



**Istituto nazionale per la valutazione del sistema educativo di
istruzione e di formazione**

WORKING PAPER N. 70/2025

**Valutare i fattori che influenzano la difficoltà degli item in prove standardizzate di
comprensione del testo: rassegna della letteratura.**

Laura Petrillo - Assegnista di ricerca INVALSI

Collana: Working Papers INVALSI

ISSN: 2611 - 5719

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the view and the official policy or position of INVALSI.

*Le opinioni espresse nei lavori sono attribuibili esclusivamente agli autori e non impegnano
in alcun modo la responsabilità dell'Istituto. Nel citare i temi, non è, pertanto, corretto
attribuire le argomentazioni ivi espresse all'INVALSI o ai suoi Vertici*

Abstract

Negli anni i test standardizzati di comprensione del testo hanno assunto un ruolo sempre più rilevante sia nell'ambito di rilevazioni su larga scala, sia in ambito scolastico e clinico. La complessità del costrutto oggetto di indagine ha portato numerosi studiosi a interrogarsi su quali siano i fattori che contribuiscono a influenzare la prestazione di bambini e adolescenti agli item di comprensione del testo. L'obiettivo principale della presente rassegna narrativa è quello di individuare e classificare le caratteristiche dei testi, le proprietà degli item e i fattori individuali dei rispondenti che, interagendo tra loro, modulano la probabilità di risposta corretta alle prove di comprensione del testo. Saranno inoltre descritti e analizzati i principali strumenti di misurazione automatica della complessità linguistica e i modelli statistici utilizzati per prevedere la difficoltà degli item. Lo studio si inserisce nel progetto *Linguistic Complexity Evaluation in Education* (LuCET) e adotta come cornice teorica il modello RAND (Snow, 2002), che considera la comprensione del testo come il risultato dinamico dell'interazione tra lettore, testo e compito.

Parole chiave: comprensione del testo, difficoltà degli item, difficoltà del testo, modelli esplicativi di risposta all'item, età scolare

Introduzione

La presente rassegna della letteratura si propone di approfondire gli studi sui fattori che possono influenzare la difficoltà di *processing* del testo scritto in studenti della scuola primaria e secondaria, con particolare riferimento ai test standardizzati di comprensione. La rassegna si contestualizza all'interno del progetto *Linguistic Complexity Evaluation in education* (LuCET; Agrusti, Cardinaletti, Desimoni, Pallotti, & Montemagni, 2024), Progetto di ricerca di Rilevante Interesse Nazionale (PRIN) Finanziato dall'Unione Europea-*Next Generation EU* (CUP del Progetto: F53D23005210006; Codice del Progetto: 2022KPNY3B; Bando PRIN 2022).

I test standardizzati per la valutazione della comprensione nella lettura tipicamente richiedono di leggere più testi stimolo e di rispondere a item associati a ciascun testo, al fine di rilevare il costrutto oggetto di indagine, come definito dal quadro teorico di riferimento del test. La complessità insita nel processo di comprensione di un testo ha portato, negli anni, numerosi autori a interrogarsi su quali fattori contribuissero a spiegare la variabilità nella prestazione degli studenti agli item dei test di comprensione.

Come evidenziato da numerosi autori nel corso degli anni (per es., Sheehan & Mislevy, 1990; Lumley, Routitsky, Mendelovits & Ramalingam, 2012; Ferrara, Steedle & Frantz, 2022), lo studio dei fattori che predicano la difficoltà degli item di comprensione nella lettura, basato su modelli psicolinguistici che descrivono il processo attivo di costruzione del significato di un testo, nonché dai risultati della ricerca sperimentale sull'argomento, può svolgere un ruolo chiave sia nello studio della validità delle misure, sia per la costruzione di nuovi strumenti. È, inoltre, importante per coloro che sono interessati all'interpretazione degli esiti dei test stessi, in quanto può promuovere la consapevolezza dei possibili ostacoli linguistici e cognitivi alla comprensione del testo, anche al fine di approntare percorsi di potenziamento mirati.

Si deve, inoltre, considerare che le prove standardizzate di comprensione del testo hanno un ruolo fondamentale non solo nei percorsi di valutazione delle abilità di lettura che possono essere realizzati in ambito scolastico o clinico, ma anche nell'ambito di rilevazioni standardizzate su larga scala nazionali, come per esempio le prove INVALSI in Italia o il *National Assessment of Educational Progress* (NAEP) negli Stati Uniti, e in quelle comparative internazionali, quali il *Progress in International Reading Literacy Study* (PIRLS) promosso dalla IEA e il *Programme for International Student Assessment* (PISA) dell'OCSE. Lo studio dei fattori che influenzano la difficoltà delle domande può concorrere, in tali contesti, non solo al miglioramento nel processo di sviluppo e validazione degli strumenti di rilevazione stessi, ma anche ad

aumentare le nostre conoscenze sulle qualità dei sistemi di istruzione, nonché a fornire un quadro sempre più completo delle competenze degli studenti.

Il primo obiettivo della presente rassegna consiste nell'analizzare gli studi riguardanti i fattori che influenzano la difficoltà degli item di prove standardizzate di comprensione del testo; un'esplorazione dei fattori cognitivi, linguistici e contestuali che ne determinano la difficoltà, come la complessità linguistica dei testi, le caratteristiche delle domande e la relazione tra le caratteristiche individuali degli studenti e la loro prestazione nelle prove di comprensione. Saranno inoltre esaminati i principali modelli di analisi statistica utilizzati in letterature per esaminare tali relazioni, in un'ottica di integrazione tra modelli psicometrici e cognitivi.

Il modello teorico di riferimento

Il modello euristico di RAND (Snow, 2002), sviluppato da Catherine Snow e il RAND *Reading Study Group* (RRSG), è il quadro di riferimento di numerosi studi che si sono preposti l'obiettivo di studiare quali caratteristiche del testo, degli item e del rispondente influenzano la difficoltà degli item di test standardizzati di comprensione del testo (e.g., Rahman & Mislevy, 2017; Toyama, 2019; 2021; Wang, 2021).

Il RRSG definisce la comprensione del testo come

*“the process of simultaneously
extracting and constructing meaning through interaction and involvement with written language”*
(Snow, 2002, pag.13)

Questa definizione sottolinea che la comprensione è un processo attivo in cui il lettore interagisce con il testo per costruirne il significato (Snow, 2002). Caratteristica chiave del modello è la concettualizzazione della comprensione nella lettura come interazione tra tre elementi principali, che agiscono all'interno di un contesto più ampio.

