



UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Psicologia Generale

**Corso di laurea triennale in Scienze psicologiche
cognitive e psicobiologiche**

Tesi di laurea Triennale

**Il test della Torre di Londra di
Shallice: differenze tra i metodi di
scoring, criticità e versione
computerizzata**

**The Shallice Tower of London Task: different scoring methods, criticalities and
computer version**

Relatore
Prof. Luca Stefanutti

Laureanda:
Sofia Fresch
Matricola:
2013374

Anno Accademico 2022/2023

SOMMARIO

INTRODUZIONE	3
1.IL TEST DELLA TORRE DI LONDRA DI SHALLICE (1982)	3
2.ASPETTI CRITICI: IL LIVELLO DI DIFFICOLTÀ	5
2.1.Determinare il livello di difficoltà	5
2.2.Altri fattori che influiscono sulla difficoltà e sulla performance	9
2.3.Un possibile set di problemi strutturalmente bilanciato	12
2.4.La versione computerizzata del test	13
3.I METODI DI SCORING	15
3.1.I principali metodi di scoring utilizzati	15
3.2.Utilizzare un punteggio unico o dei punteggi parziali	19
3.3.Influenze sistematiche determinate dalla scelta del metodo di scoring	20
CONCLUSIONE	23
BIBLIOGRAFIA	24

INTRODUZIONE

Il test della Torre di Londra (Tower of London o ToL) è uno strumento utilizzato, sia in ambito clinico sia di ricerca, per valutare le capacità di pianificazione e problem solving.

Tuttavia, la considerevole diffusione di questo test non è accompagnata da altrettanta chiarezza in merito ad alcuni aspetti della sua struttura interna e all'interpretazione dei dati che esso fornisce: nello specifico, le questioni emerse all'incirca negli ultimi vent'anni riguardano rispettivamente la misurazione della difficoltà di un problema all'interno dello spazio della ToL e la scelta del metodo di calcolo del punteggio più adeguato. Molti studiosi hanno messo in luce come questa mancata univocità in merito a questi aspetti abbia portato alla produzione di un corpus di lavori, riguardanti la ToL, la cui comparabilità è quantomeno discutibile. Obiettivo di questa tesi è dunque fornire una panoramica di questi dibattiti e portare in luce, per quanto possibile, quegli aspetti la cui chiarificazione potrebbe contribuire ad un utilizzo più consapevole di questo strumento.

1. IL TEST DELLA TORRE DI LONDRA DI SHALLICE (1982)

Il test della Torre di Londra è stato prodotto e utilizzato per la prima volta dal neuropsicologo inglese Timothy Shallice nel 1982: si tratta di una versione alternativa di un test già esistente, chiamato Torre di Hanoi e costituito da una piattaforma di tre bastoncini su cui sono presenti tre dischi di dimensioni diverse, che vanno riposizionati da una conformazione iniziale ad una finale predeterminata. Shallice dichiarò di preferire il proprio test rispetto alla Torre di Hanoi, in quanto quella di Londra è stata costruita in modo tale da permettere di creare esercizi con livelli di difficoltà graduati e permette inoltre di proporre una maggiore varietà di problemi differenti (Berg e Byrd, 2002).

Il funzionamento del test in questione è il seguente: si ha un supporto da cui si stagliano tre bastoncini di dimensioni diverse, ossia capaci di accogliere rispettivamente tre, due o una pallina o sferetta (diversamente dalla versione precedente). Le palline utilizzate presentano tre colori diversi, solitamente blu,

verde e rosso. L'esercizio richiede che il partecipante ridisponga le palline a partire da una conformazione iniziale, sempre uguale, alla conformazione finale richiesta, rispettando il numero minimo di mosse indicato al momento della somministrazione. Il tutto va eseguito rispettando alcune regole fondamentali: è possibile muovere solamente una pallina per volta e, una volta presa una pallina, essa va necessariamente riposizionata in uno dei bastoncini e non può essere appoggiata o messa da parte. Al termine di ogni esercizio, le sfere vengono riposizionate nella conformazione iniziale, per poi passare all'esercizio successivo, con un aumento progressivo del numero minimo di mosse richiesto per i singoli esercizi.

Il test della Torre di Londra è uno strumento che nasce per l'assessment della pianificazione, la quale rientra tra le funzioni esecutive, ossia quelle funzioni di "alto livello" o metacognitive che si occupano di gestire altre funzioni cognitive minori, nonché le emozioni e l'attenzione, necessarie alla messa in atto di comportamenti finalizzati (Etnier and Chang, 2009). Carlin et al. (2000) hanno descritto i processi cognitivi coinvolti in questo test come "un meccanismo di anticipazione volto a generare molteplici sequenze di eventi ipotetici e di conseguenze di questi ultimi, tra cui lo sviluppo di complessi di eventi strutturati che possano guidare il passaggio da uno stato iniziale ad uno stato finale, l'anticipazione di eventi futuri legati all'esecuzione e il riconoscimento dell'avvenuto raggiungimento dell'obiettivo". Esiste inoltre uno studio (Unterrainer et al., 2004) i cui risultati ci permettono di affermare che il test della Torre di Londra misura prevalentemente le abilità di pianificazione e di problem solving, non essendo sufficientemente correlato ad altre funzioni cognitive.

Negli anni che hanno seguito il lavoro di Shallice, è stato prodotto un considerevole numero di versioni alternative del test: la cosiddetta "ToL DX", "ToL-R", "4-rod ToL", "five disk Tower of London" e altre ancora. Alcune di queste varianti possono essere definite come delle "isomorfe" della versione di Shallice del 1982, ossia si servono di un supporto che presenta un aspetto diverso rispetto all'originale ma lo spazio problema rimane inalterato (lo spazio problema si può definire come una sorta di mappa che indica il modo di muoversi da ogni

posizione legittima possibile all'interno del test ad un'altra, il tutto sempre attraverso mosse legittime) (Berg e Byrd, 2002).

La scelta all'interno di questo lavoro è stata quella di porre l'attenzione, per quanto possibile, esclusivamente a quegli studi riguardanti la versione originale del test.

