dell'esito ottenuto correggendo eventualmente la risposta. Ciò porta a suggerire una relazione tra le funzioni esecutive e la capacità di stima che può essere approfondita attraverso l'analisi delle relazioni fra i punteggi dei test di stima e quelli di strumenti per la valutazione delle funzioni esecutive. Viene evidenziata una forte relazione tra la prestazione di stima cognitiva rappresentata dai punteggi ottenuti al CET (versione Axelrod e Millis, 1994) ed il funzionamento esecutivo valutato attraverso il test di fluenza verbale Controlled Oral Word Association Test (COWAT - detto anche FAS, Strauss, Sherman, e Spreen, 2006) e il Test della Figura Complessa di Rey (RCFT, Meyers e Meyers, 1995), secondo quanto riportato dagli studi di Spencer e Johnson-Greene (2009) condotti su un campione di pazienti neurologici. Si evidenziano relazioni tra il dominio delle funzioni esecutive e la capacità di stima cognitiva a livello delle variabili di pianificazione e fluenza verbale, misurate rispettivamente dal RCFT (Meyers e Meyers, 1995) e dal COWAT (Strauss et al., 2006). Uno studio recente di Margraf, Bachmann, Schwandner, Gottschalk, e Seidel (2009) mostra come l'utilizzo del CET risulti essere poco sensibile nel discriminare partecipanti sani e pazienti con problematiche a carico delle funzioni esecutive (nel

caso particolare considerando pazienti con ischemia cerebrale (SVID) rispetto all'utilizzo di altri strumenti valutativi come il già citato ToL (Shallice, 1982) o il Test dell'Orologio (CLOX, Royall, Cordes, e Polk, 1998). Secondo uno studio di D'Aniello, Scarpina, Albani, Castelnuovo, e Mauro (2015) il CET sarebbe associato alla componente di recupero dell'informazione a livello semantico delle funzioni esecutive: gli autori riportano nel loro lavoro la somministrazione ad un gruppo di pazienti affetti dal morbo di Parkinson del CET (versione Della Sala et al., 2003), del Mini Mental State Examination (MMSE, Folstein, Folstein, e McHugh, 1975) e del Test dell'Orologio (Mondini, Mapelli, Vestri, e Bisiacchi, 2003) per il funzionamento cognitivo generale, oltre che una serie di prove standardizzate per la valutazione delle funzioni esecutive tra cui la Frontal Assesment Battery (FAB, Dubois, Slachevsky, e Isella, 2000) per la rilevazione di sindromi disesecutive frontali, il Digit Span all'indietro (Monaco, Costa, Caltagirone, e Carlesimo, 2013) per misurare la memoria di lavoro verbale, il test di Weigl (Goldstein e Scheerer, 1941), il Test di Fluenza Verbale fonemica e semantica (Novelli et al., 1986) e il Test di Stroop (Caffarra, Vezzadini, Dieci, Zonato, e Venneri, 2002) per verificare la resistenza alle interferenze e la capacità di inibizione dei partecipanti. Non vengono evidenziate correlazioni tra i punteggi del CET (Della Sala et al., 2003) e quelli del MMSE (Folstein et al., 1975), mentre vengono riscontrate tra il Test dell'Orologio (Mondini et al., 2003), la FAB (Dubois et al., 2000), il Digit Span all'indietro (Monaco et al., 2013) e il Test di Fluenza Verbale semantica (Novelli et al., 1986); in particolare la relazione con quest'ultimo strumento suggerisce l'importanza della memoria di lavoro nel favorire l'accesso e il recupero dell'informazione semantica, la quale contribuisce in maniera importante al processo di stima cognitiva (Brand et al., 2003). La relazione con il CET non si è riscontrata, però, con la componente fonemica del Test di

Fluenza Verbale (Novelli et al., 1986): entrambi i test sono sensibili alla misura delle capacità esecutive, ma la componente fonemica, rispetto a quella semantica, sarebbe meno correlata al recupero dell'informazione. Viene rilevata un'assenza di relazione con il CET anche per il Test di Weigl (Goldstein e Scheerer, 1942) come ulteriore conferma in quanto la prova richiede abilità di ragionamento visuospatiale e intelligenza fluida. Un'ultima relazione riscontrata, con i punteggi del CET, è quella con la FAB (Dubois et al., 2000) che potrebbe suggerire come lo strumento di stima considerato possa rappresentare un indice di funzionamento cognitivo generale. Alla luce delle relazioni dei punteggi dei diversi strumenti evidenziate a livello sperimentale nello studio di D'Aniello et al. (2015), sarebbe opportuno definire lo strumento del CET come sensibile solo ad alcune componenti delle funzioni esecutive, ed in particolare a quelle coinvolte nel processo di stima cognitiva, come la capacità di memoria di lavoro verbale e quella di conoscenza semantica (D'Aniello et al., 2015).

Nel complesso emergono risultati contrastanti sul rapporto tra funzioni esecutive e la capacità di stima cognitiva: i diversi studi a riguardo dovrebbero includere nelle loro analisi e considerazioni anche il coinvolgimento delle diverse componenti delle funzioni esecutive, o di altri domini cognitivi, riguardo la facoltà di stima così da ampliare le prove a favore o sfavore della relazione tra i due tipi di test; questo chiarirebbe meglio i processi e le risorse utilizzate e aiuterebbe a verificare i risultati fino ad ora ottenuti sulla sensibilità di strumenti come il CET. Quest'ultimo, ad oggi risulta essere rappresentativo del dominio delle funzioni esecutive a seconda delle richieste del compito considerato (Gansler, Varvaris, Swenson, e Schretlen, 2014). Approfondimenti sulla sensibilità e sull'utilizzo di questi strumenti acquisiscono importanza nel momento in cui si considerano pazienti in cui i deficit possono occorrere sia a

livello delle funzioni esecutive in maniera generale che alla specifica capacità di stima. Indagini in tal senso sarebbero utili alla pratica clinica in modo tale da garantire un'opportuna diagnosi differenziale e non sovrapporre quadri clinici che andrebbero in realtà distinti.

2.2 Studi condotti sulle popolazioni cliniche

Analizzando le popolazioni cliniche neurologiche e non, si nota come le difficoltà nella capacità di stima sarebbero caratteristiche tipiche in primis dei pazienti con lesioni frontali, secondo quanto riportato dagli studi di Shallice ed Evans (1978) che sottolineano le prestazioni deficitarie di pazienti con patologie frontali alle domande di stima rispetto a partecipanti sani di controllo, pazienti con lesioni temporali o diencefaliche. Conferme in tal senso si hanno nel contributo di Della Sala, MacPherson, Philips, Sacco, e Spinnler (2004) in cui si evidenzia la correlazione tra prestazioni deficitarie al CET (Della Sala et al., 2003), sia nei punteggi attribuiti agli item che alla presenza di bizzarrie, per pazienti frontali confrontati ad individui sani. Vi è la possibilità che pazienti con problematiche frontali mostrino difficoltà nella prestazione di test che misurano la facoltà di stima, ma che non presentino punteggi fuori dalla norma a test che valutano le funzioni esecutive in generale. Questi strumenti di valutazione della capacità di stima possono allora essere utili ad una diagnosi più accurata nell'evidenziare problematiche specifiche che altrimenti non sarebbero rilevate, diminuendo il numero di falsi negativi: un caso specifico può essere quello della doppia dissociazione evidenziata a livello della sindrome di Korsakoff, secondo quanto riportato dagli studi di Taylor ed O'Carrol (1995). Si evidenzia come pazienti affetti da questa patologia possano mostrare difficoltà nell'eseguire test di stima, come il CET (utilizzata una versione ridotta a 10 domande, Shoqeirat, Mayes, MacDonald, Meudell, e Pickering, 1990), senza ottenere però un riscontro simile considerando pazienti con altri disturbi frontali, considerati nello studio: gli autori hanno raccolto i dati delle prestazioni di 370 pazienti con diverse patologie neurologiche di varia eziologia (sindrome Korsakoff, trauma cranico, tumore cerebrale, aneurisma, demenza, encefalite) o con disturbi psichiatrici (ansia e depressione) senza riscontrare a livello sperimentale una

sensibilità del CET (Shoqeirat et al., 1990) nel rilevare problematiche a livello dei lobi frontali, confrontando quanto raccolto da questi pazienti con le prestazioni di un gruppo di partecipanti sani di 150 individui. I pazienti affetti dalla sindrome di Korsakoff sono risultati essere gli unici ad ottenere delle prestazioni deficitarie nei punteggi di stima cognitiva, senza trovare una correlazione della sensibilità dello strumento ai più generali disturbi frontali. Questi risultati sono supportati anche da quanto evidenziato dalle ricerche di Leng e Parkin (1988) in cui pazienti affetti dalla sindrome di Korsakoff sono risultati essere deficitari nella prestazione di stima rispetto a pazienti con altre problematiche di tipo frontale, con uniche eccezioni di un partecipante con lesioni a livello orbitofrontale, causate da aneurisma all'arteria anteriore, e alcuni pazienti con encefaliti causate da Herpes Simplex i quali hanno ottenuto anch'essi prestazioni fuori dalla norma. Questo studio suggerisce come la regione orbitofrontale potrebbe essere la più indicata a rispecchiare una difficoltà nella capacità di stima, ma con poca validità dei risultati date le ridotte dimensioni del campione clinico, di soli 13 individui, e senza definire una precisa localizzazione anatomica delle lesioni (come accaduto anche per Shallice ed Evans, 1978). Shoqeirat et al. (1990) propongono una replica della ricerca di Leng e Parkin (1988) con lo scopo di ampliare il campione sperimentale in termini di numerosità e verificare la possibile conferma o smentita dei risultati. In questo caso, emerge come i partecipanti affetti dalla sindrome di Korsakoff vengono discriminati dai soggetti di controllo considerando i punteggi ottenuti sia al test di stima cognitiva (CET, Shoqeirat et al., 1990) che a quelli del funzionamento esecutivo (WCST, Grant e Berg, 1948; COWAT, Strauss et al., 2006; RCFT, Meyers e Meyers, 1995); essi non vengono, però, discriminati a livello di prestazione dalle altre due popolazioni cliniche considerate, pazienti con encefalite causata da Herpes Simplex e con aneurisma a carico

dell'arteria cerebrale anteriore, in contrasto con quanto riportato da Leng e Parkin (1988).

La sindrome di Korsakoff si manifesta con la presenza di un'atrofia dei corpi mammillari che a volte si estende anche a livello della corteccia frontale (Squire, Amaral e Press, 1990; Jernigan, Schafer, Butters e Cermak, 1991) di interesse alla capacità di stima: secondo Taylor ed O'Carrol (1995) a spiegare l'eccezione dei pazienti affetti da questa patologia nella prestazione di stima potrebbe essere l'incapacità di questi pazienti di inibire le risposte bizzarre data la presenza di un deficit nella fase di monitoraggio della correttezza della risposta. Ciò potrebbe essere relato alla presenza di confabulazioni, peculiare segno clinico del disturbo, nonostante in questi pazienti siano in genere preservate le componenti di memoria a lungo termine e semantica (Taylor e O'Carrol., 1995) utili in altre fasi del processo di stima, come nella capacità di attingere alle conoscenze immagazzinate. Leng e Parkin (1993) propongono, però, come si possa manifestare in pazienti con sindrome di Korsakoff anche un deficit a livello della memoria semantica che completerebbe il quadro clinico, capace di giustificare le prestazioni deficitarie di questi pazienti al compito di stima cognitiva. Il coinvolgimento della memoria semantica nella stima cognitiva sarebbe confermato dal contributo di Della Sala et al. (2004) che verificano, in un gruppo di 175 partecipanti sani, la presenza di una correlazione debole ma significativa tra i punteggi ottenuti al CET (Della Sala et., 2003) e nel General Knowledge of the World test (GKW, Mariani, Sacco, Spinnler, e Venneri, 2002) che verifica conoscenze generali, accidentali, basate sull'esperienza e scolastiche dell'individuo.

Una possibile soluzione ai deficit di memoria che condizionano le prestazioni dei pazienti può essere l'utilizzo del test di stima di Brand, Kalbe e Kessler *Test zum kognitiven Schätzen* (TKS,

2002b): in questo strumento per ogni quesito, 16 in totale, vengono presentate delle carte che raffigurano di volta in volta l'oggetto citato nella domanda di stima (una carta per domanda). In questo modo, rispetto all'utilizzo del CET, verrebbe diminuito il carico in memoria della rappresentazione dell'oggetto da stimare e di cui si devono valutare gli attributi per arrivare ad inferire una certa misura, a seconda dei casi, di peso, dimensione, quantità o tempo (quattro categorie dello strumento in cui sono suddivisi gli item). Questo test potrebbe rappresentare un tentativo di valutare nei pazienti frontali mnesici la componente esecutiva della capacità di stima in maniera isolata, senza che i deficit di memoria possano influenzare la prestazione cognitiva dei pazienti, soprattutto in riferimento alla sindrome di Korsakoff e al morbo di Alzheimer.