In primo luogo, il lettore, agente attivo nel processo di ricostruzione del significato di un testo. Alcune caratteristiche del lettore possono influenzare la sua capacità di capire un testo scritto. Esse includono, per esempio, le abilità cognitive, la motivazione, l'autoefficacia percepita, le conoscenze e le esperienze pregresse. Nel modello, è specificato che tali caratteristiche possono variare sia tra i lettori (differenze tra gli individui), nonché per lo stesso lettore in funzione delle caratteristiche del testo o della specifica attività di lettura (differenze entro gli individui).

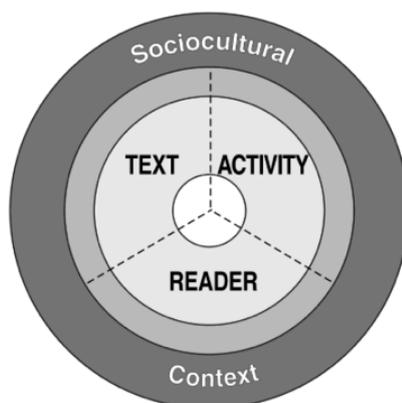
Il secondo elemento del modello è rappresentato dal testo stesso, concepito come lo stimolo linguistico oggetto di elaborazione. Il lettore, secondo il modello, ricostruisce rappresentazioni mentali del testo a più livelli. La

facilità o difficoltà di un testo può dipendere da più fattori inerenti al testo stesso (per esempio, il lessico utilizzato, la complessità sintattica, il contenuto, il genere, etc.), oppure dall'interazione tra caratteristiche del testo e conoscenze pregresse e abilità del lettore, nonché dalle attività in cui il lettore è coinvolto.

Il terzo elemento riguarda l'attività di lettura, che comprende più elementi interconnessi, quali lo scopo della lettura o il compito che il lettore deve svolgere, i processi di elaborazione del testo e i risultati stessi dell'attività. Lo scopo della lettura può variare durante il processo di lettura, anche alla luce dell'interazione in essere tra lettore e testo. I processi di elaborazione di un testo variano in funzione delle capacità del lettore e delle caratteristiche del testo. Inoltre, i diversi aspetti di elaborazione del testo variano in funzione del tipo di lettura, che può spaziare da una lettura ricreativa ad una lettura analitica volta all'estrazione di informazioni specifiche. La tipologia di attività influenza le strategie cognitive adottate dal lettore, modulando la direzione delle risorse attentive e determinando il tipo di elaborazione del testo.

È fondamentale sottolineare che questi tre elementi interagiscono in maniera dinamica e interdipendente all'interno di un contesto socioculturale più ampio, che influenza ed è influenzato a sua volta dagli altri elementi del modello (Snow, 2002).

Sulla base di tale quadro teorico di riferimento, si evidenzia come, nell'analizzare i fattori che possono influenzare la prestazione degli studenti a prove standardizzate di comprensione del testo, sia importante adottare un approccio multidimensionale, in cui si tenga conto dell'interazione tra fattori relativi allo studente, fattori riconducibili al testo stimolo della prova stessa, nonché al tipo di compito sotteso agli item della prova (e.g., Toyama, 2019; Ferrara et al., 2022). Inoltre, potrebbe essere anche importante considerare fattori contestuali, quale il contesto socioculturale, in un'ottica multilivello.



*Figura 1 - A Heuristic for Thinking About Reading Comprehension
Tratto da: Snow, 2002*

Le caratteristiche del testo

In letteratura, numerosi studi si sono prefissati di approfondire il ruolo delle caratteristiche del testo stimolo sulla prestazione dei rispondenti nelle prove di comprensione nella lettura. Il testo si caratterizza per proprietà intrinseche quali la complessità lessicale e sintattica, rilevabili attraverso indici di frequenza e sofisticazione del vocabolario, nonché attraverso l'articolazione e profondità delle strutture frasali. La coesione testuale, ovvero l'insieme dei legami linguistici superficiali che connettono le parti del testo, e la coerenza, intesa come l'unità semantica e logica sottostante, insieme alla struttura organizzativa e al genere testuale (espositivo, narrativo, argomentativo, ecc.), definiscono il quadro formale e contenutistico che il lettore deve decodificare e integrare per costruire il significato (McNamara, Crossley & McCarthy, 2009).

Gli studi che si sono focalizzati sulle caratteristiche del testo hanno preso in considerazione diversi aspetti, rilevati con strumenti diversi. Alcuni autori hanno utilizzato tassonomie (e.g., generi testuali) oppure rilevato caratteristiche dei testi stimolo tramite strumenti qualitativi, per esempio tramite griglie (Freedle & Kostin, 1992; Valencia, Wixson, Ackerman & Sanders, 2017). In alcuni casi, tali caratteristiche fanno riferimento solo al testo (Freedle & Kostin, 1992), in altri, all'interazione tra testo e domande (e.g., Valencia et al., 2017). Altri studi hanno utilizzato caratteristiche linguistiche rilevate attraverso strumenti di analisi automatizzata (e.g., Sano, 2016; Toyama, 2019), come Coh-Metrix (Graesser, McNamara, & Kulikowich, 2011) e TextEvaluator (Sheehan, Kostin, Napolitano, & Flor, 2014).

Tra le classificazioni dei testi più utilizzate nella letteratura sulla relazione tra caratteristiche del testo e abilità di comprensione, emerge il genere testuale. Tipologie del testo diverse richiedono strategie cognitive altrettanto differenti e che ne potrebbero influenzare la comprensione. Alcuni studi hanno dimostrato che i testi narrativi risultano meno difficili rispetto ai testi espositivi (Best, Floyd, McNamara, 2008; McNamara, Ozuru & Floyd, 2011). I testi narrativi e i testi espositivi si distinguono non solo per le loro caratteristiche formali, ma anche per gli argomenti trattati e gli scopi comunicativi (Wang, 2021). Dal punto di vista delle caratteristiche testuali, i testi narrativi tendono a presentare parole ad alta frequenza, frasi più complesse e una coesione relativamente bassa, mentre i testi espositivi sono caratterizzati da un vocabolario più difficile, frasi meno complesse e una maggiore coesione (McNamara, Graesser & Louwerse, 2012). Per quanto riguarda lo scopo, i testi narrativi generalmente raccontano esperienze personali o esprimono emozioni attraverso un linguaggio più familiare, mentre i testi espositivi hanno l'obiettivo di spiegare concetti o fornire nuove informazioni, spesso utilizzando una terminologia più tecnica (Medina & Pilonieta, 2006).