2. ASPETTI CRITICI: IL LIVELLO DI DIFFICOLTÀ

2.1. Determinare il livello di difficoltà

Può essere utile, sia per i clinici che per i ricercatori, riuscire a misurare con una certa accuratezza la difficoltà di un problema, affinché il test sia sensibile alle differenti abilità di pianificazione delle varie popolazioni analizzate. La modalità principale di valutare la difficoltà di un problema all'interno del sistema della Torre di Londra è quello di considerare il numero minimo di mosse in cui è necessario che il problema stesso sia completato (Berg e Byrd, 2010).

Un considerevole numero di studi, tuttavia, ha messo in luce come alcune proprietà strutturali del compito abbiano influenze sistematiche nei processi cognitivi legati alla pianificazione (Kaller, 2004) e come, dunque, la loro presenza modifichi intrinsecamente il livello di difficoltà del compito. In altre parole, in molte circostanze mantenere inalterato il numero minimo di mosse non è una misura sufficiente per mantenere costante il livello di difficoltà. (Berg e Byrd, 2010).

Per esplorare adeguatamente il problema è necessario esplicitare alcuni concetti riguardanti la struttura e le caratteristiche del compito:

- Goal hierarchy: l'ovvietà con cui una pallina deve essere messa a posto prima delle altre. Ad esempio, in un problema che ha come conformazione iniziale quella delle tre palline disposte nello stesso bastoncino, la goal hierarchy ha il suo livello minimo di ambiguità, in quanto risulta ovvio che la prima pallina da utilizzare sia quella posizionata sopra alle altre due. Il massimo livello di ambiguità si ha invece in quei problemi che presentano una pallina per ciascun bastoncino.
- Start position hierarchy: l'ovvietà con cui una pallina deve essere mossa per prima. Non si tratta, in questo caso, della pallina che per prima deve

raggiungere la sua posizione finale affinché il problema sia risolto correttamente, ma della prima pallina che deve essere afferrata e spostata nel momento in cui il problema ha inizio.

- Suboptimal alternative: percorso di risoluzione che permette di risolvere il problema con una o due mosse in più rispetto a quelle previste.
- Subgoaling: una richiesta di subgoal consiste nella necessità di produrre una mossa essenziale alla soluzione ottimale, ma in cui la palla non si trova nella sua posizione finale.
- Soluzione ottimale: una soluzione al problema che rientri nel numero di mosse richiesto. Può essere singola oppure avere delle alternative ottimali (Newman, 2009).

Alcuni studi affermano che, nei problemi da tre mosse, l'accuratezza nella risoluzione cala da più del 95% all'83% nel caso in cui la goal hierarchy sia ambigua e sia presente un determinato tipo di richieste subottimali. Si nota inoltre un aumento significativo dei tempi di pianificazione nel caso di goal hierarchy ambigua, presente singolarmente o combinata rispettivamente a alternative subottimali o richieste di subgoal, nonché nel caso in cui siano combinate alternative subottimali e richieste di subgoal.

Secondo un altro studio (Newman, 2009), nei problemi da sei mosse l'accuratezza è influenzata dalla goal hierarchy ambigua nel caso in cui ci sia una sola soluzione ottimale. La goal hierarchy ambigua ha un'influenza anche sui tempi di pre-pianificazione.

La goal hierarchy produce inoltre delle modificazioni significative dell'attività cerebrale visibili tramite risonanza magnetica funzionale a livello di corteccia prefrontale sinistra, corteccia cingolata anteriore, corteccia precentrale bilaterale e gangli della base, da sola o nella sua interazione con il numero di soluzioni ottimali; questo secondo scenario si ha in quanto la goal hierarchy interferisce con la pianificazione, condizionando l'abilità di determinare il percorso ottimale.

Un terzo studio (Newman e Pittman, 2007) indaga ulteriormente gli effetti della goal hierarchy e del numero di percorsi di soluzione ottimali all'interno di una serie di problemi con un numero minimo di mosse pari a sei.

I risultati mostrano che il numero di tentativi risolti all'interno del numero minimo di mosse è stato significativamente influenzato da questi due fattori, su molteplici livelli.

Il primo livello è quello dell'accuratezza: i problemi con varie soluzioni ottimali hanno elicitato un numero maggiore di soluzioni corrette rispetto ai problemi che presentavano un solo percorso ottimale; questi ultimi, infatti, tendono a generare una maggiore quantità di errori in quanto, se si ha di fronte la possibilità di un solo procedimento corretto, c'è una maggiore probabilità di scegliere un procedimento che non porterà alla soluzione, rispetto a quanto potrebbe succedere in quei problemi che offrono più soluzioni. Oltre ad un maggior numero di errori, questa tipologia di problema ha determinato anche un maggiore utilizzo di mosse extra.

Anche i problemi non ambigui hanno determinato un maggior numero di problemi risolti correttamente e, allo stesso modo, quelli ambigui hanno elicitato maggiori mosse extra.

Inoltre, si è notata un'interazione tra questi due fattori, che ha mostrato che la goal hierarchy influenza in modo maggiore i problemi con molteplici soluzioni ottimali rispetto a quelli con una sola soluzione.

Il secondo livello di analisi, ossia il tempo di pianificazione (che va dalla presentazione del problema al momento in cui si muove la prima pallina) è stato influenzato dalla goal hierarchy, che ha determinato un tempo di pianificazione più prolungato in caso di ambiguità, ma non dal numero di soluzioni ottimali, né dall'interazione dei due fattori in esame.

Il livello che riguarda il cosiddetto movement time (che va dal momento in cui si muove la prima pallina fino al momento in cui il problema viene risolto, nel caso in cui il soggetto riesca a risolverlo) è stato influenzato da entrambi i fattori, ma è interessante notare come il numero di soluzioni ottimali abbia influenzato questa misura solamente in specifici punti del problema: quelli che corrispondevano all'inizio di un cosiddetto "subgoal chunk", ossia una sequenza di mosse che porta al completamento di un subgoal. Un'assunzione che viene fatta per spiegare questo fenomeno è che la sequenza di mosse che costituisce un subgoal chunk sia programmata tutta insieme e che i picchi nel tempo siano legati

alla programmazione del chunk successivo. Di conseguenza, anche se i partecipanti programmano l'intera sequenza di mosse, queste sono divise in raggruppamenti più piccoli.