Considerando pazienti affetti dalla patologia di Alzheimer appena citata, si possono considerare gli studi di Goldstein e collaboratori (1996) che hanno considerato un gruppo di pazienti affetti da questo disturbo comparandoli ad un altro gruppo di partecipanti sani nel rispondere alla versione originale del CET a 15 item di Shallice ed Evans (1978), per poi andare a valutare le differenze di risposta per tutte le tipologie di domande, sia quantitative che qualitative. I pazienti hanno mostrato prestazioni molto inferiori rispetto all'altro gruppo di partecipanti sani, fornendo un numero più elevato di risposte considerate estreme e bizzarre. Questi risultati sono stati confermati anche da Della Sala et al. (2004) nel confronto tra pazienti affetti da Alzheimer e controlli sani, appaiati per età e scolarità. Sempre Goldstein et al. (1996) provano ad introdurre una facilitazione a livello sperimentale nel tentativo di verificare un eventuale cambiamento nella prestazione di questi pazienti, comparandoli sempre ad un gruppo di controllo: vengono definite quattro alternative di risposta per ciascuna delle 15 domande di stima proposte (sempre nella versione di Shallice ed

Evans, 1978) di cui solo una corrisponde a quella corretta (target) mentre le altre tre rappresentano dei distrattori (non target). Anche in questo caso si osservano prestazioni inferiori per i pazienti rispetto ai controlli nonostante la presenza della scelta multipla, sia nelle domande qualitative che in quelle quantitative: i partecipanti con morbo di Alzheimer tendono a scegliere più facilmente l'alternativa di risposta sbagliata (non target) tra le quattro disponibili. È stato condotto uno studio anche sulla possibilità di utilizzare strumenti di valutazione della stima cognitiva per diagnosticare precocemente uno stato di Mild Cognitive Impairment (MCI) che può eventualmente evolvere col tempo nella manifestazione del morbo di Alzheimer (Petersen, 2016): Levinoff et al. (2006) confrontano tra loro gruppi di anziani privi di patologie neurologiche, con morbo di Alzheimer e con MCI (rispettivamente di 40, 40 e 73 partecipanti) somministrando una versione ibrida del CET con 23 item tratti sia dalla versione di Shallice ed Evans (1978) che da quella di Axelrod e Millis (1994), con l'aggiunta di alcuni item definiti ex novo da parte degli autori. I risultati riportano una prestazione peggiore al compito di stima nei pazienti con Alzheimer rispetto a quelle degli altri due gruppi sperimentali; non si evidenzia una capacità del test di discriminare fra partecipanti sani e quelli affetti da MCI, ma solo tra pazienti con malattia di Alzheimer e anziani sani.

Da queste osservazioni sulle diverse popolazioni cliniche, nasce la possibile ipotesi alternativa che il CET possa essere utilizzato non solo come specificatamente sensibile a problematiche esecutive frontali ma anche come riferimento testistico per il funzionamento cognitivo generale, valutando risorse di altri domini cognitivi reclutate nei processi di stima cognitiva, come attenzione e memoria (Taylor ed O'Carrol 1995).

Si potrebbero utilizzare strumenti di neuroimmagine funzionale, come PET o SPET, per andare ad indagare le aree di attivazione

del processo di stima e non basare le proprie considerazioni solo sulla localizzazione anatomica delle lesioni e sull'osservazione del comportamento del paziente durante la valutazione (Taylor ed O'Carrol, 1995). Una tale impostazione clinica potrebbe evitare all'esaminatore di sovrapporre lesioni focali di tipo frontale a problematiche più generali che potrebbero coinvolgere anche altre regioni dell'encefalo, come sindromi da disconnessione cerebrale.

2.3 Test neuropsicologici per la valutazione

Il primo test in ordine cronologico che tratta la capacità di stima cognitiva è la versione originale del Cognitive Estimation Task (CET) di Shallice ed Evans del 1978, per misurare le capacità di ragionamento e automonitoraggio: essa fa riferimento agli studi condotti principalmente nel Novecento su pazienti con lesioni a carico dei lobi frontali, in particolare a livello della regione dorsolaterale, osservando in questi individui particolari difficoltà nella capacità di giudizio, oltre che all'adattarsi a situazioni non consuete. Il principale riferimento teorico, su cui si basano queste ricerche, è la teoria di Luria del 1966 sulle capacità di pianificazione in relazione al funzionamento dei lobi frontali, in riferimento alle capacità cognitive ad alto livello che controllano i programmi cognitivi a più basso livello, attribuiti alle regioni posteriori dell'encefalo. Il caso esemplare che viene presentato nel *paper* di Shallice ed Evans (1978) è quello del paziente "J.S.", viene riportato il suo inusuale deficit selettivo nel dare risposte a domande centrate sulla capacità di fornire delle stime, a seguito di un'importante lesione frontale destra causata da un'esplosione: al quesito "Quanto misura il palazzo più alto di Londra?" egli replica con una misura tra i 18 000 e i 20 000 piedi (circa tra i 5400 e 6100 metri). Questo è solo un esempio delle risposte particolarmente bizzarre fornite dal paziente associate ad una ferma convinzione sulla veridicità delle stesse, anche nel momento in cui veniva sottolineata dagli esaminatori l'assurdità delle misure fornite: egli o non replicava o, seppur riluttante, modificava di poco la misura data nonostante rimanesse fuori luogo rispetto al confronto con la realtà. Gli autori, dato il caso singolare del paziente "J.S.", affinano alcune domande per testare la capacità di selezione del piano cognitivo da attuare per arrivare alla risposta corretta: i 15 quesiti proposti nella versione di Shallice ed Evans (1978), scelti tra i 26 iniziali secondo la minor varianza delle risposte, fanno riferimento ad

una conoscenza comune in modo che non venga richiesto un esagerato impiego di risorse cognitive; pur stimolando la capacità di ragionamento logico, senza evocare risposte facilmente fruibili dall'esperienza o dalla memoria del soggetto. Esempi di domande sono "Quanto misura l'altezza di un bus a due piani?" oppure "Quanto è lunga mediamente una cravatta?", per rispondere servono delle considerazioni basate su attributi e riferimenti di cui si hanno conoscenza, a favorire una risposta e contribuire al controllo del valore espresso, eventualmente correggendosi. Shallice ed Evans (1978) decidono di selezionare domande che possano suscitare difficoltà in pazienti con problemi a carico dei lobi frontali escludendo potenziali problematiche legate al dominio dell'intelligenza: spesso prestazioni fuori dalla norma degli individui erano proprio imputate a problematiche di quoziente intellettuale. I risultati sperimentali condotti su 96 pazienti neurologici (Shallice ed Evans, 1978) confermano una prestazione peggiore per pazienti con lesioni frontali rispetto a quelli con problematiche nelle regioni posteriori (a confermare l'ipotesi di partenza della teoria di Luria, 1966); non sono state osservate differenze per pazienti con lesioni localizzate nell'emisfero destro o sinistro. Inoltre, viene sottolineata una sostanziale dissociazione a livello cognitivo tra funzioni frontali ed intelligenza: non si hanno correlazioni tra la prestazione deficitaria dei soggetti frontali al CET e i punteggi ottenuti tramite la somministrazione delle Matrici di Raven (Raven, 1938). Gli autori riportano anche come le variabili di bassa scolarità e di età avanzata possano favorire una prestazione deficitaria. I limiti di questa prima versione sono evidenti: la mancanza di valori normativi per la correzione dei punteggi grezzi, la numerosità ridotta del gruppo di controllo di solo 25 partecipanti, l'eccessiva varietà nella tipologia delle domande poste e l'esagerato attaccamento del contenuto dei quesiti al contesto specifico culturale di

riferimento (quello britannico) che limita la scelta dei pazienti a cui somministrare lo strumento.

Un passo in avanti viene compiuto dallo strumento di Stime dei Tempi e dei Pesi (STEP, Nichelli et al., 2002) che parte da premesse analoghe a quelle di Shallice ed Evans (1978) considerando l'importanza della corteccia prefrontale e il suo ruolo di esecutivo centrale, richiamando anche esplicitamente al lavoro dei due autori, proponendo però nuove domande a comporre gli item. Gli autori osservano due principali tipi di critiche alla versione del CET del 1978: la prima fa riferimento a come il test nella versione originale sia troppo dipendente dagli aspetti contestuali e culturali del tempo e luogo in cui è stato redatto, con domande come “Quanti anni ha la persona più anziana della Gran Bretagna in questo momento?”; la seconda riguarda la mancata distinzione delle domande di stima secondo diversi tipi di categorie e capacità. Ciò porta Nichelli et al. (2002) a distinguere 20 domande appartenenti alla categoria “peso” (esempio: “Quanto tempo ci mette... Ad allacciarsi le scarpe?”) e 20 domande a quella del “tempo” (esempio: “Quale è il peso... Di una sedia di un tavolo da pranzo?”), per un totale di 40 item. Vengono eliminate e distinte da queste nuove categorie domande troppo specifiche e di tipo qualitativo riportate nella versione di Shallice ed Evans (1978, esempio: “Quale è il pesce più grande del mondo?”) selezionando quelle quantitative e misurabili che permettono di definire dei punteggi normativi. I quesiti posti da Nichelli et al. (2002) fanno riferimento sia ad esperienze dirette che indirette vissute dal soggetto: le domande delle due macrocategorie di “peso” e “tempo” vengono suddivise a metà, in una con domande che riguardano l'azione attiva da parte del soggetto, come stimare il peso di un oggetto d'uso quotidiano o un intervallo di tempo definito dall'intervento della persona. Nell'altra metà si considerano azioni di tipo passivo, dovendo rispondere o sul

peso di oggetti usualmente solo osservati e mai sollevati (come il paraurti posteriore di un'utilitaria) oppure circa intervalli di tempo non direttamente determinati dall'agire della persona e mai misurati (come il tempo impiegato ad aspettare che il semaforo diventi verde). Per la prima volta viene inserito l'indice degli elementi di bizzarria: esso ha la funzione di rimarcare risposte particolarmente fuori dai range di riferimento sottolineando quelli che sono potenziali campanelli d'allarme di una condizione critica del soggetto, favorendo l'utilizzo del test nella pratica clinica, mantenendo sempre viva l'attenzione dell'esaminatore all'interpretazione qualitativa dei punteggi. La distinzione delle domande di "peso" e "tempo" permette anche di osservare le potenziali dissociazioni categoriali a carico della regione prefrontale a seguito di lesioni, come definito dal modello teorico gerarchico di Grafman (1989; Grafman e Hendler, 1991), sulle rappresentazioni sia sul piano semantico che episodico della pianificazione a livello delle funzioni cognitive: aspetti circa attenzione e memoria fanno riferimento alla più complessa categoria del tempo ed informazioni stato-cinestesiche a quella del peso. Esse riflettono l'idea del modello teorico che vi sia un'architettura categoriale legata alla pianificazione cognitiva che coinvolge la corteccia prefrontale. A sottolineare l'importanza di questa concezione categoriale nel processo cognitivo di stima, nello strumento di Nichelli e collaboratori (2002) sono stati identificati a livello statistico i valori normativi sia per i singoli subtest che per le differenze dei subtest nel confronto tra le due diverse categorie.

Nelle critiche poste alla prima versione del 1978 di Shallice ed Evans, si può citare anche la necessità di estendere il range attribuito ai valori di normalità delle risposte, spesso molto variabili nella popolazione sana di controllo (Axelrod e Millis, 1994) andando a ridurre l'individuazione di eventuali falsi positivi, per non definire anormale ciò che non lo è. La

soluzione a quest'ultima osservazione può essere quella di utilizzare il metodo dei percentili, utilizzato in varie versioni come in quella di Bullard et al. del 2004, di MacPherson et al. del 2014 o di Scarpina, D'Aniello, Mauro, Castelnuovo, e MacPherson, del 2015 (vedi in seguito). Gli elementi critici della prima versione del 1978 sono stati analizzati anche dalla revisione di Axelrod e Millis del 1994: tra le migliorie apportate dagli autori nel loro strumento si segnalano l'ampliamento del campione di controllo a 164 partecipanti sani, solo 25 nel campione preliminare di Shallice ed Evans (1978), considerando anche maggiori variazioni di scolarità seppur con una media elevata di circa 16 anni (aggiornamento del campione britannico ad opera anche di O'Carrol, Egan e Mackenzie, 1994). Anche in questo caso per favorire l'assegnazione dei punteggi si sono utilizzate solo domande di tipo quantitativo in modo da standardizzare e verificare quanto una prestazione possa deviare dai valori di normalità. Questa versione propone anche l'aggiunta delle unità di misura in ciascun item del test, nel tentativo di evitare che il soggetto possa essere limitato nella propria prestazione cognitiva dall'incapacità di rilevare quella opportuna da utilizzare o a problematiche di conversione tra diversi ordini di grandezza. I risultati del primo studio condotto da Axelrod e Millis (1994) sul campione dei soggetti sani hanno fatto emergere correlazioni solo parziali tra i punteggi ottenuti al test e la variabile della scolarità (soprattutto nella differenza tra individui con licenza di scuola superiore ed individui con più di 16 anni di scolarità) e assenti con la variabile dell'età. Il secondo studio condotto dagli autori ha considerato 30 pazienti ambulatoriali di controllo privi di lesioni cerebrali e difficoltà cognitive, selezionati grazie ad un ampio screening neuropsicologico, confrontati ad altri 30 pazienti con trauma cranico grave, classificato secondo la Glasgow Coma Scale (Teasdale e Jennett, 1974) con un punteggio inferiore a 9 nelle prime 24 ore post trauma e a distanza di minimo un mese e

massimo tre anni e quattro mesi dall'incidente. I risultati riportano una prestazione sensibilmente inferiore nei traumatizzati cranici rispetto ai pazienti di controllo, ciò giustificato sia da deficit di memoria che di risoluzione dei problemi.