Tra le caratteristiche del testo prese in esame e rilevate (anche) attraverso strumenti qualitativi, tra le altre, emergono la struttura del testo, il livello di concretezza o astrattezza delle informazioni presentate, il numero di informazioni nuove introdotte in un testo, la presenza di caratteristiche letterarie e retoriche specifiche. Testi con strutture più complesse o meno familiari rendono più difficile per il lettore seguire il filo del discorso (Graesser et al., 2011). I contenuti concreti e specifici facilitano la comprensione, mentre quelli astratti tendono ad aumentare la difficoltà di elaborazione (Freedle & Kostin, 1992). La quantità di informazioni uniche o nuove contenute in un passaggio influenza anch'essa la complessità: testi che introducono molte informazioni nuove richiedono un maggiore carico cognitivo al lettore (Freedle & Kostin, 1992). Inoltre, la letteratura ha evidenziato come l'impiego di frasi che posticipano il nucleo informativo della frase è associato ad un aumento dell'incertezza e della difficoltà di elaborazione (Freedle & Kostin, 1992). Altri autori si sono concentrati sull'interazione tra testo e domande: Valencia et al. (2017), per esempio, hanno mappato in che misura la difficoltà delle domande è associata al grado in cui la struttura (genere, organizzazione), il formato, il linguaggio e lo stile di un testo sono fattori chiave per rispondere alle domande stesse.

Nella letteratura sul ruolo delle caratteristiche del testo rispetto alla difficoltà delle domande a esso associate, emerge un crescente numero di studi che ha preso in esame caratteristiche del testo estratte con diversi strumenti di analisi testuale. Tali strumenti forniscono indici linguistici a livello di parola, frase e paragrafo e/o intero testo. Dal 1920 sono stati sviluppati diversi strumenti quantitativi per valutare la leggibilità e la complessità di un testo (Klare, 1984): come evidenziato da Toyama (2019) in letteratura si possono distinguere tre generazioni di strumenti.

Gli strumenti di prima e seconda generazione si basano su misurazioni quantitative del contenuto del testo da analizzare, quali il numero di sillabe per parola, il numero di parole per frase, e il numero di frasi per periodo, per calcolare indici di leggibilità. Gli strumenti di prima generazione erano basati su calcolo manuale o tramite tabelle di conversione. Tra i più noti per la lingua inglese c'è il "Flesch-Kincaid Grade Level" (Kincaid, Fishburne, Rogers, & Chissom, 1975), che utilizza la lunghezza delle parole e delle frasi per determinare il livello di istruzione necessario per comprendere un testo. Gli strumenti di seconda generazione, con l'aumento della potenza di calcolo, hanno automatizzato l'analisi e permesso di valutare grandi corpora di testi. Tuttavia, hanno continuato a concentrarsi su parole e frasi, come i loro predecessori. Tra i più diffusi in ambito internazionale ci sono il "Lexile Framework for Reading" (Schnick & Knickelbine, 2007) e il "Degrees of Reading Power" (DRP, Koslin, Zeno, & Koslin, 1987). Un indice di leggibilità tarato sulla lingua italiana e utilizzato per lo studio della relazione tra leggibilità di un testo e comprensione (e.g., Tarchi, 2012) è il "Gulpease" (Lucisano e Piemontese, 1988), elaborato dal Gruppo universitario linguistico pedagogico (GULP). L'indice ha un valore che oscilla tra 0 e 100, dove il valore "100" indica la leggibilità più alta. Può



essere interpretato attraverso una scala che mette in relazione i valori restituiti dalla formula con il grado di scolarizzazione del lettore.

Negli ultimi anni sono emersi strumenti di analisi linguistico-computazionale avanzati, di terza generazione, che non fanno affidamento unicamente su caratteristiche generali e formali del testo, quali la lunghezza della frase e la lunghezza delle parole, ma si basano su teorie multidimensionali, multilivello e interattive per l'analisi del linguaggio, del discorso e della comprensione nella lettura (Toyama, 2019), maggiormente in coerenza con il modello RAND (Snow, 2002). Tra i più noti in ambito internazionale ci sono “Coh-Metrix” (Graesser et al., 2011) e “TextEvaluator” (Sheehan et al, 2014). Coh-Metrix (Graesser et al., 2011) è uno strumento che utilizza il *natural language processing* per l'analisi del testo rispetto a più caratteristiche linguistiche, in coerenza con un modello multilivello della comprensione (Graesser et al. , 2011). Il programma estrae un numero elevato di caratteristiche linguistiche, sintetizzate in otto indici ottenuti empiricamente attraverso l'Analisi delle Componenti Principali (ACP) su un corpora di testi. Tali indici sono la narratività, la coesione referenziale, la semplicità sintattica, la concretezza delle parole, la coesione causale, la coesione verbale, la coesione logica e la coesione temporale. È ipotizzato che tali componenti influenzino la facilità di comprensione di un testo (McNamara, Graesser, McCarthy, & Cai, 2014; Graesser et al., 2011). Per esempio, si ipotizza che testi con un maggior grado di narratività siano più facili da comprendere. Per quanto riguarda la semplicità sintattica, l'ipotesi è che frasi brevi e strutture semplici facilitano la lettura; mentre per quanto riguarda la concretezza, si ipotizza che parole evocative e tangibili rendano il testo più chiaro. TextEvaluator (Sheehan et al, 2014) è uno strumento di analisi testuale, simile a Coh-Metrix, basato su teorie che considerano la lettura come un processo attivo in cui i lettori costruiscono rappresentazioni mentali coerenti del testo. I lettori, dunque, non si limitano a decodificare parole e frasi, ma cercano attivamente di comprendere il significato globale del testo, collegando le informazioni tra loro e integrandole con le proprie conoscenze ed esperienze (Alderson, 2000; Gernsbacher, 1990; Just & Carpenter, 1987). A differenza delle formule tradizionali di leggibilità, TextEvaluator valuta la complessità di un testo attraverso otto dimensioni linguistiche che offrono un quadro più ricco e articolato. Anche nel caso di TextEvaluator, così come in Coh-Metrix, le dimensioni sono state estratte attraverso l'ACP. Esse sono relative al vocabolario accademico, alla complessità sintattica, alla concretezza, alla non familiarità delle parole, allo stile conversazionale/interattivo, al grado di narratività, alla coesione e all'argomentazione. Per esempio, il vocabolario accademico è un indice di quanto il linguaggio si avvicini a quello tipico dei testi scolastici o scientifici, distinguendolo da quello narrativo o informale. La coesione lessicale viene valutata attraverso la ripetizione di parole chiave e l'uso di connettivi, elementi che rendono il testo più coeso e interpretabile come un messaggio coerente. Una caratteristica distintiva di TextEvaluator è la sua capacità di ridurre il *bias* legato

al genere testuale, che porta solitamente a sovrastimare la difficoltà dei testi informativi e a sottostimare quella dei testi narrativi. Per questo, il sistema utilizza tre modelli predittivi distinti per testi narrativi, informativi e misti (Sheehan et al., 2014; Sheehan, 2016). Per quanto riguarda la lingua italiana, possiamo annoverare, per esempio, READ-IT (Dell'Orletta, Montemagni, Venturi, 2011). Tale strumento è basato su una combinazione di tratti linguistici che spaziano tra diversi livelli di descrizione linguistica, lessicale, morfo-sintattico e sintattico, e consente una valutazione della leggibilità articolata su due livelli, il documento e la singola frase