Una conclusione simile è stata tratta da Unterrainer et al. nel 2003, i quali hanno affermato che, durante la fase di pre-pianificazione, i partecipanti non costruiscano precisamente l'intera sequenza di mosse da svolgere per risolvere il compito, ma si occupino semplicemente di tenere a mente alcune cosiddette "manovre chiave", ossia poche mosse fondamentali. L'intera e precisa sequenza di mosse viene completata in itinere, ossia durante la fase di esecuzione del compito.

Berg e Byrd nel 2010 hanno inteso investigare nuovamente questi aspetti problematici del test della Torre di Londra, in uno studio che contenesse una disamina comprensiva delle precedenti scoperte e che andasse ad indagarle in maggiore profondità: per fare ciò, si sono servite innanzi tutto di un'analisi di una quantità e varietà di caratteristiche del problema maggiori di quelle utilizzate in precedenza (mosse minime, start hierarchy, goal hierarchy, numero di percorsi di soluzione ottimale, move choices), in modo tale da comprendere cosa porti determinate tipologie di esercizio ad essere particolarmente ostiche da risolvere; hanno infine utilizzato problemi con un numero minimo di mosse variabile.

Utilizzando quattro diverse classi di misurazione (successo nelle soluzioni, efficienza nelle mosse, inteso come numero di mosse extra e come optimal move score, cioè mosse che allontanano dalla soluzione sottratte alle mosse che avvicinano alla soluzione, tempo trascorso fino alla prima mossa e, infine, tempo di soluzione dopo la prima mossa) lo studio ha dato ulteriore conferma di molti dei risultati fondamentali raggiunti precedentemente da altri studi: i risultati confermano che la tipologia di start hierarchy, quella di goal hierarchy e il numero di percorsi di soluzione ottimali si possono considerare dei predittori della difficoltà di un problema all'interno del sistema della Torre di Londra.

2.2. **Altri fattori che influiscono sulla difficoltà e sulla performance**

Altri studi hanno fornito prova di ulteriori fattori che influenzano sistematicamente la performance al test.

Ampliando il tema delle proprietà del compito precedentemente citate, uno studio del 2010 ha ipotizzato che queste caratteristiche, ad esempio il numero minimo di mosse richieste o il numero di possibili alternative nella prima mossa, non influenzino solamente il livello di difficoltà dell'esercizio, ma possano anche implicare combinazioni di abilità cognitive differenti a seconda del caso (Koppenol-Gonzalez, Bouwmeester, e Boonstra, 2010). Generalmente, la memoria di lavoro e l'inibizione sono considerate le componenti delle funzioni esecutive che danno il contributo più importante alla performance nel test della Torre di Londra (Goel e Grafman, 1995; Huizinga, Dolan, e Van der Molen, 2006; Miyake et al., 2000): tuttavia, si ipotizza che i processi cognitivi misurati dal test possano dipendere in buona parte dalle istruzioni generali e dalle specifiche restrizioni fornite al momento del test, dunque è necessario investigare quali proprietà del compito sono correlate alle diverse abilità cognitive. Grazie anche ad uno studio di Huizinga et al. (2006), si è giunti a poter affermare che l'inibizione gioca un ruolo importante quando i problemi vengono volontariamente risolti entro il numero minimo di mosse indicato, ma che non abbia nessuna influenza sulle mosse che eccedono il numero minimo, né il segmento di tempo precedente all'esecuzione della prima mossa. Di conseguenza, vari ricercatori hanno affermato che la capacità del test di valutare le abilità di pianificazione è più alta quando non si fornisce al partecipante la possibilità di ritrattare una mossa effettuata (Goel e Grafman, 1995; Murij e DeLuca, 1998), altrimenti la performance potrebbe dipendere più dall'inibizione che dalla pianificazione, poiché solamente i partecipanti che inibiscono la tendenza a muovere la pallina verso la sua posizione finale riescono a pianificare più attentamente le mosse e a risolvere i problemi in modo efficace (Koppenol-Gonzalez, Bouwmeester, e Boonstra, 2010).

La memoria di lavoro, invece, risulta influenzare in modo consistente la fase di esecuzione, in quanto il piano precedentemente studiato deve essere tenuto a mente mentre le palline vengono mosse.

Un aspetto poco investigato riguarda l'utilizzo di strategie per lo svolgimento del test: le persone sottoposte al test potrebbero servirsi di strategie diverse, ad esempio per prove ed errori (trial-and-error) oppure cercando di pianificare l'intero percorso di risoluzione prima di aver effettuato la prima mossa. Le strategie costituiscono un ulteriore elemento in grado di influenzare la performance al test, in quanto alcune strategie specifiche risultano più efficaci di altre, influenzando la durata del tempo di pianificazione e il punteggio finale (Koppenol-Gonzalez, Bouwmeester, e Boonstra, 2010).

Lo stesso studio del 2010 ha individuato e classificato quattro categorie di partecipanti sulla base delle strategie utilizzate: la prima categoria è formata da partecipanti che non adottano una strategia efficace, svolgono la prima mossa troppo in fretta e commettono molti errori; tutto ciò si riflette in un tempo di pianificazione breve e in un punteggio basso, nonché su una bassa influenza delle caratteristiche del problema sulla performance. Il secondo gruppo presenta un'abilità di pianificazione efficace, con il tempo di pianificazione più lungo fra tutte le categorie e con la maggiore influenza delle proprietà strutturali del problema sulla performance. Ciò avviene perché i partecipanti con un tempo di pianificazione consistente hanno più probabilità di notare le proprietà del problema e dunque di esserne influenzati. La terza e la quarta categoria condividono un punteggio medio, ma la quarta si caratterizza per avere un tempo di pianificazione più lungo (contro quello della terza che risulta essere medio) portando ad una maggiore influenza delle proprietà del problema, da cui sono meno distratti rispetto ai partecipanti del terzo gruppo. Ciò si può riflettere in alcune importanti considerazioni sulle varie modalità di calcolo del punteggio, di cui si discuterà più avanti.