Il riferimento principale con campione italiano del Cognitive Estimation Task (CET) è la versione di Della Sala e collaboratori del 2003: la prestazione al test rappresenta un indicatore affidabile del potenziale malfunzionamento a carico delle regioni frontali, nel momento in cui si rilevino prestazioni con un cospicuo numero di risposte che non rispettano i range di normalità. In questa versione viene riproposto l'utilizzo dell'indice degli elementi di bizzarria: viene assegnato il punteggio di 1 per segnalare la presenza di una risposta bizzarra, oppure di 0 in sua assenza; il punto viene assegnato a stime fuori luogo che superano o il limite inferiore o quello superiore di valori soglia particolarmente bassi, da un lato, o alti, dall'altro. Questi punteggi vengono attribuiti per ogni item andandoli poi a sommare tra di loro in modo da ottenere un punteggio totale degli elementi di bizzarria che va da un minimo di 0 ad un massimo di 21, esattamente il numero totale delle domande dello strumento. Questa scala fornisce un'indicazione qualitativa delle risposte bizzarre fornite che va ad aggiungersi a quella quantitativa dei punteggi attribuiti agli item. Ciascuna domanda ha tre intervalli numerici di riferimento secondo cui vengono attribuiti i punteggi: il primo in cui la risposta viene considerata nella norma e si attribuisce un punteggio grezzo di 0 (tra 0-30 % di sopra o sottostima della risposta corretta), un secondo range più ampio a cui si assegna il punteggio grezzo di 1 (tra il 31-90 %), ed un terzo intervallo considera i valori superiori ed inferiori degli estremi stabiliti nel secondo range (oltre il 90% in più o in meno). Non è detto che una risposta che ottiene l'assegnazione del punto come elemento di bizzarria corrisponda per forza ad

un punteggio grezzo di 2, dipende da item ad item; sommando i diversi grezzi di ogni risposta si ottiene un valore totale che fa riferimento alla prestazione del soggetto, deficitaria tanto più il totale ottenuto sarà elevato. L'obiettivo dello strumento resta quello di misurare la capacità dell'individuo di pianificare e monitorare la strategia di ragionamento che porta alla produzione delle risposte: ciascuna domanda non suscita nell'esaminato una risposta immediata e precisa se non a seguito di riflessioni e pensieri astratti; si utilizzano domande tipo "Quanti cammelli ci sono in Italia?" (domanda a titolo di esempio che presta anche il titolo al *paper* di riferimento, Della Sala et al., 2003). Però, si riscontrano alcuni quesiti che vanno contro questi principi teorici, per esempio: "Quante settimane ci sono in un anno?" o ancora "Dopo che l'acqua bolle qual è il tempo necessario per cucinare un uovo sodo?"; sono item che richiedono un maggior accesso a conoscenze acquisite nel tempo tramite l'esperienza, come nell'organizzare impegni e consultando il calendario per il numero delle settimane nel primo caso, e maturando esperienza in cucina per il tempo necessario alla preparazione di un uovo sodo nel secondo esempio. La possibilità di predisporre in memoria di determinate informazioni può portare il soggetto a rispondere correttamente in modo rapido e senza particolari sforzi cognitivi rispetto ad altre domande, *bypassando* l'impiego del processo di ragionamento astratto a cui ipoteticamente mira l'utilizzo dello strumento. Per quanto riguarda la somministrazione del test ai controlli sani i punteggi raccolti sembrano essere correlati più a variabili come il genere e la scolarità piuttosto che all'età, con prestazioni tendenzialmente migliori in maschi e soggetti con elevata scolarità (correlazioni assenti per l'età anche negli studi del 2004 sempre di Della Sala et al.). Questa versione del CET (Della Sala et al., 2003) resta ad oggi probabilmente una delle più utilizzate nella pratica clinica per misurare la capacità di stima e tra i più affidabili nella fase di valutazione

neuropsicologica per constatare il funzionamento delle funzioni esecutive, potenzialmente compromesso sia da patologie di tipo psichiatrico che di tipo neurologico. In genere questo test viene somministrato insieme ad altre batterie più complete composte da diverse singole prove: non basta il CET o un qualsiasi altro strumento di valutazione della capacità di stima come unica evidenza per la definizione di problematiche delle capacità esecutive: esso è sensibile solo ad alcune delle sottocomponenti del dominio cognitivo di appartenenza. È più corretto somministrare anche batterie come la FAB (Dubois et al., 2000) o la Behavioural Assessment of Dysexecutive Syndrome (BADS, Wilson et al., 1996): la prima per una valutazione generale delle funzioni esecutive e la seconda per la diagnosi della sindrome disesecutiva; entrambe si compongono di singole prove, tra cui alcune dedicate proprio alla misura della capacità di stima. Nello studio di Gillespie, Evans, Gardener, e Bowen del 2002 non si trovano, però, correlazioni tra i punteggi del CET e quelli ottenuti nella parte di valutazione della stima della BADS (Wilson et al., 1996).

Altro strumento da considerare è il Biber Cognitive Estimation Test (BCET) di Bullard e collaboratori pubblicato nel 2004: esso mantiene la suddivisione categoriale operata già nello STEP (Nichelli et al., 2002), aggiungendo alle categorie di tempo e peso quelle riguardanti distanza e quantità, con 5 item per categoria a comporre un totale di 20 quesiti. Nella scelta delle domande gli autori scartano quelle potenzialmente influenzate dall'alto o basso livello di scolarità, limitando i casi in cui il soggetto possa essere a conoscenza della risposta o avvantaggiato nel ragionamento, influenzando la prestazione cognitiva. Il gruppo di ricerca di Bullard et al. (2004) ha anche condotto uno specifico studio su pazienti con demenza, morbo di Alzheimer e di Parkinson, confrontando le loro prestazioni con quelle di individui privi di deficit neurologici: a tutti i

soggetti sperimentali è stato somministrato il MMSE (Folstein et al., 1975) per una valutazione globale del funzionamento cognitivo ed intellettuale, cercando di individuare possibili *outlier* indagando anche la storia medica dei partecipanti di controllo. I punteggi deficitari al BCET dei soggetti sperimentali mostrano difficoltà sia per i pazienti con patologia di Alzheimer che quelli affetti da Parkinson, a differenza delle prestazioni nella norma dei controlli sani. Risultati importanti si ottengono analizzando le correlazioni tra MMSE (Folstein et al., 1975), BCET e le categorie che compongono quest'ultimo: i punteggi dei due diversi test paiono correlati tra loro sia per i controlli sani che per i pazienti con demenza, mostrando una buona validità convergente. Inoltre, per i primi si osserva una correlazione tra il punteggio al MMSE (Folstein et al., 1975) e la specifica categoria della quantità, mentre per i pazienti con Parkinson e Alzheimer si hanno relazioni, oltre che ancora con la quantità, anche con le categorie di peso e di tempo. Questi risultati confermano possibili problematiche delle funzioni esecutive proprie dei pazienti che soffrono delle sopra citate forme di demenza (La Rue, 1992; Goldstein et al., 1996); si sono anche riscontrate influenze dei punteggi per la variabile del genere, a differenza di quanto osservato nel gruppo di controllo. Risultati significativi sull'utilizzo del BCET sono stati evidenziati anche in altri studi nel rilevare prestazioni deficitarie di popolazioni cliniche, come nel caso di pazienti psichiatrici schizofrenici (Jackson, Fein, Essock, e Mueser, 2001) o con forme di autismo nelle prime fasi dell'infanzia (Liss, Fein, Bullard, Robins, e Waterhouse, 2000). Secondo Jackson et al. (2001) il BCET rappresenta anche un possibile predittore dell'esito dell'*outcome* riabilitativo, oltre che utile per rilevare problematiche esecutive e frontali di pazienti con trauma cranico o con ADHD. (Bullard et al., 2004).

Per completare questa rassegna è importante citare anche una delle versioni più recenti del CET ad opera di Scarpina et al. pubblicato nel 2015: essa ha il merito di aver contribuito ad un aggiornamento della taratura che consideri un campione normativo italiano, composto da 227 volontari con età compresa tra i 19 e 91 anni e con un range di scolarità tra i 5 e i 24 anni. La versione è composta da 9 diversi item con la possibilità di due forme parallele (A e B) per eventuali somministrazioni ripetute nel tempo che possano anche aiutare a contrastare l'eventuale effetto pratica ed apprendimento nel tempo da parte dell'esaminato. Le correlazioni statistiche dei punteggi sono simili a quelle evidenziate per il CET nella versione di Della Sala et al. (2003): assenti per età e scolarità, secondo la variabile genere si hanno punteggi molto inferiori, quindi migliori prestazioni, per i maschi rispetto alle femmine sia nella versione A che in quella B, con una differenza nelle prestazioni tale da necessitare una correzione dei punteggi grezzi. Anche in questo caso gli item sono stati costruiti in modo da prescindere dagli aspetti culturali specifici, viene rimosso l'indice degli elementi bizzarria presente in alcune versioni precedenti, non vengono previsti limiti di tempo per la risposta o vincoli di unità di misura che rimangono a giudizio del soggetto, e viene data la possibilità di ritrattare la propria risposta qualora il partecipante non sia convinto della stessa.

La versione di Scarpina et al. (2015), cronologicamente più aggiornata, prende spunto dalla versione con taratura inglese di MacPherson et al. di un anno precedente pubblicata nel 2014: essa è la prima a proporre una doppia forma parallela del test di stima, ripresa dalla versione a taratura italiana del 2015, con 9 item sia per la forma A che per quella B. La versione finale dello strumento è stata ricavata tramite due esperimenti: il primo riguarda la somministrazione di una prima bozza del test composta da 58 diverse domande somministrata a 184 volontari

sani britannici; i quesiti vengono suddivisi nelle categorie di numero, peso, lunghezza, area e velocità dando come istruzioni ai soggetti di rispondere con la soluzione che paia a loro più corretta, pur dichiarando come una risposta certa ed esatta realmente non ci sia. Viene chiesto rispondere attraverso l'utilizzo di una cifra quantitativa, senza imporre particolari vincoli di unità di misura come nella versione di Scarpina et al. (2015); le risposte vengono solo successivamente corrette in maniera omogenea secondo un'unica unità di misura per ogni item, in modo da escludere punteggi distorti da errori di conversione. Gli item che vanno a comporre le due versioni finali del test sono 18 in quanto vengono scartati alcuni considerati non significativi, in termini di coerenza interna, oppure a seguito dell'esame della relazione tra il punteggio dei singoli item e il punteggio globale del test. Per il secondo esperimento il campione è composto da 24 pazienti con lesioni frontali focali, evidenziate attraverso strumenti di neuroimmagine come TC e MRI, e privi nella loro storia medica di forme di epilessia precoce pre-lesione (soggetti con manifestazioni post-lesione potevano essere inclusi nel campione), afasie o più in generale patologie neurologiche o psichiatriche, selezionando per lo più pazienti con tumori cerebrali di vario tipo ed eziologia (18 su 24). Le prestazioni di questi pazienti sono state confrontate con quelle di 48 partecipanti sani selezionati dal campione del primo studio, come gruppo di controllo; un totale di due partecipanti sani per ogni paziente, appaiati per età, genere e scolarità. Rispetto alla versione successiva di Scarpina et al. (2015), si evidenziano correlazioni statistiche discordanti: MacPherson et al. (2014) ottengono una forte correlazione negativa tra il punteggio ottenuto al test e le variabili di età e di scolarità, nel considerare i controlli sani; ciò significa che anziani ed individui altamente scolarizzati tendono ad ottenere punteggi inferiori che, secondo il metodo di *scoring* attribuito, evidenziano una migliore

prestazione cognitiva. Anche in questa versione si rilevano migliori prestazioni da parte di individui maschi, differenza secondo la variabile del genere che non viene ritrovata nel confronto fra pazienti frontali e controlli sani: la condizione patologica porta ad una riduzione drastica della prestazione a prescindere dal genere dell'individuo, in entrambe le forme parallele del test (solo nella forma B confrontando punteggi di femmine sane e non si rilevano differenze, comunque poco significative). Tra le variabili con cui si è trovata una correlazione negativa si può citare anche l'intelligenza: si apre un interessante dibattito su come effettivamente essa possa o meno influenzare la prestazione cognitiva di stima. Secondo Scarpina et al. (2015) l'età non inciderebbe sulla prestazione in quanto sarebbe associata alla componente dell'intelligenza cristallizzata che riguarda aspetti legati alle nozioni apprese nel tempo e all'esperienza che tendono a resistere maggiormente all'avanzare dell'età dell'individuo e all'invecchiamento biologico (Cattell, 1963). Questa in contrasto con la componente di intelligenza fluida associata per lo più alla capacità di adattarsi a nuove situazioni e problemi da risolvere, oltre che a sistemi quali la memoria di lavoro (distinzione dei due tipi di intelligenza secondo il modello bifattoriale di Cattell del 1963). La posizione di Scarpina et al. (2015) critica il lavoro di MacPherson et al. (2014) secondo cui sarebbe vista in senso negativo l'influenza della variabile dell'età nei confronti della capacità di stima cognitiva che non dovrebbe basarsi su conoscenze già apprese, ma sulla novità e sulla stimolazione di network inediti. Altro contributo a riguardo è quello di Gansler et al. (2014): essi sostengono la capacità della componente cristallizzata dell'intelligenza di predire la prestazione del soggetto, dati i risultati ottenuti dalla somministrazione di una versione revisionata del CET originale di Shallice ed Evans del 1978 (CET-R, Gansler et al., 2014) somministrato ad oltre duecento controlli sani. Viene considerata la variabile

dell'intelligenza come valore unico e le sue relazioni con le sottocomponenti di intelligenza cristallizzata, fluida e di *visual processing*, esaminando per ciascuna di esse la relazione con la prestazione al test. Questa analisi ha evidenziato anche una relazione con la componente fluida dell'intelligenza, seppur meno robusta rispetto a quella cristallizzata. Posizioni discordanti suggeriscono come il dibattito possa essere ulteriormente arricchito da altri contributi in futuro; allo stesso tempo la questione sottolinea l'importanza di andare a stimolare i diversi aspetti del processo cognitivo coinvolto nella capacità di stima in modo da favorire un minor accesso a ciò che il soggetto conosce già in maniera consolidata e a favorire la stimolazione delle diverse componenti dell'intelligenza, senza intenderla solo come variabile unica.

3. Nuovo test di stima cognitiva: Think or Try (ToT)

Il lavoro proposto considera il nuovo test per la valutazione della stima cognitiva Think or Try (ToT, in appendice) che è stato costruito a partire da una preliminare raccolta dati proposta ad oltre 1300 individui in due diverse forme A e B composte ciascuna da 10 domande sulla capacità di stima.