Tra le caratteristiche quantitative che influenzano la difficoltà nella comprensione del testo, la lunghezza delle frasi emerge come un fattore fondamentale. Diverse ricerche sulla lingua inglese hanno evidenziato che testi con frasi più brevi risultano generalmente più facili da comprendere, mentre frasi lunghe e complesse aumentano il carico cognitivo richiesto al lettore, ostacolando la costruzione di significati coerenti (e.g., Toyama, 2019). Un altro aspetto fondamentale è rappresentato dalla frequenza d'uso delle parole. Lessici composti da parole di alta frequenza, cioè termini più comuni e familiari, favoriscono una lettura più fluida e immediata, mentre l'impiego di vocaboli rari o tecnici tende ad aumentare la difficoltà del testo, specialmente per i lettori con un repertorio linguistico meno ampio (Crossley, Greenfield & McNamara, 2008). La complessità sintattica costituisce inoltre una variabile rilevante: strutture sintattiche articolate, caratterizzate da subordinate annidate o inversioni nell'ordine canonico degli elementi della frase, richiedono al lettore operazioni più complesse e sono associate a maggiori difficoltà di comprensione (Graesser, et al., 2011). Anche la coesione testuale, intesa come l'insieme dei legami linguistici che collegano tra loro frasi e paragrafi, gioca un ruolo significativo. Testi con un alto livello di coesione, sia a livello superficiale (connettivi logici, anafore) sia a livello profondo (relazioni causali e temporali esplicite), risultano generalmente più facili da comprendere poiché facilitano la costruzione di rappresentazioni mentali integrate (Kulesz, Francis, Barnes & Fletcher, 2016). Alcuni studi (McNamara, Louwerse, McCarthy, & Graesser, 2010) dimostrano che aumentare la coesione di un testo migliora la comprensione più efficacemente rispetto alla semplice riduzione della lunghezza delle frasi o all'uso di parole più comuni. Britton e Gülgöz (1991) hanno evidenziato che gli studenti comprendono meglio i testi migliorati in termini di coesione rispetto a quelli rivisti secondo le tradizionali formule di leggibilità. L'aumento della coesione spesso porta a testi più leggibili perché introduce connettivi e riferimenti espliciti, riducendo la necessità di inferenze basate sulla conoscenza pregressa (McNamara et al., 2014). Infine, la densità lessicale, ossia il rapporto tra parole di contenuto (nomi, verbi, aggettivi) e parole funzionali (articoli, congiunzioni, preposizioni), è un ulteriore indicatore di complessità. Un'alta densità lessicale implica una maggiore concentrazione di informazioni, il che può incrementare la difficoltà per il lettore nel mantenere ed integrare le informazioni durante la lettura (Uccelli, Dobbs, & Scott, 2013).

Le caratteristiche delle domande

Data la complessità del costrutto in esame, le prove di comprensione del testo tipicamente contengono item che fanno riferimento a più aspetti di tale competenza, per garantire un'adeguata validità di contenuto. Per esempio, come descritto nel quadro teorico di riferimento (Mullis & Martin, 2015; 2019), gli item delle indagini PIRLS ed ePIRLS si propongono di valutare quattro processi cognitivi della *reading literacy*, ossia ricavare informazioni e concetti esplicitamente espressi nel testo, fare inferenze semplici, interpretare ed integrare informazioni e concetti, analizzare e valutare il contenuto, la lingua e gli elementi testuali. Nell'indagine PISA (per PISA 2018, vedi OECD, 2019), rivolta a quindicenni, sono individuate tre macrocategorie di processi cognitivi (oltre alla fluenza nella lettura, preconditione funzionale agli altri tre processi). La prima fa riferimento alle attività di localizzare le informazioni all'interno di un testo o selezionare il testo rilevante tra più testi (o una parte del testo tra più parti). La seconda categoria è relativa ai processi di comprensione del significato, con riferimento sia all'acquisizione della rappresentazione letterale del significato di frasi o parti del testo, sia alla ricostruzione del significato del testo attraverso le inferenze. La terza macrocategoria concerne il valutare e riflettere sulla forma e contenuto del testo; essa include i processi di valutazione della qualità e credibilità di un testo, di riflessione sulla forma e contenuto del testo, nonché l'individuazione e la gestione delle possibili contraddizioni tra testi o parti di testo.

Con riferimento al modello RAND (Snow, 2002), è possibile ipotizzare che il tipo di processo cognitivo sotteso agli item sia una delle variabili che ne determinano la difficoltà. Tale aspetto è stato approfondito da alcuni studi condotti su dati di rilevazioni su larga scala nazionali o internazionali (per es. Kirsch & Mosenthal, 1990; Sheehan & Mislevy, 1990; Lumley et al., 2012; Valencia et al., 2017; Bulut, Bulut & Arıkan, 2022; Yousefpoori-Naeim, Bulut & Tan, 2023). Per esempio, Bulut e colleghi (2022) in un recente studio su dati ePIRLS relativi agli studenti di quarta primaria che hanno svolto la prova digitale in lingua inglese, hanno evidenziato che gli item che richiedono di analizzare e valutare criticamente un testo sono più difficili rispetto a quelli che richiedono di ricavare informazioni e concetti esplicitamente espressi nel testo, fare inferenze semplici, interpretare o integrare informazioni e concetti. Tale risultato è coerente con quanto anticipato nel quadro teorico di riferimento di ePIRLS 2016 (Mullis & Martin, 2015). Inoltre, lo studio ha evidenziato un'interazione significativa tra processo cognitivo sotteso all'item e caratteristiche degli studenti. In particolare, la differenza tra bambini e bambine, a favore di quest'ultime, è più ampia sia per gli item che richiedono di fare semplici inferenze sia per quelli che richiedono di focalizzarsi su e recuperare informazioni date nel testo. Invece, il *gap* tra studenti che parlano a casa la lingua del test e gli studenti che parlano