Un aspetto da notare riguarda i cosiddetti iso-problems: si tratta di quei problemi che richiedono una sequenza di mosse perfettamente identica per essere risolti, mentre ciò che muta da un esercizio all'altro è il colore delle specifiche palline. L'intero set di problemi della Torre di Londra consiste in 210 classi di sei problemi

ciascuna che sono mutualmente equivalenti: i sei problemi delle singole classi di equivalenza sono definiti iso-problems (Faber et al., 2009). Si tratta dunque di esercizi che, pur presentandosi differenti tra loro all'occhio dell'osservatore, sono caratterizzati dallo stesso livello di difficoltà.

Per quanto riguarda questo tipo di problemi, uno studio ha fatto notare come sia presente un effetto di miglioramento costante della performance dalla somministrazione del primo iso-problem al sesto (ovviamente inframezzati da altre tipologie di problemi) e che questo effetto sia dimostrato da molteplici misure della performance (il tempo totale richiesto per la soluzione del problema, il numero di mosse effettuate, la cosiddetta misura di efficiency e altre ancora). È stato inoltre riportato che solo un terzo dei partecipanti allo studio ha notato la presenza di questi problemi matematicamente equivalenti e, dunque, che la maggior parte dei soggetti abbia appreso le sequenze di mosse degli iso-problem in modo inconsapevole. Questi risultati permettono di parlare di iso-effect, ossia il fenomeno per cui la performance negli iso-problem migliora nel corso della loro somministrazione senza che il partecipante se ne renda conto, presumibilmente grazie a meccanismi di apprendimento non dichiarativi (Faber, 2009).

Un altro studio (Unterrainer et. Al, 2003) si è occupato degli eventuali effetti sistematici di diverse tipologie di istruzioni e dell'apprendimento sulle abilità di pianificazione e dunque sulla performance al test, attraverso cinque livelli di difficoltà (ossia utilizzando problemi dalle tre alle sette mosse minime). In primo luogo, si è notato che i partecipanti a cui era stato esplicitamente richiesto di pianificare anticipatamente le proprie mosse, dunque prima di iniziare con il movimento delle palline, hanno risolto correttamente un maggior numero di tentativi rispetto ai partecipanti a cui ciò non era stato richiesto. In secondo luogo, a metà dei partecipanti è stato fornito un cue, ossia è stato indicato precisamente e prima dell'inizio del test il numero minimo di mosse in cui avrebbero dovuto risolvere l'esercizio, mentre all'altra metà è stata data l'indicazione generica di risolverlo nel minor numero di mosse possibile. I risultati affermano che il primo gruppo, ossia quello fornito di cue, è riuscito a risolvere correttamente un maggior

numero di problemi rispetto al secondo. Per quanto riguarda l'apprendimento, in generale i partecipanti hanno mostrato un miglioramento della performance nella seconda somministrazione rispetto alla prima. Tuttavia, l'aspetto più interessante riguarda l'effetto che la presenza o meno del cue ha sull'effetto apprendimento: si è notato che il miglioramento dal primo al secondo test era significativo solo quando i partecipanti iniziavano dalla condizione senza cue per poi passare alla condizione con cue, ma non si verificava il contrario (chi cominciava con la versione cue manteneva un livello simile nella condizione non cue). Ciò è probabilmente dovuto al fatto che i partecipanti che svolgevano per primo un compito con cue entravano in contatto con una versione più semplice del test, riuscendo a raggiungere buoni risultati sin da subito e dunque non riuscendo a raggiungere risultati significativamente migliori nella più complicata versione priva di cue.

Come gli autori stessi affermano, questi risultati dimostrano che le istruzioni e i cue forniti, nonché l'apprendimento, hanno una forte influenza sulla performance ToL, e contribuiscono inoltre a spiegare parte delle divergenze e incongruenze riscontrabili nelle pubblicazioni riguardanti il test.

2.3. Un possibile set di problemi strutturalmente bilanciato

Una volta descritte le principali caratteristiche che, insieme al numero minimo di mosse, concorrono a determinare la difficoltà del problema, influenzando di conseguenza la performance in modo sistematico, emerge la necessità di realizzare una versione standardizzata del test ToL, che tenga conto dei parametri descritti, in modo tale da poter essere applicata in clinica e in ricerca e da permettere la produzione di dati altamente comparabili tra loro.

Un tentativo in questo senso è stato realizzato da Kaller, Unterrainer e Stahl tra il 2011 e il 2012 (Kaller et al., 2011; Kaller, Unterrainer e Stahl, 2012): il set di problemi in questione è stato realizzato sulla base di un'analisi teoretica comprensiva dello spazio problema della Torre di Londra, ma anche tenendo in considerazione le evidenze empiriche legate all'impatto che la struttura del

problema ha sulla pianificazione. È stato dunque calibrato secondo variazioni sistematiche di vari parametri: in particolare la precedentemente citata goal hierarchy e la search depth (il numero di mosse intermedie necessarie prima che la prima pallina venga sistemata nella sua posizione finale) e il numero di percorsi di soluzione ottimali è stato mantenuto uguale a uno, il tutto spaziando da un numero minimo di mosse di quattro a uno di sette. In linea con le intenzioni dello studio, i risultati hanno mostrato un incremento lineare quasi perfetto della difficoltà del problema all'aumentare del numero di mosse. Secondo questi studiosi, il set di problemi in questione può essere utilizzato non solo per misurare le abilità di pianificazione generali, ma anche per fornire informazioni più dettagliate su particolari aspetti di questa funzione, anche a livello di soggetti individuali. Infine, grazie alle sue buone proprietà psicometriche, può essere utilizzato sia in clinica che in ricerca, con popolazioni anche molto diverse fra loro, permettendo di migliorare in modo significativo la misurazione delle abilità di pianificazione. Gli autori hanno tuttavia incoraggiato un approfondimento del tema, con il fine di un perfezionamento sempre crescente del set di problemi utilizzati.

2.4. La versione computerizzata del test

Come testimoniato da un numero consistente di studi che ne hanno fatto uso, oltre alla classica versione “cartacea” del test della ToL, ossia quella fisicamente formata dai tre bastoncini di dimensione crescente e dalle tre palline, è possibile utilizzare delle versioni computerizzate dello stesso test. Queste possono variare in alcuni dettagli a seconda dello specifico software, ma rispecchiano gli aspetti principali della versione manuale. Un esempio può essere dato dal fatto che in alcune versioni il software è impostato in modo tale da non permettere, in modo automatico, quelle azioni non permesse dal regolamento, ad esempio non sistemare una pallina in uno dei bastoncini una volta selezionata, ma tenerla da parte, oppure muovere più di una pallina per volta.