La forma finale del test, considerata in questo lavoro, è costituita da 8 item, dopo aver selezionato quelli più significativi a partire dai 20 inizialmente considerati, in ordine di somministrazione troviamo:

- *“Quanto è lungo uno spazzolino da denti?”*
- *“Qual è la velocità massima di una gazzella?”*
- *“Quale è la distanza tra una persona e il suo cane tenuto al guinzaglio?”*
- *“Quanti petali ha una margherita?”*
- *“Quanto pesa un dizionario?”*
- *“Quanta acqua contiene un bicchiere?”*
- *“Quanto è largo un marciapiede?”*
- *“Quanto dista il sedile dal volante di un’automobile?”*

Gli item della forma finale del test hanno in comune l’oggetto della domanda con riferimenti al quotidiano considerando, per esempio, lo spazzolino da denti, il bicchiere d’acqua, il volante dell’automobile ecc.; ciò a favorire il ragionamento del partecipante alla luce della sua esperienza. Unica eccezione è quella del quesito *“Quale è la velocità massima di una gazzella?”* che è stato mantenuto nel test finale vista la sua forte validità statistica, rilevata attraverso le somministrazioni preliminari del ToT stesso.

La raccolta preliminare ha anche permesso di raccogliere dei dati utili ad evidenziare degli intervalli normativi a cui poter confrontare le risposte fornite dagli individui nelle varie

somministrazioni: lo *scoring* si basa su valori di riferimento che permettono l'attribuzione dei punteggi grezzi, utilizzando un metodo simile a quello riportato nella versione di Della Sala et al. (2003). Viene assegnato il punteggio grezzo di 0 per le risposte considerate corrette che rientrano nel primo intervallo stabilito per ogni item, ovvero entro il 30% di differenza dal valore di riferimento stabilito. Viene assegnato il punteggio di 1 per le risposte comprese tra gli estremi del primo intervallo e nuovi valori a formare un ulteriore intervallo, si considera una differenza della risposta tra il 30-90% dal valore di riferimento. Un terzo range con l'attribuzione di un punteggio grezzo di 2 viene definito per i valori superiori o inferiori a quelli estremi del secondo intervallo, con una differenza che supera il 90% dal valore di riferimento. (range specifici per ogni item, con rispettiva unità di misura, riportati in appendice).

Nel ToT non viene proposto l'indice degli elementi di bizzarria: nonostante esso sia d'aiuto nella valutazione qualitativa della prestazione dell'individuo nel sottolineare la presenza di risposte bizzarre, viene messa in discussione la sua utilità clinica considerando come la prestazione dell'individuo al CET (Della Sala et al., 2003) possa risultare deficitaria secondo i punteggi attribuiti agli elementi di bizzarria ma non secondo i punteggi totali ottenuti al test. (Riccardi et al., in preparazione).

Il punteggio totale del test dato dalla somma dei punteggi grezzi ottenuti ad ogni item, può variare da un minimo di 0 fino ad un massimo di 16, dato il numero complessivo degli 8 item che compongono il test, ciascuno con un punteggio massimo assegnabile di 2.

È importante sottolineare che più alto risulta essere il punteggio totale ottenuto al ToT e peggiore sarà la prestazione cognitiva del partecipante, con punteggio totale di 0 ottenuto dall'individuo nella prestazione migliore.

Il numero ridotto degli item permette una rapida somministrazione di screening che aiuta ad indagare in un breve tempo, di circa dieci minuti, la capacità di stima cognitiva del soggetto: questa valutazione permette nel caso in cui si riscontrino difficoltà più o meno evidenti, sia a livello dei punteggi ottenuti che della prestazione qualitativa del soggetto, di andare poi ad approfondire la fase di valutazione proponendo altri strumenti di valutazione della stima cognitiva, come il CET di Della Sala et al. del 2003, che risultano essere più dispendiosi nella durata della somministrazione rispetto al ToT.

Nella formulazione delle domande, come in altre versioni recenti del CET, non viene suggerita al partecipante l'unità di misura da utilizzare, lasciando che sia l'individuo a scegliere quella opportuna da utilizzare secondo le sue valutazioni (come riportato esplicitamente nelle istruzioni del test, vedi appendice) evitando di influenzare la prestazione cognitiva. Non vengono definiti limiti di tempo per la risposta da parte del soggetto che deve essere comunque espressa e riportata anche sotto sollecitazione dell'esaminatore. Tra gli altri accorgimenti introdotti si può citare, a livello dei singoli item, l'aggiunta dell'aggettivo "*massima*" nell'item riguardo la *velocità della gazzella*, in modo da evitare che il partecipante possa confondersi con la velocità media oppure rimanere nell'ambiguità del solo termine velocità (come nel CET Della Sala et al., 2003 o in altre versioni precedenti).

3.1 La ricerca: studio dell'effetto test-retest

Il lavoro proposto fa riferimento alla prima parte della procedura di validazione tutt'ora in corso del test Think or Try volta a verificare le diverse proprietà psicometriche proponendo lo strumento a diversi campioni di partecipanti sani, secondo un progetto di ricerca che si articola nella verifica di test-retest, equivalenza, validità convergente e *inter-rater agreement* dello strumento.

Questo elaborato si concentra sulla prima fase di test-retest, l'unica in cui è stato possibile ottenere dati completi e analizzabili in tempi utili alla stesura dell'elaborato.

Ad un primo campione è stata proposto il ToT in due diversi momenti in autosomministrazione attraverso l'invio di un link; ad un secondo campione è stato proposto il test somministrato da parte di un esaminatore, sempre in due tempi diversi, attraverso una videochiamata.

In entrambi i sottogruppi sono stati considerati due momenti di somministrazione: la prima avviene ad un determinato tempo T0, e la seconda a distanza di sei settimane circa, definendo il tempo T1. Di seguito vengono illustrati il metodo ed i materiali utilizzati, definendo la descrizione del campione dei partecipanti, la procedura metodologica, ed infine l'analisi e la discussione di dati e risultati ottenuti.

3.2 Materiali e Metodi

È stato proposto il test Think or Try con l'obiettivo di verificare alcune proprietà psicometriche dello strumento, tra cui la riproducibilità dell'utilizzo dello stesso considerando due rilevazioni in tempi differenti (T0 e T1) con l'obiettivo di verificare l'affidabilità del test.

Quest'ultima viene definita come la possibilità di ottenere lo stesso risultato in entrambi i momenti di somministrazione senza che vi siano alterazioni, se non date da dei cambiamenti attesi (Krabbe, 2017). È stata quindi correlata la misura ottenuta con la somministrazione del ToT al tempo T0 (fase test) con quella che avviene al tempo T1 (fase retest) a sei settimane di distanza. L'intervallo di tempo considerato è stato scelto in quanto considerato sufficientemente breve per evitare di supporre che il partecipante manifesti un cambiamento significativo e sufficientemente lungo per limitare l'effetto pratica.

Attraverso le analisi condotte si va ad indagare anche la presenza o assenza dell'effetto pratica: ovvero la possibilità che si manifestino dei miglioramenti della prestazione cognitiva data da una valutazione ripetuta nel tempo, utilizzando sempre lo stesso materiale testistico in diverse somministrazioni (Duff, Callister, Dennett, e Tometich, 2012). Si va ad osservare se il test proposto nella stessa modalità a distanza di tempo possa portare ad un apprendimento delle risposte alle domande di stima, segnalato da un conseguente miglioramento dei punteggi ottenuti.

Altro aspetto rilevante indagato è il confronto tra la modalità autosomministrata e quella con l'esaminatore in videochiamata, osservando possibili influenze sulla prestazione data la presenza o meno dell'esaminatore.

3.2.1 Partecipanti

Il campione totale è composto da 50 partecipanti sani di nazionalità italiana suddiviso tra 26 individui maschi (52%) e 24 partecipanti femmine (48%), con un'età compresa tra i 22 e i 61 anni ($M=38$; $DS=13,9$) e una scolarità compresa tra gli 8 e i 19 anni ($M=15,3$; $DS=3,08$).

Andando a descrivere i due campioni di 25 individui secondo la modalità autosomministrata o in videochiamata con esaminatore si registrano: per la forma in autosomministrazione 14 maschi e 11 femmine (M 56%, F 44%), di un'età compresa tra i 23 e i 56 anni ($M=39,7$; $DS=12,9$), una scolarità dagli 8 ai 19 anni ($M=15,0$; $DS=3,83$). Per la modalità in videochiamata con esaminatore si registrano 12 maschi e 13 femmine (M 48%, F 52%), un'età dei partecipanti compresa tra i 22 e i 61 anni ($M=36,4$; $DS=15,0$) e una scolarità tra gli 11 e i 18 anni ($M=15,6$; $DS=2,12$).

I partecipanti considerati sono sani e non appartenenti a popolazioni cliniche particolari: l'accertamento dell'assenza di patologie rilevanti è avvenuto attraverso un'autodichiarazione in calce al consenso informato, con cui si è andato ad escludere la potenziale presenza nei campioni considerati di individui con disturbi neurologici (come Sclerosi Multipla, malattia di Alzheimer, malattia di Parkinson, Sclerosi Laterale Amiotrofica), psichiatrici (Schizofrenia, Psicosi, Depressione Maggiore) e/o dipendenze (Alcolismo e Tossicodipendenza) evitando di considerare partecipanti con precedenti diagnosi certificate da uno specialista. L'autodichiarazione è stata prevista affinché la prestazione cognitiva del partecipante non venisse influenzata da condizioni cliniche potenzialmente capaci di compromettere il funzionamento esecutivo dell'individuo e, di conseguenza, la sua prestazione.

Gli individui sono stati informati, attraverso il consenso informato, circa le modalità di somministrazione, i tempi della ricerca, il

trattamento e la conservazione dei dati sensibili secondo le norme vigenti, la possibilità di ritirarsi in qualunque momento, la garanzia del trattamento anonimo dei dati di tipo confidenziale con il codice identificativo come unico riferimento, le finalità statistiche di ricerca e non cliniche dello studio, e la possibile restituzione dei grezzi per rivolgersi poi ad uno specialista.

I dati sono stati considerati validi solo nel momento in cui il partecipante abbia preso parte ad entrambi i momenti di somministrazione (T0 e T1) ed entro i tempi stabiliti.

3.2.2 La procedura

Si è proceduto nel somministrare lo strumento a ciascuno dei due gruppi di partecipanti secondo la modalità assegnata a ciascun campione: per la fase di autosomministrazione tramite l'invio di un link *Google Form* contenente la raccolta preliminare dei dati, il consenso informato e la certificazione di assenza di patologie; questa modalità considera il primo gruppo di 25 partecipanti a cui viene indicato il codice identificativo da riportare prima della compilazione dei dati e delle risposte.

La fase di somministrazione con esaminatore in videochiamata fa riferimento alla somministrazione dello strumento da parte dell'esaminatore che avviene a distanza via *webcam* al secondo gruppo di 25 partecipanti, necessitando della presenza contemporanea di partecipante ed esaminatore. Quest'ultimo, prima di somministrare il ToT, raccoglie direttamente i dati preliminari riportati a voce dal partecipante. In questo caso il consenso firmato è stato firmato precedentemente alla data di somministrazione e senza dover richiedere al partecipante il codice identificativo, inserito direttamente dall'esaminatore.

In ciascuna delle somministrazioni sono state impartite le istruzioni del test, via link riportate in forma scritta e in forma orale dall'esaminatore in videochiamata, le quali precisano come le domande proposte non abbiano una risposta esatta, dovendo fare

una stima precisa della risposta reputata corretta riportando anche l'unità di misura, dove possibile.

In alcuni casi, durante le videochiamate, è stato necessario sottolineare, a seguito di difficoltà riportate dai partecipanti nel rispondere, di fare una media delle rappresentazioni disponibili in memoria dei vari oggetti riscontrati nella propria esperienza, e di fornire quindi una risposta che le raccolga nel considerare, ad esempio, i *dizionari* esistenti evitando di pensare a quello tascabile di dimensioni ridotte, ma nemmeno uno di grandezza smisurata ai livelli di un'enciclopedia.

Spesso i partecipanti in risposta alle domande, invece di stimare una cifra, fanno constatazioni del tipo “Dipende dal bicchiere” (nell'item “*Quanta acqua contiene un bicchiere?*”), “Ci sono molti tipi di marciapiede” (alla domanda “*Quanto è largo un marciapiede?*”), “Il guinzaglio è anche allungabile” (in risposta a “*Qual è la distanza tra una persona e il suo cane tenuto al guinzaglio?*”); in questi casi l'esaminatore ha sollecitato il soggetto a ragionare, senza eccedere nei dettagli, e a considerare un valore medio che raccolga le diverse possibilità riscontrate nel quotidiano. Questo è applicabile solo nella modalità di somministrazione con la presenza dell'esaminatore in videochiamata, ma non in quella proposta in autosomministrazione. Analogamente per le unità di misura che, soprattutto nelle modalità in autosomministrazione, vengono omesse: anche in questi casi le risposte in cui compariva solo il valore assoluto senza alcun riferimento all'unità di misura o all'ordine di grandezza (discorso a sé per l'item “*Quanti petali ha una margherita?*” in cui non si ha un'unità di misura) è stato attribuito il punteggio grezzo massimo di 2 all'item. In videochiamata con l'esaminatore presente è stato possibile verificare la risposta del partecipante anche in termini di unità di misura, chiedendo conferma circa la risposta del partecipante solo nei casi in cui gli ordini di grandezza risultassero non congruenti o confusi a seguito dell'indecisione del partecipante (soprattutto per

l'item della "*Quanta acqua contiene un bicchiere?*"); o valutando con il punteggio grezzo di 2 nel caso fossero omessi, sempre senza suggerire quella da utilizzare. Nella modalità autosomministrata sono state considerate valide le risposte composte da un giudizio qualitativo che riportassero, però, una cifra stimata come per esempio "Più o meno 18 cm." (alla domanda "*Quanto è lungo uno spazzolino da denti?*"), o ancora "Dipende, in media 1,5 m." (nell'item "*Quanto è largo un marciapiede?*").