prevalentemente un'altra lingua è più ampia per gli item che richiedono di analizzare e valutare criticamente un testo, rispetto alle altre categorie di item. Un effetto significativo del processo cognitivo sotteso all'item è stato anche evidenziato in studio condotto da Yousefpoori-Naeim, Bulut e Tan (2023) sui dati PISA 2018 relativi al Canada. In particolare, gli autori hanno evidenziato che, per i quindicenni, gli item che afferiscono alla categoria 'riflettere e valutare' sono più difficili rispetto a quelli che richiedono il recupero di informazioni date nel testo, e che tale effetto è leggermente più forte per gli studenti che parlano a casa la lingua del test. Altri dati sull'importanza del processo cognitivo sotteso all'item nel determinarne la difficoltà derivano da studi che non si basano su analisi di dati raccolti in indagini su larga scala, ma su test standardizzati di comprensione in commercio, utilizzati tipicamente in ambito scolastico o clinico. Per esempio, Wang (2021), in uno studio sui fattori che influenzano la difficoltà degli item di due test standardizzati, il *Gates-MacGinitie Reading Test-4° edizione* (GMRT-4; MacGinitie et al., 2000) e la scala di comprensione del testo della *Wechsler Individual Achievement Test- 3° Edizione* (WIAT-III; Wechsler, 2009) ha classificato le domande in letterali e inferenziali. Nella classificazione utilizzata dall'autore, le prime richiedono il reperimento diretto di informazioni esplicitamente enunciate, mentre le seconde richiedono la capacità di integrare informazioni testuali con inferenze basate su conoscenze pregresse. I risultati dello studio, condotto su bambini statunitensi, hanno mostrato che le domande letterali sono meno difficili di quelle inferenziali per i testi con bassa coesione testuale, mentre un pattern opposto si osserva per i testi con altra coesione testuale. Toyama (2019), nelle sue analisi di secondo livello sui dati del *ReadingPlus InSight Assessment* raccolti su studenti degli Stati Uniti e del Canada, ha testato una serie di modelli per la previsione della difficoltà degli item in cui sono incluse numerose variabili a livello di testo, rispondente e item. Tra esse, anche il tipo di elaborazione cognitiva richiesta dalla domanda. L'autore ha classificato ex-post gli item in quattro categorie, con uno schema di codifica basato sul modello della comprensione del testo di Kintch (1988). La prima categoria include gli item che richiedono un recupero letterale di informazioni date in singole frasi del testo; la seconda agli item che richiedono l'elaborazione/ristrutturazione di informazioni reperibili in frasi contigue entro lo stesso paragrafo. La terza e quarta categoria di item implicano un maggior livello di elaborazione: richiedono infatti allo studente di integrare parti distanti del testo attraverso inferenze ponte, oppure di integrare, sempre attraverso processi inferenziali, il testo con la propria conoscenza pregressa. Lo studio ha evidenziato che, per il test in esame, le domande che richiedono il recupero letterale delle informazioni del testo sono più complesse rispetto alle domande inferenziali. Tale risultato è in contrasto con la letteratura sull'argomento. Tuttavia, come evidenziato dall'autore stesso, questo potrebbe essere legato dalle caratteristiche del test in esame, in cui allo studente non è data la possibilità di tornare al testo, con un maggior carico sulla memoria delle domande di recupero delle informazioni.

Oltre al processo cognitivo sotteso all'item, numerose sono le caratteristiche degli item presi in esame della letteratura sull'argomento (Ferrara et al., 2022). Per esempio, Lumley, Routitsky, Mendelovits e Ramalingam (2012), in uno studio sull'indagine PISA hanno evidenziato come il grado di astrazione delle informazioni richieste per rispondere alle domande, il grado di similarità tra risposta corretta e altre informazioni concorrenti e il livello di familiarità (per lo studente) delle informazioni richieste sono tra le variabili maggiormente predittive del livello di difficoltà degli item, insieme ad altri fattori relativi ai processi cognitivi sottesi agli item, quale il tipo di compito e la necessità di integrare le informazioni presenti nel testo con conoscenze esterne a esso.

Le caratteristiche degli item prese in esame nel lavoro di Lumley e colleghi (2012) e da altri studi sull'argomento possono essere classificate in variabili che riguardano la relazione tra testo e domande e variabili che fanno riferimento a caratteristiche relative alla domanda stessa (Ferrara et al., 2022).

Tra le prime, per esempio, un ruolo particolarmente rilevante è assunto dal tipo di informazione del testo che lo studente deve elaborare per rispondere correttamente all'item, come il grado di astrazione o la necessità di elaborare figure retoriche, metafore, giochi di parole o tratti specifici del linguaggio letterario. Il dato sulla rilevanza del grado di astrazione è confermato da più studi sull'argomento (per es. Freedle & Kostin, 1992; Kirsch & Mosenthal, 1990; Lumley et al., 2012; Valencia et al., 2017; Toyama, 2019). Le domande che si riferiscono a informazioni concrete (come eventi, azioni, personaggi, luoghi) risultano generalmente più semplici rispetto a quelle che richiedono l'identificazione di concetti astratti, come motivazioni, implicazioni, relazioni causali o logiche. Valencia et al. (2017) hanno, inoltre, evidenziato il ruolo dell'uso di caratteristiche retoriche e stilistiche del testo nella formulazione dell'item. Alcune domande, infatti, richiedono ai rispondenti di elaborare figure retoriche, metafore, giochi di parole o tratti specifici del linguaggio letterario. Questo aspetto può aumentare significativamente la difficoltà, specialmente quando lo stile del testo è particolarmente elaborato o poco frequente nella lingua d'uso quotidiano.

Un'altra variabile indagata rispetto alla relazione tra item e testo è il grado di sovrapposizione tra le parole, le frasi e i concetti contenuti nella domanda e quelli presenti nel testo. In generale, una maggiore corrispondenza (es. domande che riprendono testualmente porzioni del testo) è associata ad una minore difficoltà, perché permette al lettore di risalire facilmente all'informazione richiesta. Al contrario, una bassa corrispondenza (es. parafrasi, riformulazioni o sinonimie) implica una maggiore richiesta inferenziale. Questo è stato confermato in vari studi (Sheehan & Mislevy, 1990; Gorin & Svetina, 2008; Sano, 2016) che mostrano come gli item più impegnativi siano quelli che richiedono un'elevata elaborazione semantica per collegare la domanda al contenuto del testo. Inoltre, la quantità di contenuto informativo di un testo o di una domanda è stata individuata come un elemento importante: un maggior numero di idee specifiche, categorie informative o

contenuti unici rende l'elaborazione più complessa (Drum, et al. 1981). Inoltre, anche la posizione dell'informazione rilevante all'interno del testo può influire sulla difficoltà dell'item. Secondo Freedle & Kostin (1992), le domande che richiedono informazioni situate all'inizio del brano (come la prima frase del primo o del secondo paragrafo) sono tendenzialmente più semplici. Informazioni distribuite o collocate in posizioni meno salienti richiedono invece un maggiore sforzo di recupero.