Vista questa duplice possibilità di somministrazione e visto il fatto che spesso queste due tipologie di strumento vengono utilizzate in modo intercambiabile (McKinlay e McLellan, 2010) ci si può chiedere se l'utilizzo di una versione piuttosto che un'altra possa determinare differenze sistematiche nella performance.

Nonostante la scarsità di lavori realizzati sull'argomento, uno in cui questa domanda viene indagata è stato realizzato da McKinlay e McLellan nel 2010: nello specifico, i due autori si sono chiesti se la modalità di presentazione del test, ossia elettronica oppure manuale, possa influenzare la performance ToL. Lo studio è stato realizzato tenendo conto del fatto che il numero di mosse non costituisce l'unico fattore di difficoltà.

In seguito all'assegnazione randomica dei partecipanti ad una oppure all'altra delle due versioni indagate, gli autori hanno potuto constatare che la modalità di presentazione non influenza la performance in modo significativo, che dunque le due versioni possono essere utilizzate in modo equivalente e che i risultati prodotti possono essere comparati fra loro. Tuttavia, un limite di questo studio riguarda il campione utilizzato, costituito di soli studenti: gli autori ipotizzano che altri gruppi di pazienti o adulti possano risultare maggiormente influenzati dalla modalità di presentazione, in quanto meno familiari con i dispositivi elettronici. Si tratta dunque di una possibilità da approfondire.

3. I METODI DI SCORING

Una questione importantissima, a proposito del test della Torre di Londra, è la scelta del metodo di calcolo del punteggio, ossia di valutazione della performance. Sia in ambito clinico che di ricerca, ne sono stati utilizzati di varie tipologie e in molteplici varianti, spesso senza tenere in considerazione la possibilità che il metodo di scoring utilizzato potesse non essere adeguato allo scopo, non essendo adatto a cogliere il particolare costrutto che si intende indagare, oppure non permettendo la comparazione dei risultati con quelli di altri studi.

3.1. I principali metodi di scoring utilizzati

Un buon punto di partenza, per iniziare a comprendere gli aspetti principali che ci permettono di valutare l'adeguatezza della scelta di un metodo di scoring, è riportare una panoramica delle principali misure esistenti. Una categorizzazione in questo senso è stata realizzata da Berg e Byrd nel 2002: i due studiosi hanno confrontato i vari metodi fino ad allora presenti in letteratura dividendoli in quattro gruppi principali, ossia il "successo o accuratezza della risoluzione", "efficienza della risoluzione", "velocità di performance e di pianificazione" e "regole infrante".

Il successo o accuratezza nella soluzione, altrimenti definito accuratezza, riflette il successo del partecipante nel portare le palline dalla configurazione iniziale a quella finale e si traduce nella proporzione oppure nella percentuale di problemi risolti. Questa tipologia di misura può essere applicata secondo due approcci: il primo basa la definizione di "successo" sul numero di problemi risolti al primo tentativo oppure entro un numero prestabilito di tentativi, il secondo considera un problema come risolto con successo se viene completato entro un determinato limite di tempo, senza porre attenzione al numero di mosse o tentativi in eccesso.

Un vantaggio del primo di questi due approcci consiste nel fatto che contribuisce a mettere l'accento sulla necessità di una pianificazione svolta in modo particolarmente cauto; il secondo approccio, invece, permette uno sguardo sui pattern di soluzione messi in atto dal partecipante, sui suoi errori tipici e sui modi di riparare a questi errori.

L'efficienza della risoluzione corrisponde al numero di tentativi svolti per portare a termine l'esercizio. Si calcola semplicemente contando il numero di mosse fatte per arrivare alla soluzione, oppure come il numero di mosse realizzate in più rispetto al numero indicato per una soluzione ottimale.

Le misure di velocità sono rilevanti in quanto la velocità dei processi cognitivi, tra cui la pianificazione, è ritenuta essere in relazione inversa con la loro accuratezza. Tuttavia, queste misure sono soggette ad una minore precisione nel caso in cui il compito sia somministrato manualmente e il tempo sia calcolato tramite cronometro. Queste imprecisioni possono essere largamente ridotte tramite l'utilizzo di versioni computerizzate del test.

La misura fondamentale di velocità è il *total solution time*, che va dal momento della presentazione del problema a quello in cui la soluzione è raggiunta. È una misura sensibile alla difficoltà del problema, ma contiene diverse componenti temporali che riflettono processi cognitivi diversi: la prima componente è detta *first move time*, che va dalla presentazione del problema al momento in cui è attuata la prima mossa e per questo è altrimenti definita *planning time*. A sua volta si compone di una parte cognitiva, la pianificazione delle mosse, e di una parte motoria, ossia il lasso di tempo necessario allo spostamento della prima pallina: secondo gli autori, tuttavia, è bene non separare le due componenti, in quanto parte della pianificazione potrebbe avvenire comunque dopo che la pallina è già stata afferrata. È importante inoltre notare che nel caso di popolazioni cliniche, bambini o altri gruppi con difficoltà motorie, la porzione di tempo dedicata alla componente motoria potrebbe essere considerevolmente più lunga di quella di un gruppo di controllo adulto.

Una volta controllato il fattore dell'abilità motoria, ci si chiede cosa comunichi questa specifica misura. A proposito del *first move time*, le interpretazioni

riportate sono due: per Levin et al. (1996), nonché Kafer e Hunter (1997), questa misura aumenta all'aumentare della meticolosità con cui il partecipante pianifica le sue mosse, tuttavia Kafer e Hunter suggeriscono anche che potrebbe essere collegata alla difficoltà del singolo partecipante nel formulare un piano, ed essere dunque indice di una pianificazione inefficace.