I dati raccolti nelle due modalità sono stati riportati all'interno di un file Excel dove sono stati uniformati per unità di misura prima delle analisi, secondo l'ordine di grandezza maggiormente presente nelle varie risposte raccolte per ciascun item (unità di misura di "moda" come, ad esempio, metri per la *larghezza del marciapiede*, centimetri per la *distanza tra sedile e volante dell'automobile* o per la *misura dello spazzolino* ecc.). Una volta resi i dati omogenei si è poi andati a considerare solo il valore assoluto delle misure, per praticità di analisi, considerando le stime registrate come variabili numeriche continue e non di tipo nominale.

3.2.3 Analisi dei dati e risultati

Le analisi dei dati sono state condotte utilizzando il software Jamovi versione 2.3 (The Jamovi project, 2022).

Al fine di verificare le proprietà psicometriche dello strumento si è andati preliminarmente ad indagare la possibilità di considerare nelle analisi tutto il campione dei 50 partecipanti coinvolti nello studio, a prescindere dalle modalità di somministrazione in autosomministrazione e con esaminatore in videochiamata.

Come prima cosa si è voluto verificare se i due sotto-campioni fossero omogenei per le loro caratteristiche sociodemografiche (età e scolarità), così da poter essere confrontabili. A tal scopo è stato utilizzato un test-t di Student a campioni indipendenti secondo la

modalità di somministrazione. I risultati indicano che i due campioni non differiscono in modo significativo per età ($t(48) = 0,829$; $p = 0,411$) e scolarità ($t(48) = -0,639$; $p = 0,526$) portando a considerare i due campioni comparabili secondo le variabili considerate, data un'omogeneità a livello della loro composizione. Utilizzando sempre il t-test a campioni indipendenti, si verifica la possibilità di comparare i punteggi totali del ToT, ottenuti ai tempi T0 e a T1 (variabili dipendenti) secondo la modalità di somministrazione. Al tempo T0 si ottiene un risultato non significativo ($t(48) = 0,612$; $p = 0,543$), mentre per il punteggio totale a T1 risulta essere significativo ($t(48) = 2,056$; $p = 0,045$). Da questi risultati emerge come le due modalità non risultano tra di loro equivalenti data la non omogeneità dei punteggi ottenuti. A livello metodologico saranno considerate le analisi separate secondo le modalità di somministrazione, con i rispettivi campioni di 25 partecipanti ciascuno. Per considerare il campione globale i due sottogruppi considerati avrebbero dovuto svolgere entrambe le modalità di somministrazione.

Dato quanto emerso, è stato utilizzato un metodo di esplorazione visiva qualitativa dei grafici descrittivi della distribuzione dei punteggi totali del ToT considerando i due tempi di somministrazione (T0 e T1); si confrontano le differenze tra le due modalità grazie all'esplorazione visiva degli istogrammi e dei *box plot*. Nonostante il t-test per campioni indipendenti non abbia riportato una differenza significativa, nei punteggi medi ottenuti nelle due modalità a T0 ($M = 4,20$ autosomministrazione contro $M = 3,80$ con esaminatore in videochiamata) dagli istogrammi risulta una distribuzione dei punteggi differente tra le due modalità. In particolare, i partecipanti con somministrazione con esaminatore in videochiamata mostrano meno punteggi estremi rispetto a quelli dell'autosomministrazione, i quali ottengono punteggi totali anche superiori a 8 (*in autosomministrazione* riportato in Fig 1.1.).

Sempre in autosomministrazione, si ha una moda spostata a 3 rispetto a quella di 2 della modalità di somministrazione con esaminatore in videochiamata (*con esaminatore* riportato in Fig 1.1.). Da notare anche come in autosomministrazione non vi siano punteggi in corrispondenza dello zero, ed in entrambe le modalità non si abbiano punteggi che raggiungono il massimo punteggio ottenibile di 16. Al tempo T1 si rileva a livello descrittivo una media dei punteggi totali che risulta essere minore nella somministrazione in videochiamata con esaminatore ($M=3,12$) rispetto alla modalità autosomministrata ($M=4,40$). Tale differenza, emersa in maniera significativa nel t-test a campioni indipendenti, risulta evidente anche osservando a livello grafico la distribuzione dei punteggi che risulta molto differente con un andamento per la modalità autosomministrata (*in autosomministrazione*, riportato in Figura 1.2.) che replica quello di una curva gaussiana: meno casi di prestazioni particolarmente accurate, ma con anche la presenza di casi che raggiungono il punteggio totale di 10 (range da 0 a 10), cosa che non avviene per la distribuzione della somministrazione in videochiamata con esaminatore (*con esaminatore*, riportato in Figura 1.2.). Nella somministrazione con esaminatore in videochiamata si nota un range di punteggio totale più ristretto (minimo e massimo con valori da 0 a 7) ed un elevato numero di casi in corrispondenza dello zero. Ancora una volta non si rilevano punteggi massimi di 16 in entrambe le modalità, probabilmente data la considerazione di campioni di partecipanti sani.

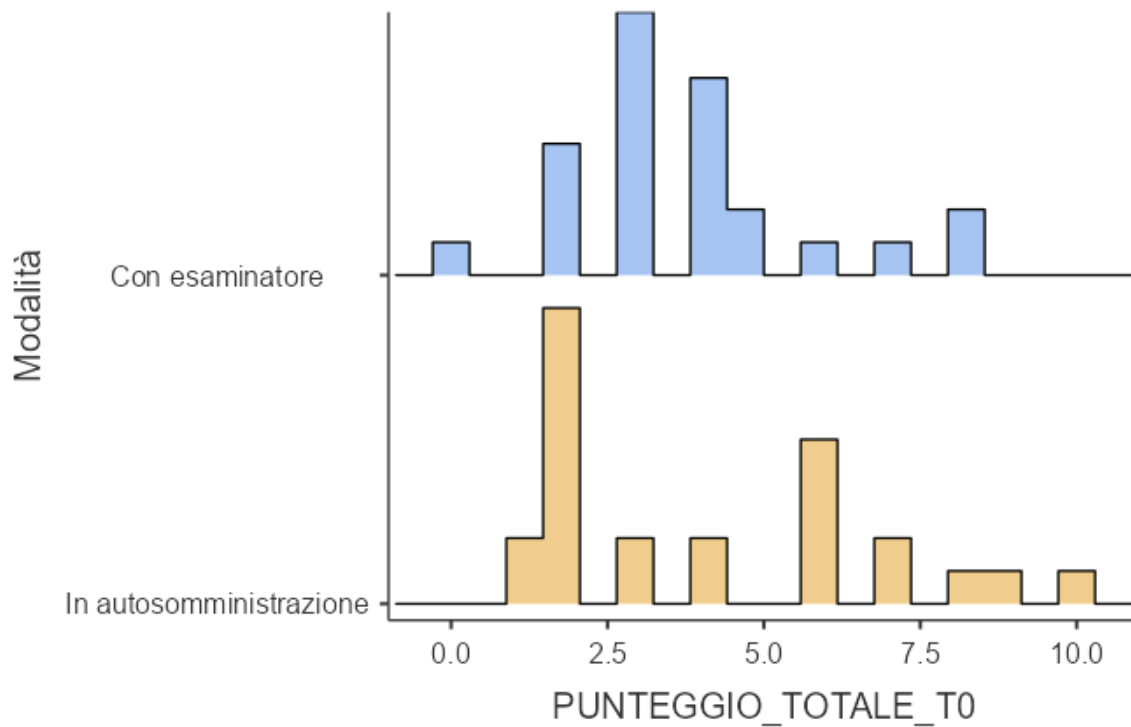


Figura 1.1. Istogramma descrittivo della distribuzione dei punteggi totali ottenuti al tempo T0 secondo le due modalità

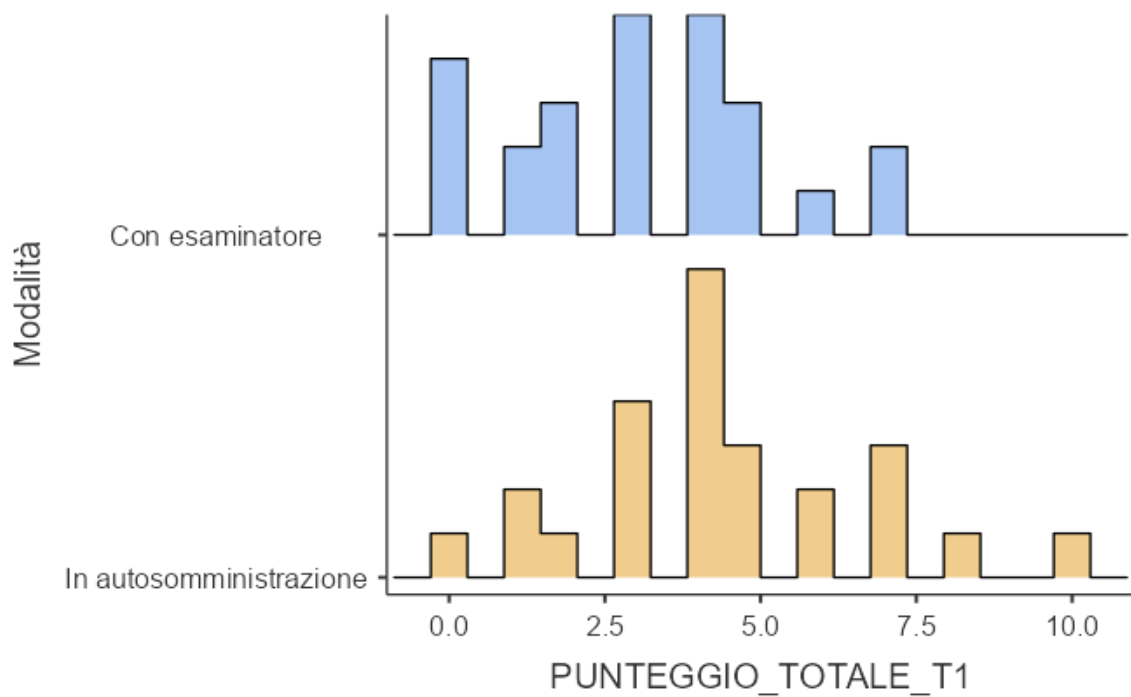


Figura 1.2. Istogramma descrittivo della distribuzione dei punteggi totali ottenuti al tempo T1 secondo le due modalità.

Conferme si ottengono dall'analisi ed osservazione dei *box plot*: a T0 si evidenzia una maggiore variabilità e deviazione standard nelle risposte della modalità autosomministrata (Figura 2.1); anche i quartili sono differenti, con un'asimmetria evidente nella modalità con esaminatore in videochiamata data una mediana in corrispondenza del 25° percentile. Nella modalità autosomministrata si nota una maggior distribuzione dei punteggi totali ottenuti con un minimo, anche in questo caso, in corrispondenza dello 0, ma con un massimo che arriva sino a 10. Al tempo T1 osservando i *box plot* si osserva una media dei punteggi totali inferiore ed una maggior variabilità per la modalità autosomministrata (*box* più grande) mentre per la somministrazione con esaminatore in videochiamata la mediana è esattamente al centro del *box* (Figura 2.2).

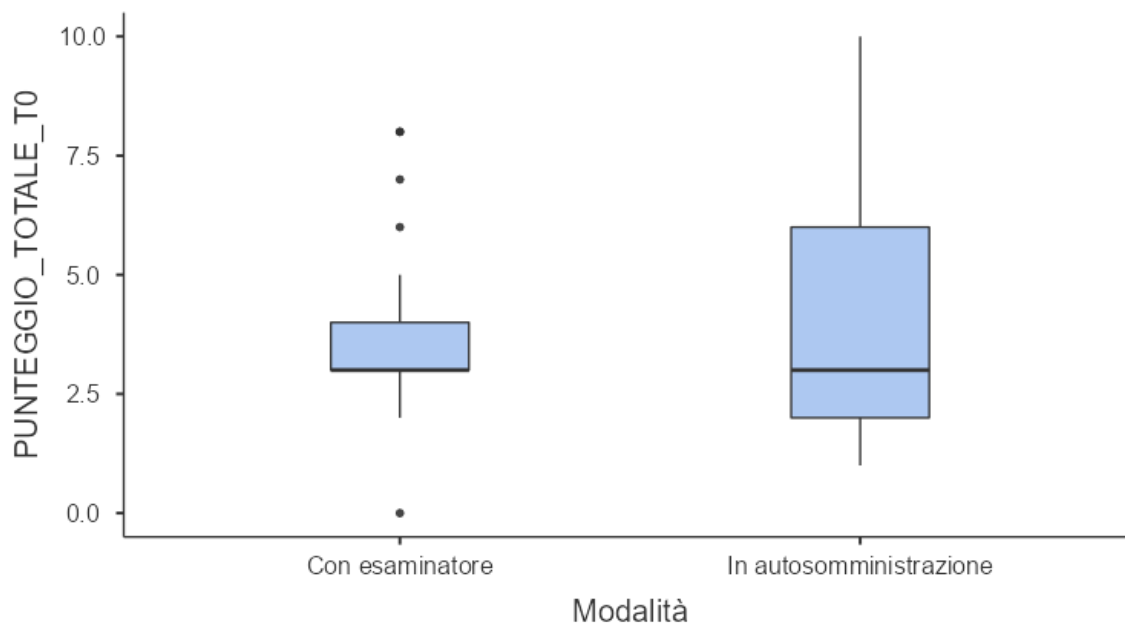


Figura 2.1. *Box plot* per il punteggio totale al tempo T0 per modalità in autosomministrazione e con esaminatore in videochiamata.