Altre variabili prese in esame in letteratura fanno riferimento a vari parametri che quantificano la complessità linguistica delle domande stesse. Secondo Sano (2016), la difficoltà di uno stem può essere stimata attraverso vari parametri, come la frequenza media dei lemmi, la varietà delle parti del discorso presenti e la complessità sintattica. In particolare, una minore frequenza delle parole usate e una maggiore variabilità grammaticale sono associate a un aumento della difficoltà, perché richiedono al rispondente competenze linguistiche più raffinate per decodificare la richiesta. In linea con tale risultato, Toyama (2019) ha evidenziato che la leggibilità della domanda è uno tra i fattori che contribuiscono a predire la probabilità di risposta corretta.

Le domande dei test standardizzati di comprensione, inoltre, spesso si differenziano per formato (per es., a scelta multipla o a risposta aperta), variabile che si ipotizza possa influenzare la difficoltà delle domande stesse. Recenti studi indicano che l'effetto del formato della domanda varia in funzione di alcune caratteristiche dei rispondenti (e.g., Valencia et al., 2017; Bulut et al., 2022; Yousefpoori-Naeim et al., 2023). Per esempio, Bulut e colleghi (2022) hanno osservato che le domande a risposta aperta sono più difficili rispetto a quelle a scelta multipla, e che tale differenza è più ampia per i bambini che parlano a casa una lingua diversa da quella del test, rispetto ai loro pari. Non è emerso nello studio, invece, un'interazione significativa tra formato delle domande e genere degli studenti. Tale interazione, invece, è stata riscontrata su adolescenti canadesi di quindici anni da Yousefpoori-Naeim e colleghi (2023) nel loro studio sui dati PISA 2018: l'effetto del formato è più ampio per i ragazzi rispetto alle ragazze, con uno svantaggio maggiore per gli item a risposta aperta.

Infine, altri studi hanno esaminato fattori specifici relativi alle domande a scelta multipla. L'opzione corretta in una domanda a scelta multipla può presentare caratteristiche linguistiche che ne influenzano la difficoltà. Ad esempio, studi come quelli di Drum, Grover e Flake (1981) e Sano (2016) hanno evidenziato che la presenza nella risposta corretta di parole rare oppure di parole non presenti nel testo, può rappresentare un ostacolo per il lettore. La difficoltà aumenta quando la risposta corretta contiene un vocabolario poco frequente, o che richiede un'elevata familiarità con termini accademici o tecnici. Un altro dato ricorrente in letteratura riguarda l'effetto della posizione dell'opzione corretta. Infatti, Freedle & Kostin (1992) hanno osservato che le risposte corrette collocate verso la fine dell'elenco tendono ad essere selezionate più

facilmente: questo effetto può incidere sulla probabilità di rispondere correttamente alla domanda. Inoltre, i distrattori giocano un ruolo importante nella determinazione della difficoltà di un item. Quando questi sono formulati in modo plausibile e hanno una forte somiglianza lessicale o semantica con il contenuto del testo, l'item diventa più difficile. Alcuni studi (Drum et al., 1981; Mosenthal, 1996; Gorin & Svetina, 2008) hanno utilizzato indicatori come la similarità semantica per quantificare la plausibilità dei distrattori: distrattori troppo simili all'opzione corretta aumentano l'ambiguità della scelta e la probabilità di errore.

Le caratteristiche dei rispondenti

Dopo aver esaminato le caratteristiche del testo e degli item che influenzano la difficoltà delle prove di comprensione, è importante considerare anche i fattori legati ai rispondenti. Le caratteristiche individuali degli studenti giocano, infatti, un ruolo importante nell'interazione con il testo e nella probabilità di rispondere correttamente alle domande. Tra questi, la conoscenza del vocabolario generale rappresenta uno dei predittori più significativi della comprensione del testo, in quanto consente ai lettori di accedere in modo efficiente al significato delle parole e di costruire rappresentazioni mentali coerenti (Spencer, Gilmour, Miller, 2019). Toyama (2019) evidenzia come la padronanza lessicale sia particolarmente rilevante per affrontare testi complessi, poiché facilita l'integrazione tra le informazioni esplicite e le inferenze richieste.

I risultati di uno studio su studenti americani dagli 11 ai 20 anni (Kulesz, 2014) mettono in luce l'importanza del vocabolario e delle conoscenze pregresse come principali predittori della comprensione del testo, indipendentemente dall'età o dal livello scolastico degli studenti. I lettori che possiedono un vocabolario ricco e un ampio bagaglio di conoscenze dimostrano una maggiore probabilità di rispondere correttamente agli item di comprensione, rispetto a coloro che presentano competenze linguistiche e conoscitive meno sviluppate. Questi risultati sono coerenti con la letteratura teorica e sperimentale esistente, secondo cui la costruzione di una rappresentazione mentale coerente del testo dipende dalla profondità e dall'ampiezza delle conoscenze del lettore, comprese quelle lessicali (Perfetti, 2007; Vellutino, 2003). Una padronanza accurata delle forme e dei significati delle parole contribuisce non solo alla rapidità e correttezza della decodifica, ma anche alla capacità di attribuire significato a parole nuove e di integrare concetti all'interno della rappresentazione testuale in costruzione. Le conoscenze pregresse, d'altra parte, facilitano la comprensione del contenuto, permettendo di colmare eventuali lacune informative e promuovendo l'integrazione tra il testo e la conoscenza enciclopedica immagazzinata nella memoria semantica (van den Broek, 2012).

Numerosi studi dimostrano che lettori dotati di un repertorio lessicale ampio e di conoscenze pregresse pertinenti ottengono risultati significativamente migliori nei compiti di comprensione (Ouellette, 2006;



Perfetti, 2007; Ozuru, Dempsey, & McNamara, 2009, Tarchi, 2011). La disponibilità immediata di concetti familiari e termini noti consente infatti una più rapida assimilazione delle informazioni testuali, nonché una loro integrazione con le rappresentazioni mentali già esistenti. Questo processo si rivela particolarmente utile nella lettura di testi meno coesi, dove le connessioni logiche non sono sempre esplicitate: in tali casi, la conoscenza pregressa agisce da supporto inferenziale, colmando eventuali lacune nella coesione testuale (Kendeou & van den Broek, 2007). Inoltre, la familiarità con un ampio repertorio lessicale e concettuale consente ai lettori di affrontare testi di maggiore difficoltà, inclusi quelli di carattere scientifico (Kulesz, 2014). Anche le conoscenze pregresse svolgono un ruolo importante: esse costituiscono uno sfondo su cui si innesta la nuova informazione e influenzano la facilità con cui il lettore costruisce connessioni tra le idee presenti nel testo. Per esempio, Tarchi (2011), in uno studio condotto su studenti del secondo anno della scuola secondaria in Italia, ha analizzato la relazione tra caratteristiche del lettore e tipologia del testo (di scienze e di storia), evidenziando come le conoscenze precedenti relative all'argomento sono un predittore significativo per entrambi i testi.