L'altra controparte del *total solution time* è il *subsequent move time*: in questo caso il tempo dedicato allo spostamento fisico delle palline è tendenzialmente maggiore rispetto alla parte cognitiva rappresentata dalla pianificazione in itinere e dal tempo dedicato a risolvere eventuali errori, e per questo è necessario cercare di separare la componente motoria da quella cognitiva; questa operazione può essere svolta con vari metodi. Un esempio viene fornito da Owen et al. (1990), i quali hanno usato una versione computerizzata del test detta "One Touch Tower of London", capace di misurare le pause che avvenivano in itinere indipendentemente dall'attività di spostamento delle palline, grazie ad un ulteriore compito di controllo motorio associato al test principale. Berg e Byrd (2002), invece, hanno semplicemente cercato di individuare le pause insolitamente lunghe tra una mossa e l'altra, considerandole come un indizio di pianificazione in itinere.

La misura detta "regole infrante", infine, riguarda quei casi in cui un partecipante viene meno a una delle tre regole fondamentali della ToL: non posizionare più palline del numero di palline permesso dalle dimensioni del bastoncino, non posizionare le palline in alcun posto se non nei bastoncini e posizionare una pallina per volta. Questi errori possono verificarsi per due motivi: i partecipanti possono dimenticare la presenza di una determinata regola o scegliere deliberatamente di non seguirla: il primo caso è quello più frequente, specialmente nel caso di pazienti neuropsicologici; tuttavia, con popolazioni non cliniche, sia di adulti che di bambini, può non essere sempre così ed è bene utilizzare delle procedure precedenti al test effettivo che permettano di verificare la corretta assimilazione delle regole da parte del partecipante.

Berg e Byrd riportano inoltre alcune misure meno considerate e utilizzate, ma tendenzialmente in grado di cogliere alcuni aspetti interessanti della

performance. Tra queste ci sono il focus dello sguardo, che può essere rivolto all'apparato fisico della ToL, oppure alla rappresentazione del risultato finale da raggiungere: è stato notato come i partecipanti più abili tendano a guardare quasi esclusivamente il supporto fisico, il che rappresenterebbe la capacità di tenere a mente la *goal position* e di mantenere l'attenzione sull'esercizio.

Un'altra misura riportata dagli autori è quella della distanza dalla soluzione: in quei contesti di somministrazione in cui sono permesse altre soluzioni oltre a quella ottimale, ogni mossa può essere analizzata utilizzando il *problem space* per determinare se ha contribuito ad avvicinare o allontanare il partecipante alla soluzione; nello specifico, va calcolata al completamento di ogni mossa, contando il numero di mosse che intercorrono tra la posizione attuale e quella finale.

Un'ulteriore misura possibile riguarda l'indice di apprendimento, ossia la quantità o il pattern di miglioramento nella performance dal primo all'ultimo problema somministrato. Curve di apprendimento più ripide, ossia processi di apprendimento più veloci, sono presumibilmente correlati ad una maggior capacità di imparare dall'esperienza con la ToL. Può essere misurato in vari modi (ad esempio verificando la riduzione del numero di tentativi necessari a raggiungere la soluzione ottimale), ma è spesso stato tralasciato in quanto il modesto numero di problemi che vengono solitamente somministrati, nonché la loro presentazione in ordine crescente di difficoltà, non permettano di valutare adeguatamente il livello di apprendimento (gli autori sostengono tuttavia che nonostante questi due aspetti sia comunque possibile notare un'evidenza di apprendimento).

L'ultima misura riportata riguarda le mosse incomplete, ossia quei momenti in cui una pallina viene afferrata ma successivamente viene rimessa nello stesso punto in cui già si trovava. Il numero di questi eventi potrebbe essere un indice di impulsività o di povertà di pianificazione, ad esempio nei bambini rispetto agli adulti.

Uno studio di Koppenol et al. (2010) invece, ha verificato quali fossero i metodi solitamente utilizzati in ambito clinico e di ricerca. Nel primo caso, ossia in clinica, specialmente con pazienti neuropsicologici, vengono riportate tre misure: la

prima, la più diffusa, è il cosiddetto total move score, ossia il numero di mosse utilizzate oltre al numero minimo richiesto per completare il compito: in ogni singolo problema, si moltiplica il numero minimo di mosse indicato per quel problema per il numero di mosse in eccesso effettuate dal partecipante. Di conseguenza, un total score basso riflette una buona pianificazione. Una seconda misura è il total correct score, ossia il numero di problemi risolti entro il numero minimo di mosse. L'ultimo gruppo di misure sono quelle di tempo: quello totale, diviso nel tempo di pianificazione e di esecuzione. Dei tre, il più indicativo è ritenuto essere il tempo totale, mentre i suoi sottocomponenti dovrebbero servire ad una sua migliore comprensione.

Per quanto riguarda la ricerca, i metodi utilizzati più frequentemente sono il total move score e il total correct score.

3.2. Utilizzare un punteggio unico o dei punteggi parziali

Una volta dispiegata questa panoramica dei principali metodi che si possono incontrare, è necessario mettere in luce un aspetto indagato da entrambi gli studi appena citati: in entrambi i lavori, nonché in altri, viene esplicitata l'inadeguatezza dell'utilizzo di un unico punteggio complessivo e viene incoraggiato l'utilizzo di varie misure. Mentre in ambito clinico un punteggio unico potrebbe essere sufficiente, in generale è bene considerare tutte le misure possibili in modo attento e selezionarle in base a ciò che si vuole investigare, propendendo per più di una misura, in modo tale da non rischiare di tralasciare informazioni importanti e da rendere la ricerca il più informativa possibile. Utilizzare più misure è inoltre un modo per verificare in modi diversi lo stesso costrutto (Berg e Byrd, 2002).

Un altro motivo per cui utilizzare una misura singola non è un modo adeguato per valutare la pianificazione riguarda le proprietà del problema di cui si è già discusso: i punteggi di tipo totale, quelli unici, non sarebbero ben correlati con il costrutto della pianificazione, in quanto le proprietà del problema come il numero minimo di mosse e la start hierarchy comportano differenze nei singoli esercizi

che non possono essere colti da una sola misura di punteggio (Koppenol et al., 2010).