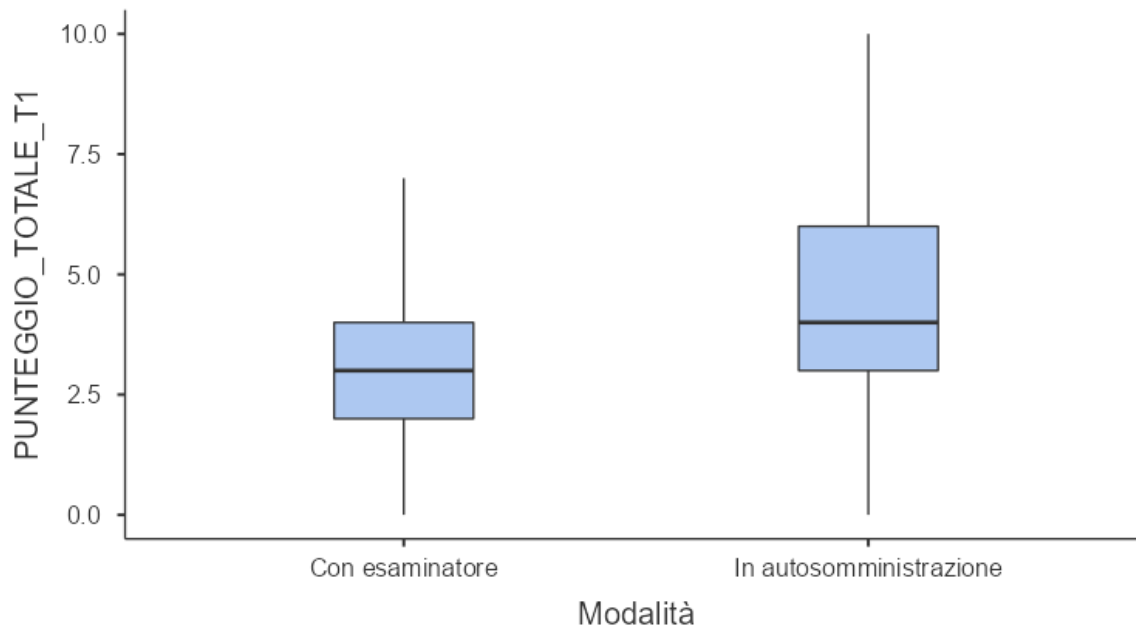


Figura 3.1. *Box plot* per il punteggio totale al tempo T1 per modalità in autosomministrazione e con esaminatore in videochiamata.

Successivamente, sono state calcolate alcune proprietà psicometriche dello strumento. Alla luce di quanto emerso dalle precedenti analisi, tali proprietà sono state calcolate separatamente per modalità di somministrazione, in quanto queste non sono comparabili.

La consistenza interna dello strumento è stata calcolata tramite l'alpha di Cronbach standardizzato considerando i due campioni distinti secondo le due modalità (in autosomministrazione e con esaminatore in videochiamata): è stata riscontrata una buona consistenza interna sia per l'autosomministrazione con un alpha di 0,69 che per la somministrazione con esaminatore in videochiamata con un valore di alpha di 0,64.

Sono state poi verificate l'affidabilità test-retest e l'effetto pratica nei due diversi setting distinti tramite l' r di Pearson. La correlazione dei punteggi totali a distanza delle due somministrazioni T0 e T1 è risultata essere significativa con

$p < 0,001$ ed un valore accettabile di $r = 0,59$ che porta a definire una buona affidabilità test-retest. I risultati ottenuti, però, devono essere confermati anche a livello di ciascun campione secondo la modalità di somministrazione: in autosomministrazione si ha un valore accettabile di $r = 0,69$ significativo con $p < 0,001$, per la somministrazione con esaminatore in videochiamata un valore di $r = 0,44$ significativo con $p = 0,028$; in entrambi i casi si registra una buona affidabilità test-retest. L'affidabilità test-retest, sempre tramite r di Pearson, è stata calcolata anche per i singoli item nelle due diverse modalità: i valori sono riportati in tabella 1.1. per la modalità autosomministrata ed in tabella 1.2 per la somministrazione con esaminatore in videochiamata. In entrambe la modalità sono emerse relazioni significative per gli item (4 su 8 in autosomministrazione e 5 su 8 con esaminatore in videochiamata) con particolare riferimento a quelli della *lunghezza del guinzaglio* e del *peso del dizionario* che mostrano una buona affidabilità test-retest in entrambe le modalità.

Auto somministrazione	Affidabilità test-retest		Effetto pratica		
	r di Pearson (tra T0 e T1)	p -value (<u>significativo</u>)	Differenza media	Test t campioni appaiati (tra T0 e T1)	p -value
<i>Spazzolino</i>	$r = 0,357$	0,087	-0,04	$t = -0,296$	0,770
<i>Gazzella</i>	$r = -0,075$	0,720	0,08	$t = 1,000$	0,327
<i>Guinzaglio</i>	<u>$r = 0,578$</u>	<u>0,002</u>	0,04	$t = 0,272$	0,788
<i>Margherita</i>	$r = -0,183$	0,382	0,12	$t = 0,531$	0,600
<i>Dizionario</i>	<u>$r = 0,495$</u>	<u>0,012</u>	-0,20	$t = -1,225$	0,233
<i>Bicchiere</i>	<u>$r = 0,404$</u>	<u>0,045</u>	0,00	$t = 0,000$	1,000
<i>Marciapiede</i>	$r = -0,021$	0,920	-0,08	$t = -0,440$	0,664
<i>Volante</i>	<u>$r = 0,505$</u>	<u>0,010</u>	-0,16	$t = -0,941$	0,356
Punteggio Totale	<u>$r = 0,689$</u>	<u>$< 0,001$</u>	-0,20	$t = -0,511$	0,614

Tabella 1.1 Affidabilità test-retest ed effetto pratica per gli 8 item del ToT e per i punteggi totali ottenuti, in modalità autosomministrata.

Al fine di indagare la possibile presenza di effetto pratica sono stati effettuati dei t-test per campioni appaiati: considerando il campione globale è emersa l'assenza di effetto pratica definito non significativo ($t(49) = 0,821$; $p=0,415$) con una differenza media dei punteggi totali del test tra T0 e T1 di 0,24; essa indica un leggero peggioramento della prestazione tra le due somministrazioni. Risultati che devono essere confermati suddividendo il campione per modalità: nella modalità in autosomministrazione, sempre tra i punteggi totali, non è emerso un effetto pratica significativo ($t(24) = -0,511$; $p=0,614$) con una differenza media dei punteggi ottenuti di -0,20 punti; essa indica un leggero miglioramento al tempo T1, rispetto alla prima somministrazione. Effetto pratica non significativo anche per la somministrazione con esaminatore in videochiamata ($t(24) = 1,607$; $p=0,121$) ma con una differenza media maggiore di 0,7 dei punteggi ottenuti che segnala un peggioramento della prestazione (confermata qualitativamente anche dall'affidabilità test-retest della modalità con esaminatore in videochiamata risultata minore rispetto in autosomministrazione). Ulteriori conferme di assenza dell'effetto pratica si rilevano considerando i singoli item, sempre con il test t a campioni appaiati, che non registra alcun effetto significativo nella modalità autosomministrata (tabella 1.1) e nella modalità di somministrazione con esaminatore in videochiamata, con l'eccezione per quest'ultima dell'item "*Quanta acqua contiene un bicchiere?*", con un effetto pratica significativo ($t(24) = 2,193$; $p=0,038$) ed un peggioramento medio della prestazione di quasi mezzo punto sul massimo ottenibile di 2 nel singolo item (0,44, vedi tabella 1.2).

Con esaminatore in videochiamata	Affidabilità test-retest		Effetto pratica		
	<i>r</i> di Pearson (tra T0 e T1)	<i>p-value</i> (significativo)	Differenza media	Test t campioni appaiati (tra T0 e T1)	<i>p-value</i> (significa tivo)
<i>Spazzolino</i>	<u><i>r</i> = 0,510</u>	<u>0,009</u>	-0,08	t = -0,811	0,425
<i>Gazzella</i>	<u><i>r</i> = 0,546</u>	<u>0,005</u>	0,04	t = 0,296	0,770
<i>Guinzaglio</i>	<u><i>r</i> = 0,629</u>	<u>< 0,001</u>	0,04	t = 0,440	0,644
<i>Margherita</i>	<i>r</i> = 0,104	0,622	0,08	t = 0,527	0,603
<i>Dizionario</i>	<u><i>r</i> = 0,547</u>	<u>0,005</u>	-0,16	t = -1,281	0,212
<i>Bicchiere</i>	<i>r</i> = 0,077	0,713	<u>0,44</u>	<u>t = 2,193</u>	<u>0,038</u>
<i>Marciapiede</i>	<u><i>r</i> = 0,598</u>	<u>0,002</u>	0,20	t = 2,000	0,057
<i>Volante</i>	<i>r</i> = -0,025	0,906	0,16	t = 0,811	0,425
Punteggio Totale	<u><i>r</i> = 0,439</u>	<u>0,028</u>	0,68	t = 1,607	0,121

Tabella 1.2. Affidabilità test-retest ed effetto pratica per gli 8 item del ToT e per i punteggi totali ottenuti, in modalità di somministrazione con esaminatore in videochiamata.

Infine, è stata indagata la possibile relazione tra i punteggi ottenuti al ToT e le variabili di età e scolarità accorpate tramite modello di regressione lineare: si procede adattando un modello di regressione lineare per verificare l'assenza o la presenza di relazione tra i punteggi totali (variabile dipendente) ottenuti nelle due modalità e le variabili di età e scolarità unite tra loro (come variabili indipendenti): per il tempo T0 il modello non è risultato significativo con $p=0,600$ e con $R^2=0,022$; età e scolarità non sono risultati essere predittori significativi rispettivamente con $p=0,466$ e $p=0,356$, con un valore β di $-0,115$ per la prima e di $-0,146$ per la

seconda. Il valore negativo di questi ultimi valori indica una prestazione leggermente migliore al test per individui con un maggior numero di anni di età e di scolarità. Anche il modello di regressione che considera il punteggio totale delle somministrazioni (variabile dipendente) a T1 è risultato essere non significativo, con $p=0,218$ e $R^2=0,063$; età e scolarità non risultano essere predittori significativi con rispettivamente $p=0,117$ e $p=0,933$, con un valore β di $-0,245$ per la prima e di $0,013$ per la seconda. Nei due modelli di regressione lineare sono stati successivamente considerati anche fattori come il genere e la modalità di somministrazione (sempre uniti ad età e scolarità) sia a T0 che a T1, senza mostrare alcuna relazione con i punteggi ottenuti al test; ad eccezione del fattore della modalità considerato al tempo T1. In quest'ultimo caso si è ottenuto un modello significativo con $p=0,040$ e un $R^2=0,164$, ad indicare il 16% della varianza spiegata che conferma i punteggi ottenuti a T1 tramite il test t di Student nella differenza tra le modalità di somministrazione, al netto di età e scolarità. Età e scolarità che, anche in questo ultimo caso, risultano essere predittori non significativi con rispettivamente $p=0,117$ e $p=0,933$.

3.3 Discussione

Prima di discutere quanto emerso dalle analisi, è giusto riportare alcuni limiti riscontrati a livello della distribuzione dell'età e della scolarità in entrambe le modalità, con un consistente numero di partecipanti di età compresa tra i 20 e i 30 anni (moda 23) oltre che una moda della scolarità elevata di 17 anni. Questo è dovuto anche alla scarsa numerosità del campione che dovrebbe essere ampliata considerando i partecipanti appartenenti alle altre fasi della procedura di validazione in corso. Sarebbe opportuno considerare anche partecipanti mediamente meno scolarizzati e più anziani, rispetto alle fasce di età e scolarità coperte dallo studio presentato.

Considerando i risultati ottenuti, non si evidenzia un'equivalenza tra le due modalità che permetta la considerazione delle analisi a livello del campione globale: osservando l'andamento grafico dei due sottogruppi, le modalità in autosomministrazione e di somministrazione con esaminatore in videochiamata, mostrano risultati differenti che potrebbero essere giustificati dalle caratteristiche dei due setting considerati. La prestazione durante la prima somministrazione del ToT (T0) risulta essere peggiore nella modalità autosomministrata rispetto a quella con esaminatore in videochiamata, con un'assenza nel primo caso di partecipanti che ottengono punteggio totale di 0 al test e quindi un'ottima prestazione. Al tempo T1 risulta come nella somministrazione con esaminatore in videochiamata il partecipante ottenga migliori punteggi rispetto a quella in autosomministrazione, sia con un maggior numero di prestazioni con punteggio totale minimo di 0 che con un minor numero di punteggi totali elevati. Probabilmente la presenza dell'esaminatore risulta essere un elemento che facilita la risposta del paziente guidandolo a quella corretta. O ancora si potrebbe considerare come il partecipante, in assenza dell'esaminatore, sia portato a rispondere in maniera più superficiale (in autosomministrazione presenza anche di risposte stravaganti ed ironiche), sia poco concentrato durante lo

svolgimento o non attento nel leggere le istruzioni (in riferimento alle omissioni di unità di misura registrate in autosomministrazione).

Dati i valori di affidabilità test-retest si osserva come il partecipante sia portato a mantenere la stessa misura data nei due tempi di somministrazione, soprattutto nella modalità autosomministrata: probabilmente questo dovuto alle minori influenze che può ricevere il partecipante nell'autosomministrazione rispetto alla modalità con l'esaminatore che presenta e somministra il test in videochiamata. I singoli item significativi con una buona affidabilità test-retest sono risultati essere *“Qual è la distanza tra una persona e il suo cane tenuto al guinzaglio?”* e *“Quanto pesa un dizionario?”*, probabilmente data una maggior convinzione della correttezza delle stime fornite dai partecipanti ai due specifici item che non ha portato a variazioni della risposta.

L'assenza dell'effetto pratica emersa dalle analisi permette potenzialmente di proporre il ToT con più somministrazioni nel tempo, senza che sia necessaria la costruzione di forme parallele o la creazione di item alternativi dello strumento, proposte invece dalle versioni di Scarpina et al. (2015) e MacPherson et al. (2014). L'assenza dell'effetto pratica potrebbe essere giustificata dal compito di stima che propone domande alle quali risulta difficile acquisire nel tempo una strategia di ragionamento e di astrazione migliore per arrivare alla risposta, se non informandosi su un valore di stima potenzialmente più accurato tra la prima e la seconda somministrazione. L'utilità delle forme ripetute nei test di stima cognitiva emerge nel considerare pazienti con problematiche frontali e non partecipanti sani: come riportano MacPherson et al. (2014), le forme parallele permettono di valutare più volte le prestazioni dei pazienti, nella pratica clinica o durante attività di ricerca, con la possibilità di misurare le capacità del paziente nel tempo e l'evoluzione delle sue prestazioni verificando eventuali miglioramenti.