Un ulteriore aspetto indagato nello studio di Kulesz (2014) riguarda il ruolo della memoria di lavoro e della fluidità di lettura come predittori della comprensione. I dati mostrano che la memoria di lavoro è un fattore significativamente predittivo della performance nei test di comprensione per gli studenti che frequentano le classi dal grado 7 al grado 9, mentre la fluidità di lettura risulta più rilevante per quelli che frequentano le classi dal grado 10 al grado 12. In parallelo, la memoria di lavoro gioca un ruolo cruciale nella gestione e organizzazioni delle informazioni durante la lettura. Essa consente di mantenere attivi e manipolabili in memoria i contenuti testuali appena processati, facilitando la realizzazione di inferenze necessarie per la coerenza globale del testo (Daneman & Hannon, 2001; Linderholm & van den Broek, 2002). Individui con una memoria di lavoro più efficiente risultano, infatti, maggiormente capaci di elaborare legami inferenziali, come le *bridging inferences*, che sono essenziali per connettere porzioni di testo non direttamente collegate (Hannon, 2012). In termini cognitivi, la fluidità di lettura, la memoria di lavoro, il QI e le capacità di ragionamento verbale e non verbale sono stati più volte identificati come fattori predittivi delle abilità di comprensione (Perfetti, Landi, Oakhill, 2005). Spencer et al. (2019) mostrano che lettori con maggiore consapevolezza morfologica, ovvero la capacità di riconoscere e manipolare le unità minime di significato nelle parole, presentano una migliore comprensione dei testi, poiché riescono a decodificare più facilmente vocaboli complessi.

Dunque, le caratteristiche cognitive quali la memoria di lavoro, le abilità analitiche di scomposizione e interpretazione (Cain, Oakhill & Bryant, 2004), ma anche la competenza inferenziale per la generazione di significato implicito e le capacità di visualizzazione mentale per la costruzione di rappresentazioni interne



(Kintsch, 1998) risultano significative nella capacità di rispondere correttamente alle domande di comprensione del testo. Parallelamente, la motivazione, intesa come la disposizione intrinseca o estrinseca verso l'atto di leggere, modulata dallo scopo specifico della lettura, dall'interesse verso il contenuto testuale e dalla percezione di autoefficacia come lettore, esercita un'influenza significativa (Guthrie, Klauda & Ho, 2013).

Inoltre, variabili sociolinguistiche quali il genere, la lingua principale parlata in famiglia e il multilinguismo sono anch'esse frequentemente indagate. Bulut et al. (2022) e Yousefpoori-Naeim et al. (2023) hanno osservato che studenti con una lingua madre diversa da quella della prova possono affrontare ostacoli legati alla padronanza del lessico accademico o alla conoscenza implicita della struttura testuale, con conseguenze sulla loro performance. Tuttavia, in alcuni contesti, il bilinguismo o il multilinguismo possono anche rappresentare una risorsa cognitiva, associata a una maggiore flessibilità mentale e competenze metalinguistiche (Yousefpoori-Naeim et al., 2023). In parallelo, variabili sociodemografiche come l'età e, soprattutto, lo status socioeconomico (SES) sono frequentemente associate alle differenze nei livelli di comprensione: rispondenti provenienti da contesti socioeconomici svantaggiati possono avere un'esposizione minore al linguaggio scritto e parlato, con effetti sulla competenza linguistica generale e sull'accesso al lessico astratto e accademico (Spencer et al., 2019).

In Italia, un'importante fonte di dati sulla relazione tra comprensione del testo degli studenti, le loro caratteristiche individuali e le caratteristiche del contesto è costituita dalle rilevazioni nazionali dell'Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione-INVALSI (per es. INVALSI, 2024). Tali rilevazioni coinvolgono tutti gli studenti di seconda e quinta primaria, dell'ultimo anno della scuola secondaria di primo grado, del secondo anno di scuola secondaria di secondo grado, che in Italia coincide con il termine dell'obbligo di istruzione, nonché dell'ultimo anno di scuola secondaria. Secondo i dati del Rapporto INVALSI 2024, il più recente nel momento della stesura del presente contributo, il genere e l'origine migratoria sono predittori unici significativi della comprensione del testo in tutti i gradi scolastici esaminati, con uno svantaggio degli studenti maschi rispetto alle studentesse, nonché degli studenti con origine migratoria, soprattutto nel caso delle prime generazioni. Emerge inoltre un effetto significativo del contesto socioeconomico e culturale della famiglia di origine del singolo studente; tale vantaggio si amplifica se si considera anche il contesto scolastico frequentato, qualora esso accolga mediamente studenti con profili socioeconomici e culturali più favorevoli. Anche la regolarità del percorso scolastico risulta significativo: gli studenti con almeno un anno di ritardo negli studi hanno prestazioni peggiori, evidenziando un legame diretto tra continuità negli studi e capacità di comprensione. Nella scuola secondaria, l'indirizzo scolastico seguito rappresenta un ulteriore elemento di differenziazione: gli studenti iscritti ai licei mostrano risultati nettamente

superiori, mentre gli alunni degli istituti professionali riportano esiti significativamente più bassi rispetto ai colleghi degli istituti tecnici, evidenziando l'impatto del curriculum di studi sullo sviluppo delle competenze di comprensione del testo. Infine, in tutti i gradi scolastici persistono forti disparità territoriali, anche a parità di condizioni individuali e scolastiche.

Principali modelli di analisi per la previsione della difficoltà degli item

Nell'esplorare la letteratura sulle diverse caratteristiche che possono influenzare la prestazione a item di comprensione del testo, emerge come gli studi sull'argomento abbiano fatto ricorso ad una varietà di modelli statistici e computazionali, alcuni dei quali consentono non solo di individuare l'influenza delle variabili a livello del rispondente, dell'item e del testo, ma anche di coglierne le interazioni a più livelli.

La letteratura ha fatto largo uso di metodi statistici tradizionali, in particolare i modelli di regressione, per identificare i predittori significativi della difficoltà degli item (per una rassegna, vedi Ferrara, 2022). In tali approcci, è tipicamente utilizzato come indicatore della difficoltà degli item il parametro relativo nei modelli della Teoria di Risposta all'Item (*Item Response Theory, IRT*), stimato prima dell'applicazione dei modelli di regressione, oppure la proporzione di risposte corrette. In alcuni casi, sono stati applicati *ex-post* modelli multilivello sui parametri di difficoltà degli item, per tenere conto della struttura annidata dei dati (item all'interno di testi). Altri studi hanno applicato modelli di *machine learning* per individuare i migliori predittori della difficoltà delle domande. Uno studio particolarmente rilevante in questa direzione è quello condotto da Sano et al. (2016), che ha applicato un modello di regressione lineare multipla a un corpus di 60 domande di comprensione, identificando 12 predittori lessicali, sintattici e semantici capaci di spiegare il 52% della varianza nella difficoltà degli item. Parallelamente, una regressione basata su alberi decisionali (*tree-based regression*) ha permesso di spiegare fino all'83% della varianza, utilizzando solo sette caratteristiche psicolinguistiche. Questo approccio si è dimostrato particolarmente utile per individuare interazioni non lineari tra variabili e per mappare soglie critiche oltre le quali le caratteristiche testuali influenzano significativamente la performance dei lettori.