Al contrario, in uno studio successivo di Berg e Byrd del 2010 si afferma che sia il metodo di punteggio complessivo da loro scelto che il punteggio distribuito su varie misure sono stati in grado di cogliere adeguatamente le variazioni in difficoltà dei singoli problemi e, di conseguenza, a fornire una valutazione ben bilanciata della performance. Tuttavia, si aggiunge anche che il principale vantaggio del punteggio unitario riguarda la sua praticità a livello clinico, quando la Torre di Londra è parte di un'ampia batteria di test e la pianificazione può essere considerata, seppur semplicisticamente, come un costrutto unico; al di fuori di questo contesto, si afferma che la varietà di aspetti che un punteggio di tipo multicomponenziale potrebbe mettere in luce sia molto più utile e interessante per i ricercatori che si occupano di questo test.

3.3. Influenze sistematiche determinate dalla scelta del metodo di scoring

Nella parte finale di questo capitolo, l'obiettivo è quello di riportare alcuni studi che testimoniano come la scelta da parte del ricercatore di una tipologia di misura del punteggio rispetto ad un'altra possa portare a differenze sistematiche nei risultati, essere inadeguata al tipo di popolazione su cui gli studi sono effettuati oppure non permettere la miglior interpretazione possibile dei dati raccolti. Questi studi dovrebbero poter contribuire a creare un quadro, seppur sommario, dei principali aspetti di cui si dovrebbe tener conto nella scelta del metodo di scoring. I risultati del primo di questi studi hanno portato ad affermare che diversi metodi di scoring potrebbero riuscire a cogliere componenti diverse della funzione esecutiva che è la pianificazione: nello specifico, ci si è chiesti che tipologia di pianificazione fossero in grado di cogliere rispettivamente il total number correct score e il first move time score. Ciò è stato realizzato confrontando i risultati al test ToL con quelli di altri due test test, il cosiddetto Crack the Code e il Cognitive Assessment System (CAS), entrambi utilizzati per la pianificazione. Per comprendere meglio questo studio è necessario riportare la cosiddetta "activity

theory”: secondo questa teoria, esistono tre livelli di pianificazione. Il primo è detto “activity”, in cui la pianificazione è definita come un metodo per realizzare il proprio obiettivo, a seconda delle preferenze individuali, ad esempio la tendenza all’accuratezza o alla velocità, all’innovazione o alla ripetizione di pattern già appresi. Questo livello influenza i due livelli successivi: il secondo è detto “action planning”, che equivale al problem solving e il cui obiettivo è realizzare uno scopo specifico, per cui viene impiegato nella fase di pre-planning e anche online per verificare se una mossa sia funzionale all’obiettivo da raggiungere. L’ultimo livello è quello detto “operations”, in cui il piano si trasforma in una vera e propria tattica o strategia e il partecipante cerca di risolvere il problema entro le limitazioni e le regole imposte dall’esercizio. Secondo Das e Georgiou (2016), il test CAS correla con il livello delle operations, mentre Crack the Code sarebbe correlato con l’action planning. Ciò è stato confermato dai risultati, i quali hanno permesso di affermare anche che il total number correct nella ToL correla con entrambi i fattori appena citati (operations e action planning), mentre il first move time correla con il first move time di Crack the Code: entrambi i first move time, infatti, possono essere considerati delle misure di pre-planning.

È interessante puntualizzare quali sono le implicazioni cliniche delle scoperte appena citate: secondo gli autori, infatti, nel caso di pazienti che presentano danni moderati al lobo frontale, la dimensione dell’operations planning potrebbe risultare parzialmente risparmiata e dunque la performance al test ToL ha una certa probabilità di migliorare grazie all’attività di riabilitazione. La dimensione dell’action planning, invece, risulterebbe meno responsiva all’allenamento: tuttavia, come affermato dagli autori stessi, si rende necessaria ulteriore ricerca a proposito di queste affermazioni (Georgiou, Li e Das, 2017).

Esiste uno studio, inoltre (Michalec et al., 2017), che si è preposto di analizzare quattro metodi di scoring differenti, di cui due proposti dallo stesso Shallice nella cornice della versione originale della ToL, e due realizzati da Anderson et al. e Krikorian et al. Un obiettivo secondario dello studio era quello di verificare l’effetto delle variabili demografiche sull’utilizzo di questi metodi: i risultati emersi non sono particolarmente significativi, soprattutto per quanto riguarda l’età, tuttavia i

dati riguardanti il genere e la scolarità hanno permesso agli autori di affermare che la versione specifica del test, i metodi di scoring e la popolazione specifica su cui il test viene effettuato sono tutti fattori che influiscono nell'interpretazione dei dati riguardanti le performance ToL e nella validità cross-culturale di questo strumento.

Un importante studio di Baker et al. (2001) si è invece concentrato sul confronto fra altri due metodi di scoring, utilizzandoli per valutare le performance sia di adulti che di bambini, e giungendo a constatare che la validità di costrutto della ToL cambia a seconda dell'età del campione. È necessario, tuttavia, concentrarsi su un altro aspetto emerso all'interno di questo studio, ossia che nei ragazzi l'uso di diversi metodi di scoring determina differenze significative su quello che possiamo dire a proposito delle tendenze evolutive nelle funzioni esecutive. Innanzi tutto, va specificato che i due metodi in questione sono detti accuracy scoring method, che attribuisce tre, due o un punto a seconda che il problema sia stato risolto entro il primo, secondo o terzo tentativo, senza limiti di tempo, e time-sensitive scoring method, in cui si sottrae il numero di tentativi falliti ad una misura temporale, il tutto entro i 60 secondi.

Per quanto riguarda i risultati dello studio, i due metodi porterebbero a supportare due ipotesi differenti: il metodo accuracy supporta l'ipotesi per cui le funzioni esecutive inizino a svilupparsi a metà dell'infanzia, con un aumento stabile fino alla prima età adulta. Il metodo time-sensitive, invece, suggerisce che la traiettoria di sviluppo delle funzioni esecutive abbia inizio ad un'età più precoce e che raggiunga i livelli tipici dell'età adulta già nelle prime fasi dell'adolescenza. Dunque, la discrepanza presente in letteratura a proposito dello sviluppo delle funzioni esecutive potrebbe essere almeno parzialmente spiegata dall'utilizzo di metodi differenti per il calcolo del punteggio.