L'unico item del ToT in cui risulta la presenza di un effetto pratica significativo è quello della *quantità d'acqua contenuta in un bicchiere*, nella modalità di somministrazione con esaminatore in videochiamata. Ciò probabilmente è da imputare agli accorgimenti o alle richieste di conferma da parte dell'esaminatore circa gli ordini di grandezza dell'unità di misura del litro (come riportato nella descrizione della procedura). L'intervento dell'esaminatore potrebbe aver portato ad un miglioramento nel tempo della stima fornita a questo particolare item.

Dai modelli di regressione lineare emerge come età e scolarità non rappresentino predittori significativi dei punteggi totali ottenuti ai due tempi di somministrazione. Questi risultati confermano in parte quanto emerso dalla letteratura: l'assenza di relazione con l'età è confermata da diverse versioni del CET (Axelrod e Millis, 1994; Della Sala et al., 2003 e 2004; Scarpina et al., 2015) dato il probabile coinvolgimento nel processo di stima cognitiva della componente cristallizzata dell'intelligenza, capace di resistere al deterioramento nel tempo (Cattell, 1963; Scarpina et al., 2015). Per la scolarità c'è un dibattito più aperto: alcuni contributi sottolineano una relazione assente (Scarpina et al., 2015), come nel caso di queste analisi del ToT, o solo parziale (Axelrod e Millis, 1994) con i punteggi dei test di stima; altri rilevano migliori prestazioni associate ad un più elevato numero di anni di scolarizzazione (Shallice ed Evans 1978; Della Sala et al., 2003; MacPherson et al., 2014). Seppur non significativa, nel ToT emerge una leggera tendenza ad ottenere migliori prestazioni all'aumentare dell'età e della scolarità del partecipante. Questo potrebbe essere giustificato da un maggior numero di nozioni acquisite nel tempo attraverso l'esperienza o la formazione didattica, con un vantaggio nelle conoscenze e nel ragionamento. Ciò, però, contrasta con le capacità di memoria, di controllo esecutivo e di ragionamento astratto, reclutate nella stima cognitiva, che peggiorano all'avanzare dell'età (Craik e Salthouse,

2008). L'assenza di influenza della variabile genere per i punteggi ottenuti al ToT emersa dalle analisi, ricordando la scarsa numerosità del campione considerato, contrasta con quanto evidenziato dalle versioni dei test di stima presentate: soprattutto nelle versioni del CET di Della Sala et al. (2003) e di Scarpina et al. (2015) si sono ottenute prestazioni migliori, anche in maniera importante, per individui maschi rispetto alle femmine.

Andando oltre agli aspetti statistici e quantitativi si possono considerare casi in cui risulta evidente l'importanza dell'analisi e dell'interpretazione qualitativa dei dati raccolti, con prestazioni di alcuni partecipanti che possono mettere in discussione lo *scoring* ottenuto dalla somministrazione del ToT. Un primo caso da citare è quello di partecipanti con punteggi totali molto ridotti che dovrebbero indicare una buona prestazione, data l'attribuzione di punteggi elevati a prestazioni fuori dalla norma; si notano, però, risposte ad alcuni item che si collocano esattamente agli estremi dell'intervallo di riferimento a cui si attribuisce il punteggio grezzo di 0. In altri casi si è registrato anche il *pattern* opposto: un punteggio totale elevato, con risposte che risultano leggermente al di fuori dal range corretto di risposta, a cui si attribuisce il punteggio di 0.

Altro caso da considerare è quello legato al punteggio ottenuto ai singoli item da parte dei partecipanti in relazione agli *hobby* o alla professione svolta nella vita di tutti i giorni, sempre in riferimento al campione considerato da questo lavoro: per esempio, un insegnante delle elementari nella raccolta dati della somministrazione con esaminatore in videochiamata, sia a T0 che a T1, ottiene un punteggio grezzo di 1 sottostimando la risposta relativa all'item sul *peso del dizionario* (a T1 unica risposta fuori range). Teoricamente il dizionario rappresenta un oggetto di cui il partecipante dovrebbe possedere una certa familiarità, data la professione svolta. Un altro caso analogo può essere quello di un partecipante che svolge la professione del meccanico che ha

sovrastimato la misura della *distanza tra il sedile di guida ed il volante dell'automobile*, con una attribuzione del punteggio grezzo di 2. Ultimo caso citabile è quello di un partecipante con la passione per i cani, svolgendo attività venatoria, che ha sottostimato la misura della *lunghezza del guinzaglio*, ottenendo un punteggio grezzo di 1 all'item di riferimento. Diviene interessante notare come la familiarità con certi oggetti legati alla propria professione o passione possa non essere un elemento di facilitazione nella capacità di stima; d'altronde anche oggetti considerati nel ToT, come lo spazzolino da denti, risultano essere d'uso quotidiano per tutti i partecipanti a prescindere dalla professione svolta. Queste osservazioni potrebbero essere influenzate dall'aspettativa dell'esaminatore sulle risposte fornite alla luce delle informazioni ricavate dalla conoscenza del partecipante o acquisite durante la presentazione del test e la raccolta dati.

Altre osservazioni possono basarsi sulle risposte date dai partecipanti osservando i loro cambiamenti, o mantenimenti, nel tempo tra la fase test e quella retest: in alcuni casi, si osserva come a distanza delle sei settimane tra le somministrazioni vi sia un completo stravolgimento della risposta, pur ricordando i limiti del campione. Un partecipante nello stimare la *distanza del sedile di guida dal volante dell'automobile* nella prima somministrazione esprime un valore di 40 centimetri, per poi riportare una stima di 1,5 metri nella seconda. O ancora per il *peso del dizionario* si registra a T0 la risposta di un partecipante di 700 grammi che, alla risposta a T1, viene incrementata fino 6 chilogrammi. Casi di stime fornite dallo stesso partecipante molto differenti tra loro in un arco di tempo relativamente breve, spesso accompagnate da ferme convinzioni da parte dei partecipanti di avere dato la stessa risposta in entrambe le somministrazioni. Questo ultimo aspetto, legato ad un ricordo certo della prima stima data in realtà distorto, è stato riscontrato non solo nei due casi particolari citati, ma anche in altri

individui al termine della seconda somministrazione seppur con risposte meno divergenti tra T0 e T1. Durante la seconda somministrazione non è stato comunicato dall'esaminatore se le domande proposte fossero o meno le stesse della prima somministrazione: questo per evitare il condizionamento dell'influenza del ricordo delle risposte date la prima volta. Importante evitare di condizionare l'individuo nel riportare a T1 le stime date in precedenza a T0 (che possono basarsi su un ricordo perturbato) oppure nel tentativo del partecipante di modificare appositamente la risposta fornita a T0 al tempo T1, basandosi sempre sulla memoria di quanto espresso nella prima somministrazione. È invece stato spesso necessario dopo la prima somministrazione sollecitare i partecipanti a non informarsi sulle possibili risposte alle domande prima di aver partecipato anche alla seconda somministrazione. Quest'ultimo aspetto non è stato inserito nelle istruzioni formali del test, ma si è reso utile a livello pratico soprattutto alla luce della particolare curiosità espressa da buona parte dei partecipanti nel conoscere le risposte corrette, già dopo aver terminato la prima somministrazione. Inoltre, è capitato che i partecipanti chiedessero un riscontro sul punteggio ottenuto oppure un confronto tra le prestazioni della prima e della seconda somministrazione: in questo senso è stato opportuno sottolineare le finalità di ricerca dello studio, volto ad indagare aspetti statistici e non di diagnosi, ricordando la possibilità di rivolgersi ad uno specialista per l'analisi dei punteggi grezzi, fornibili al partecipante qualora li richiedesse.

Altre osservazioni emergono sull'utilizzo delle unità di misura da parte dei partecipanti che, come riportato nelle istruzioni del ToT, sono a discrezione dell'individuo che risponde: in particolare nell'item in cui viene richiesta la *quantità d'acqua contenuta nel bicchiere* i partecipanti riportano risposte che variano molto secondo i diversi ordini di grandezza del litro (litri, decilitri, centilitri, millilitri). Questo rileva una confusione nella scelta

dell'unità di misura reputata corretta; molta più chiarezza è stata riscontrata per item che richiedono idealmente risposte in centimetri (e metri) o chilometri orari. È capitato anche il caso di un partecipante di professione geometra che ha fornito la misura del bicchiere in centimetri cubi sia a T0 che a T1, probabilmente in relazione alle nozioni che ha acquisito tramite la propria professione. Un ultimo aspetto rilevante delle risposte legate a questo item è stato notare in alcuni casi sia riportato il dato di “33 cl” o “330 ml”: una misura abbastanza precisa, legata probabilmente alla familiarità dei partecipanti con il formato delle lattine di alluminio, su cui sono riportate esplicitamente proprio queste specifiche cifre e unità di misura.

3.3.1 Conclusioni

Il test ToT, presentato in questo lavoro, è risultato essere uno strumento con buone proprietà psicometriche di valutazione della stima cognitiva: dalle analisi condotte nella fase di test-retest emergono una buona affidabilità test-retest ed un'assenza dell'effetto pratica; dai risultati ottenuti la modalità di somministrazione con esaminatore attraverso videochiamata risulta essere meno affidabile, ipotizzando come elementi del setting di valutazione possano aver influenzato la prestazione del partecipante, in particolare data la presenza dell'esaminatore via *webcam*. Le variabili di età, genere e scolarità non risultano esser capaci di predire il punteggio totale ottenuto dal partecipante, ad eccezione della modalità di somministrazione che porta ad una leggera influenza dei risultati. I risultati ottenuti sono incoraggianti in attesa delle analisi che emergeranno una volta conclusa la procedura di validazione; bisogna comunque sottolineare elementi del campione considerato che limitano la validità dei risultati ottenuti: una bassa numerosità del campione test-retest ed un elevato numero di partecipanti giovani con un'alta scolarità, i quali potrebbero aver condizionato quanto emerso dalle analisi.

La costruzione del ToT si è basata sulla raccolta degli elementi principali appartenenti ai diversi test neuropsicologici di valutazione della stima cognitiva presentati in letteratura. Si potrebbero somministrare in associazione al ToT prove (come il WCST, Grant e Berg, 1948; o il Test di Fluenza Verbale semantica, Novelli et al., 1986) o batterie (come la FAB, Dubois et al., 2000; o la BADS, Wilson et al., 1996) che valutano il funzionamento esecutivo generale per andare ad evidenziare possibili relazioni con i punteggi ottenuti a questo nuovo strumento di valutazione di stima.

Risulta importante interpretare la prestazione del partecipante anche a livello qualitativo per rilevare una prestazione deficitaria o nella norma, evitando di considerare solo i punteggi totali e gli aspetti numerici ricavati degli item.

Si potrebbero anche considerare, nella raccolta dei dati preliminari, la professione o gli interessi di ogni partecipante andando ad ampliare le possibili considerazioni di come la familiarità con certi oggetti o situazioni possa portare un vantaggio nella prestazione di stima ai diversi item, e a verificare eventuali relazioni tra la familiarità con i contenuti delle domande e i punteggi ottenuti.

Per quanto riguarda i singoli item si potrebbe pensare ad un item alternativo solo per la domanda riguardante la *velocità massima della gazzella*, non per motivi statistici ma per una finezza legata alla concezione delle diverse domande del ToT che hanno come oggetto comune elementi propri dell'esperienza quotidiana (si potrebbe considerare un animale domestico, per esempio).

Una volta conclusa la procedura di validazione condotta su partecipanti sani, si potrebbe utilizzare il test ToT direttamente su campioni di popolazioni cliniche, andando a confrontare i risultati ottenuti. Si possono ipotizzare come primi pazienti a cui sottoporre lo strumento quelli neurologici frontali riportati nei diversi studi, ma anche pazienti con patologia psichiatrica. Dagli studi raccolti

nell'elaborato emergono poche considerazioni per le problematiche psichiatriche, si potrebbero testare attraverso il ToT pazienti che soffrono di disturbi alimentari, o psicopatologie associate a perfezionismo come nel disturbo ossessivo-compulsivo (Gruner e Pittenger, 2017), caratterizzati da un'esagerata attenzione al dettaglio, rigidità cognitiva (valutabile col WCST, Grant e Berg, 1948) e problematiche di *set-shifting* (Marchiol et al., 2020) che portano a problematiche della capacità di integrazione delle informazioni del contesto e problematiche esecutive a livello di capacità di astrazione e processi di generalizzazione (Roberts, Barthel, Lopez, Tchanturia, e Treasure, 2011). Le risposte di questi pazienti a domande di stima potrebbero risultare particolarmente accurate, data l'eccessiva attenzione al dettaglio, oppure particolarmente bizzarre, per difficoltà ad integrare le informazioni.

Questo lavoro ha presentato il test ToT nelle sue caratteristiche principali mostrando le potenzialità che esso ha di misurare la capacità di stima cognitiva in un breve tempo di somministrazione; ulteriori considerazioni potranno essere fatte alla luce della conclusione della ricerca in corso che indaga le altre proprietà psicometriche dello strumento.