CONCLUSIONE

Gli studi riportati forniscono alcune importanti linee da seguire per un utilizzo più consapevole e proficuo del test della Torre di Londra. La prima riguarda la misurazione del livello di difficoltà dei problemi, che va effettuata con la consapevolezza che, oltre al numero minimo di mosse, esiste una serie di caratteristiche del problema, nonché di altri aspetti secondari, in grado di influenzare consistentemente questa misura. La seconda riguarda il metodo di calcolo del punteggio: è necessario verificare che la tipologia di metodo scelto, fra le tante disponibili, sia adeguata al costrutto a cui si è interessati, tenendo sempre in considerazione il fatto che questa scelta potrebbe avere delle conseguenze sistematiche sui risultati raccolti.

È, infine, auspicabile che la ricerca prosegua in una considerazione e valutazione sempre più esaustiva di tutti quegli aspetti, sia legati al livello di difficoltà che al metodo di scoring, in grado di influenzare i risultati del test e la loro interpretazione: a partire da questa maggiore consapevolezza delle caratteristiche del test, e con l'obiettivo di produrre degli studi sempre più comparabili fra loro, si potrà rendere più concreta la possibilità di arrivare a delle versioni del test e a dei protocolli di somministrazione maggiormente standardizzati, ma calibrati consapevolmente sul tipo di popolazione indagata, nonché sul costrutto d'interesse.

BIBLIOGRAFIA

Berg, W. K., & Byrd, D. L. (2002). The Tower of London spatial problem-solving task: Enhancing clinical and research implementation. *Journal of clinical and experimental neuropsychology*, 24(5), 586-604.

Berg, W. K., Byrd, D. L., McNamara, J. P., & Case, K. (2010). Deconstructing the tower: Parameters and predictors of problem difficulty on the Tower of London task. *Brain and Cognition*, 72(3), 472-482.

Carlin, D., Bonerba, J., Phipps, M., Alexander, G., Shapiro, M., & Grafman, J. (2000). Planning impairments in frontal lobe dementia and frontal lobe lesion patients. *Neuropsychologia*, 38(5), 655-665.

Das, J.P., & Georgiou, G. (2016). Levels of planning predict different reading comprehension outcomes. *Learning and Individual Differences*, 48, 24-28.

Etnier, J. L., & Chang, Y. K. (2009). The effect of physical activity on executive function: a brief commentary on definitions, measurement issues, and the current state of the literature. *Journal of Sport and Exercise Psychology*, 31(4), 469-483.

Faber, A. H., Kühnpast, N., Sürer, F., Hinz, A. M., & Danek, A. (2009). The iso-effect: Is there specific learning of Tower of London iso-problems?. *Thinking & Reasoning*, 15(3), 237-249.

Georgiou, G., & Das, J.P. (2016). What component of executive functions contributes to normal and impaired reading comprehension in young adults? *Research in Developmental Disabilities*, 49-50, 118-128.

Georgiou, G. K., Li, J., & Das, J. P. (2017). Tower of London: What level of planning does it measure?. *Psychological Studies*, 62, 261-267.

Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in "planning" functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia*, 33(5), 623-642.

Huizinga, M., Dolan, C. V., & Van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, 44(11), 2017-2036.

Kafer, K. L., & Hunter, M. (1997). On testing the face validity of planning/problem-solving tasks in a normal population. *Journal of the International Neuropsychological Society*, 3(2), 108-119.

Kaller, C. P., Unterrainer, J. M., Rahm, B., & Halsband, U. (2004). The impact of problem structure on planning: Insights from the Tower of London task. *Cognitive Brain Research*, 20(3), 462-472.

Kaller, C. P., Unterrainer, J. M., & Stahl, C. (2012). Assessing planning ability with the Tower of London task: psychometric properties of a structurally balanced problem set. *Psychological Assessment*, 24(1), 46.

Karen, K., Segalowitz Baker, S. J., & Ferlisi, M. C. (2001). The effect of differing scoring methods for the Tower of London task on developmental patterns of performance. *The Clinical Neuropsychologist*, 15(3), 309-313.

Koppenol-Gonzalez, G. V., Bouwmeester, S., & Boonstra, A. M. (2010). Understanding planning ability measured by the Tower of London: An evaluation of its internal structure by latent variable modeling. *Psychological assessment*, 22(4), 923.

Levin, H. S., Fletcher, J. M., Kufera, J. A., Harward, H., Lilly, M. A., Mendelsohn, D., ... & Eisenberg, H. M. (1996). Dimensions of cognition measured by the Tower of London and other cognitive tasks in head-injured children and adolescents. *Developmental Neuropsychology*, 12(1), 17-34.

McKinlay, Audrey & McLellan, Tracey. (2011). Does mode of presentation affect performance on the Tower of London task?. *Clinical Psychologist*. 15.

Michalec, J., Bezdicek, O., Nikolai, T., Harsa, P., Jech, R., Silhan, P., ... & Shallice, T. (2017). A comparative study of Tower of London scoring systems and normative data. *Archives of Clinical Neuropsychology*, 32(3), 328-338.

Miyake, A., Emerson, M. J., & Friedman, N. P. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. In *Seminars in speech and language* (Vol. 21, No. 02, pp. 0169-0183). Copyright© 2000 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel.:+ 1 (212) 584-4663.

Murji, S., & DeLuca, J. W. (1998). Preliminary validity of the cognitive function checklist: Prediction of Tower of London performance. *The Clinical Neuropsychologist*, 12(3), 358-364.

Newman, S. D., & Pittman, G. (2007). The Tower of London: A study of the effect of problem structure on planning. *Journal of Clinical and Experimental Neuropsychology*, 29(3), 333-342.

Newman, S. D., Greco, J. A., & Lee, D. (2009). An fMRI study of the Tower of London: a look at problem structure differences. *Brain research*, 1286, 123-132.

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia*, 28(10), 1021-1034.

Unterrainer, J. M., Rahm, B., Leonhart, R., Ruff, C. C., & Halsband, U. (2003). The Tower of London: The impact of instructions, cueing, and learning on planning abilities. *Cognitive Brain Research*, 17(3), 675-683.

Unterrainer, J. M., Rahm, B., Kaller, C. P., Leonhart, R., Quiske, K., Hoppe-Seyler, K., ... & Halsband, U. (2004). Planning abilities and the Tower of London: is this task measuring a discrete cognitive function?. *Journal of clinical and experimental neuropsychology*, 26(6), 846-856.