4. Bibliografia

- Anderson, S. W., Damasio, H., Jones, R. D., e Tranel, D. (1991). Wisconsin card sorting test performance as a measure of frontal lobe damage. *J Clin Exp Neuropsychol*, 13 (6), 909-922.
- Anderson, V., Northam, E., Hendy, J., e Wrenall, J. (2001). *Developmental neuropsychology: A clinical approach*. New York: Psychology Press.
- Axelrod, B. N., e Millis, S. R. (1994). Preliminary Standardization of the Cognitive Estimation Test. *Assessment*, 1 (3), 269-274.
- Brand, M., Kalbe, E., e Kessler, J. (2002a). Quantitative and qualitative differences in cognitive estimation of patients with probable Alzheimer's disease from healthy controls - What are the differences? *Memory and Emotions*. 425-429
- Brand, M., Kalbe, E., e Kessler, J. (2002b). *Test zum kognitiven Schätze*. Göttingen: Hogrefe
- Brand, M., Kalbe, E., Fujiwara, E., Huber, M., e Markowitsch, H. J. (2003). Cognitive estimations in patients with probable Alzheimer's disease and alcoholic Korsakoff patients. *Neuropsychologia*, 41, 575-584.
- Borkowsky, J. G., e Burke, J. E. (1996) Theories, models and measurements of executive functioning: An information processing perspective. *Attention, memory and executive functions (Lyon G. R., e Krasnegor N. A.)*. Baltimore: Paul H. Books.
- Bullard, S. E., Fein, D., Gleeson, M. K., Tischer, N., Mapou, R. L., e Kaplan, E. (2004). The Biber Cognitive Estimation Test. *Archives of Clinical Neuropsychology*, 19, 835-846.
- Caffarra, P., Vezzadini, G., Dieci, F., Zonato, F., e Venneri, A. (2002). Una versione abbreviata del test di Stroop: dati normativi nella popolazione italiana. *Rivista di Neurologia*, 12 (4), 111-115.

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

Craik, F. I. M., e Salthouse, T. A. (2008). *The Handbook of Aging and Cognition*. New York: Psychology Press.

D'Aniello, G. E., Scarpina, F., Albani, G., Castelnuovo, G., e Mauro, A. (2015). Disentangling the relationship between cognitive estimation abilities and executive functions: a study on patients with Parkinson's disease. *Neurol Sci*, 36 (8), 1425-1429

Della Sala, S., MacPherson, S. E., Phillips, L. H., Sacco, L., e Spinnler, H. (2003) How many camels are there in Italy? Cognitive estimates standardised on the Italian population. *Neurol Sci*, 24, 10-15.

Della Sala, S., MacPherson, S. E., Philips, L. H., Sacco, L, e Spinnler, H. (2004). The role of semantic knowledge on the cognitive estimation task. *J Neurol*, 251, 156-164.

Dubois, B., Slachevsky, A., e Isella, V. (2000). The FAB: frontal assessment battery at bedside. *Neurology*, 55, 1621-1626.

Duff, K., Callister, C., Dennett, K., e Tometich, D. (2012). Practice effects: a unique cognitive variable. *Clin Neuropsychol*, 26 (7), 1117-1127.

Folstein, M. F., Folstein, S. E., e McHugh, P. R. (1975). "Minimal state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.*, 12 (3), 189-198.

Freund, B., Gravenstein, S., Ferris, R., Burke, B. L., e Shaheen, E. (2005). Drawing Clocks and Driving Cars. *J Gen Intern Med*, 20, 240-244

Gansler, D. A., Varvaris, M., Swenson, L., e Schretlen, D. J. (2014). Cognitive estimation and its assessment. *Journal of Clinical and Experimental Neuropsychology*, 36 (6), 559-568.

Gillespie, D. C., Evans, R. L., Gardener, E. A., e Bowen, A. (2002). Performance of older adults on tests of cognitive estimation. *J Clin Exp Neuropsychol*, 24 (3), 286-293.

Goldstein, F. C., Green, J., Presley, R. M., O'Jile, J. R., Freeman, A., Watts, R., e Green, R. C. (1996). Cognitive estimation in patients with Alzheimer's disease. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, 9, 35-42.

Goldstein, K. H., e Scheerer, M. (1941). Abstract and concrete behavior: an experimental study with special tests. *Psychol Monogr*, 53 (2), no. 239.

Grafman, J. (1989). *Plans, actions and mental sets: The role of the frontal lobes. In Integrating theory and practice in clinical neuropsychologyed.* Hove: Erlbaum.

Grafman, J., e Hendler, J. (1991). Planning and the brain. *Behav Brain Sci*, 14, 563-564.

Grant, D. A., e Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in Weigl-type card-sorting problem. *Journal of Experimental Psychology*, 38, 404-411.

Gruner, P., e Pittenger, C. (2017). Cognitive inflexibility in obsessive-compulsive disorder. *Neuroscience*, 345, 243-255.

Hobson, P., e Leeds, L. (2001). Executive functioning in older people. *Reviews in Clinical Gerontology*, 11, 361-372.

Jackson, C. T., Fein, D., Essock, S. M., e Mueser, K. T. (2001). The effects of cognitive impairment and substance abuse on psychiatric hospitalizations. *Community Mental Health Journal*, 37, 303-312.

- Jernigan, T. L., Schafer, K., Butters, N., e Cermak, L. S. (1991). Magnetic Resonance Imaging of alcoholic Korsakoff patients. *Neuropsychopharmacology*, 4, 175-186.
- Jurado, M. B., e Rosselli, M. (2002). The Elusive Nature of Executive Functions: A Review of our Current Understanding. *Neuropsychol Rev*, 17, 213-233.
- Krabbe, P. F. M. (2017). *The Measurement of Health and Health Status*. Amsterdam: Academic Press Books.
- La Rue, A. (1992). *Aging and neuropsychological assessment*. New York: Plenum Press.
- Leng, N. R. C., e Parkin, A. J. (1988). Double dissociation of frontal dysfunction in organic amnesia. *British Journal of Clinical Psychology*, 27, 359-362.
- Leng, N. R. C., e Parkin, A. J. (1993). *Neuropsychology of the Amnesic Syndrome*. Hove: Erlbaum.
- Levinoff, E. J., Verret, L., Akerib, V., Phillips, N. A., Babins, L., Kelner, N., e Chertkow, H. (2006). Cognitive estimation impairment in Alzheimer Disease and Mild Cognitive Impairment. *Neuropsychology*, 20 (1), 123-132.
- Liss, M., Fein, D., Bullard, S., Robins, D., e Waterhouse, L. (2000). Cognitive estimation in individuals with pervasive developmental disorders. *Journal of Autism and Developmental Disabilities*, 30, 613-618.
- Luria, A. R. (1966). *Higher Cortical Functions in Man*. London: Tavistock.
- Luria, A. R. (1973). *The Working brain: An introduction to neuropsychology*. New York: Basic.
- MacPherson, S. E., Wagner, G.P., Murphy, P., Bozzali, M., Cipolotti, L., e Shallice, T. (2014). Bringing the Cognitive

Estimation Task into the 21st Century: Normative Data on Two New Parallel Forms. *PLoS ONE*, 9 (3), e92554.

Marchiol, F., Lionetti, F., Luxardi, G. L., Cavallero, C., Roberts, M. E., e Penolazzi B. (2020). Cognitive inflexibility and over-attention to detail: The Italian validation of the DFlex Questionnaire in patients with eating disorders. *Eur Eat Disorders*, 1-16.

Margraf, N., Bachmann, T., Schwandner, W., Gottschalk, S., e Seidel, G. (2009) Bedside screening for executive dysfunction in patients with subcortical ischemic vascular disease. *Int J Geriatr Psychiatry*, 24, 1002-1009.

Mariani, C., Sacco, L., Spinnler, H., e Venneri, A. (2002). General knowledge of the world: a standardised assessment. *Neurol Sci*, 23, 161-175.

Meyers, J.E., e Meyers, K. R. (1995). *The Meyers scoring system for the Rey complex figure and recognition trial: professional manual*. Florida: Psychol Assess Resour.

Monaco, M., Costa, A., Caltagirone, C., e Carlesimo, G. A. (2013). Forward and backward span for verbal and visuo-spatial data: standardization and normative data from an Italian adult population. *Neurol Sci*, 34, 749-754.

Mondini, S., Mapelli, D., Vestri, A., e Bisiacchi, P. S. (2003). *Esame neuropsicologico breve: una batteria di test per lo screening neuropsicologico*. Milano: Raffaello Cortina Editore.

Nichelli, P., Leone, M., Caronna, A., Imbornone, E., Alberoni, M., Zuffi, M., e Venneri, A. (2002). *Nuova Rivista di Neurologia*, 12 (2), 37-42.

Novelli, G., Papagno, C., Capitani, E., Laiacona, M., et al. (1986). Tre test clinici di memoria verbale a lungo termine: Taratura su soggetti normali. *Archivio di Psicologia, Neurologia e Psichiatria*, 47, 477-506.

- O'Carroll, R., Egan, V., e MacKenzie, D. M. (1994). Assessing cognitive estimation. *British Journal of Clinical Psychology*, *33*, 193-197.
- Petersen, R. C. (2016). Mild cognitive impairment. *Continuum Dementia*, *22* (2), 404-418.
- Raven, J. C. (1938). *Guide to the Progressive Matrices sets A, B, C, D and E*. London: H. K. Lewis.
- Roberts, M. E., Barthel, F. M. S., Lopez, C., Tchanturia, K., e Treasure, J. L. (2011). Development and validation of the Detail and Flexibility Questionnaire (DFlex) in eating disorders. *Eating Behaviors*, *12* (3), 168-174.
- Royall, D. R., Cordes, J. A., e Polk, M. (1998). CLOX: an executive clock drawing task. *J Neurol Neurosurg Psychiatry*, *64* (5), 588-594.
- Scarpina, F., D'Aniello, G. E., Mauro, A., Castelnuovo, G., e MacPherson, S. E. (2015). How many segments are there in an orange: normative data for the new Cognitive Estimation Task in an Italian population. *Neurol Sci*, *36*, 1889-1895.
- Shallice, T., ed Evans, M. E., (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex*, *14*, 294-303.
- Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society London*, *298*, 199-209.
- Shoqeirat, M. A., Mayes, A., MacDonald, C., Meudell, P., e Pickering, A. (1990) Performance on tests sensitive to frontal lobe lesions by patients with organic amnesia: Leng & Parkin revisited. *Br J Clin Psychol*, *29*, 401-408
- Spencer, R. J., e Johnson-Greene, D. (2009). The cognitive estimation test: psychometric limitations in neurorehabilitation populations. *J Clin Exp Neuropsychol*, *31* (3), 373-377.

- Squire, L. R., Amaral, D. G., e Press, G. A. (1990). Magnetic resonance imaging of the hippocampal formation and mammillary nuclei distinguish medial temporal lobe and diencephalic amnesia. *Journal of Neuroscience*, *10*, 3106-3117.
- Strauss, E., Sherman E. M. S., e Spreen, O. (2006). *A compendium of neuropsychological tests*. Oxford: Oxford University Press.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J Exp Psychol*, *18*(6), 643-662.
- Stuss, D. T., e Alexander, M. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological Research*, *63*, 289-298.
- Stuss, D. T., Alexander, M., Floden, D., Binns, M. A., Levine, B., McIntosh, A. R., et al. (2002). Fractionation and localization of distinct frontal lobe processes: Evidence from focal lesions in humans. *Principles of frontal lobe function (Stuss D. T., e Knight R. T., pp. 392-407)*. New York: Oxford University Press.
- Taylor, R., ed O'Carrol, R. (1995). Cognitive estimation in neurological disorders. *British Journal of Clinical Psychology*, *34*, 223-228.
- Teasdale, G., Jennett, B. (1974). Assessment of coma and impaired consciousness. A pratical scale. *The Lancet*, *2* (7872), 81-84.
- The Jamovi project (2022). *jamovi* (Version 2.3) [Computer Software]. Retrived from <https://www.jamovi.org>.
- Wagner, G. P., MacPherson, S. E., Parente, M. A. M. P., e Trentini, C. M. (2011). Cognitive estimation abilities in healthy and clinical populations: the use of Cognitive Estimation Test. *Neurol Sci*, *32*, 203-210.
- Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., ed Evans, J. J. (1996). *Behavioural assessment of dysexecutive syndrome*. London: Harcourt Assessment.

Zelazo, P. D., Carter, A., Reznick, J., e Frye, D. (1997). Early development of executive functions: A problem-solving framework. *Review of General Psychology*, 1, 198-226.

Zelazo, P. D., e Müller, U. (2002). Executive function in typical and atypical development. *Blackwell handbook of childhood cognitive development* (Goswami, U., pp. 445–469). Malden: Blackwell.

5. Appendice: Think or Try (ToT)

Codice: _____ Data: _____ Et : _____ Genere: _____ Scolarit : _____

Le verranno proposte delle domande che non hanno una risposta precisa.

Le chiedo di fare una stima di quella che secondo lei potrebbe essere una risposta corretta indicando anche, dove possibile, la relativa unit  di misura.

DOMANDA	RISPOSTA	RANGE	PUNTEGGIO
Quanto � lungo uno spazzolino da denti?		0 = 15-22 cm 1 = 10-14 o 23-25 cm 2 = <10 o >25 cm	
Qual � la velocit� massima di una gazzella?		0 = 50-100 km/h 1 = 30-49 o 101-120 km/h 2 = <30 o >120 km/h	
Qual � la distanza tra una persona e il suo cane tenuto al guinzaglio?		0 = 80-180 cm 1 = 30-79 o 181-250 cm 2 = <30 o >250 cm	
Quanti petali ha una margherita?		0 = 15-40 1 = 10-14 o 41-70 2 = <10 o >70	
Quanto pesa un dizionario?		0 = 1-2 kg 1 = 0.5-0.99 o 2.01-3 kg 2 = <0.5 o >3 kg	
Quanta acqua contiene un bicchiere?		0 = 150-250 ml 1 = 100-149 o 251-350 ml 2 = <100 o >350 ml	
Quanto � largo un marciapiede?		0 = 80-180 cm 1 = 50-79 o 181-250 cm 2 = <50 o >250 cm	
Quanto dista il sedile dal volante di un'automobile?		0 = 40-60 cm 1 = 30-39 o 61-80 cm 2 = <30 o >80 cm	

Punteggio Totale _____