Alcuni studi più recenti (Toyama, 2019; Bulut et al., 2022; Yousefpoori-Naeim et al., 2023; Kulesz et al., 2016; Wang, 2021) hanno adottato un approccio che coniuga i modelli psicometrici con i modelli multilivello: i modelli esplicativi di risposta all'item (*Explanatory Item Response Model, EIRM*; De Boeck & Wilson, 2004). A differenza dei modelli descrittivi dell'IRT, in cui le risposte agli item sono utilizzate per stimare i parametri di difficoltà degli item (ed eventuali altri parametri) e l'abilità degli studenti, i modelli EIRM incorporano covariate a livello degli item e/o dei rispondenti, consentendo di verificare in che misura tali



variabili contribuiscono a spiegare la variabilità nella probabilità di rispondere correttamente agli item. Nel caso in cui nel modello siano introdotte solo le covariate a livello dei rispondenti, allora si ottiene un modello di regressione latente o *person explanatory model*. Nel caso in cui, invece, il modello include variabili corrispondenti alle caratteristiche degli item, ma non le covariate a livello dei rispondenti, si ottiene un *item explanatory model* (noto anche come *Linear Logistic Test Model*; LLTM). Introducendo nello stesso modello sia covariate a livello di item, sia covariate a livello dei rispondenti, si ha il *doubly explanatory model* (noto anche come *latent regression-LLTM*).

Tale approccio si è affermato come uno strumento efficace nell'analisi delle prove di comprensione. Alcuni autori (per es. Gorin e Svetina, 2008) hanno utilizzato l'*item explanatory model* per studiare il ruolo di alcune caratteristiche degli item e del testo stimolo nel predire la probabilità di rispondere correttamente a domande di comprensione del testo. Studi più recenti si sono orientati su modelli *doubly explanatory* (Toyama, 2019; Bulut et al., 2022; Yousefpoori-Naeim et al., 2023; Kulesz et al., 2016; Wang, 2021), offrendo una rappresentazione articolata dei fattori alla base della prestazione agli item di comprensione del testo. In particolare, uno dei maggiori punti di forza di questo approccio è la possibilità di indagare se l'effetto delle caratteristiche degli item o dei testi stimolo varia in funzione delle caratteristiche dei rispondenti, in piena coerenza con il modello euristico di RAND (Snow, 2002).

Conclusioni

La presente rassegna si è prefissata di descrivere e sintetizzare la letteratura sui possibili predittori della variabilità nella prestazione a domande di comprensione del testo, in coerenza con il modello euristico di RAND (Snow, 2002). I risultati dei più recenti studi sull'argomento suggeriscono l'importanza di trovare un punto d'incontro tra tradizioni di ricerca differenti, che si sono focalizzate negli anni prevalentemente sui uno dei tre principali elementi chiave del modello RAND (Snow, 2002), al fine di esplorare le interazioni tra tali elementi. La complessità del costruito oggetto di indagine e le numerose variabili che possono essere individuate a livello del testo, dell'item e del rispondente suggeriscono, inoltre, la necessità di adottare un approccio multidisciplinare.

Si evidenzia inoltre che l'uso di variabili differenti attraverso i diversi studi, nonché di modi diversi di operazionalizzare la stessa variabile, da una parte costituisce un elemento di ricchezza nella letteratura sull'argomento, dall'altra, tuttavia, suggerisce l'importanza di studiare il ruolo di un set di variabili identificate come particolarmente rilevanti su più test e popolazioni, al fine di ottenere evidenze sempre più sistematiche e generalizzabili.



Un aspetto critico emerso dall'analisi della letteratura riguarda la distribuzione geografico-linguistica degli studi esistenti. La grande maggioranza delle ricerche considerate è stata condotta in contesti di lingua inglese. Questa predominanza comporta un evidente squilibrio nella rappresentazione linguistica e culturale, con il rischio che le conclusioni tratte siano parzialmente influenzate dalle specificità linguistiche e didattiche dell'inglese come L1 (Bulut et al., 2022; Toyama, 2019; Spencer et al., 2019). Tale squilibrio pone una questione metodologica e teorica rilevante: le caratteristiche del testo, degli item e dei rispondenti che influenzano la comprensione potrebbero manifestarsi in modo diverso in lingue con strutture differenti, come l'italiano. Pertanto, si evidenzia la necessità di approfondire l'analisi dei fattori che influenzano la difficoltà degli item di comprensione del testo anche in contesti non anglofoni. Infine, una maggiore comprensione dei fattori che influenzano la difficoltà degli item prove di comprensione del testo permette di costruire strumenti di valutazione più sensibili e di orientare interventi mirati a sostenere lo sviluppo di competenze di comprensione.

Bibliografia

- Agrusti, F., Cardinaletti, A., Desimoni, M., Pallotti, G., & Montemagni, S. (2024). 2. LingUistic Complexity Evaluation in educaTion (LuCET). *CADMO*, (2024/1).
- Alderson, J. C. (2000). *Assessing reading*. Cambridge, UK: Cambridge University Press.
- Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology*, 29(2), 137-164. <https://doi.org/10.1080/02702710801963951>
- Britton, B. K., & Gülgöz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology*, 83(3), 329–345.
- Bulut, H. C., Bulut, O., & Arikan, S. (2022). Evaluating group differences in online reading comprehension: The impact of item properties. *International Journal of Testing*, 23(1), 10–33. <https://doi.org/10.1080/15305058.2022.2044821>
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's Reading Comprehension Ability: Concurrent Prediction by Working Memory, Verbal Ability, and Component Skills. *Journal of Educational Psychology*, 96(1), 31–42. <https://doi.org/10.1037/0022-0663.96.1.31>
- Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, 42(3), 475–493.
- Daneman, M., & Hannon, B. (2001). Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT. *Journal of Experimental Psychology: General*, 130, 208–223. <http://dx.doi.org/10.1037/0096-3445.130.2.208>
- De Boeck, P., Cho, S. J., & Wilson, M. (2016). Explanatory item response models. *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications*, 247-266.



- Dell'Orletta, F., Montemagni, S., and Venturi, G. (2011). READ-IT: assessing readability of Italian texts with a view to text simplification. In *Proceedings of Second Workshop on Speech and Language Processing for Assistive Technologies*, pages 73–83, Edimburgo, UK, July
- Drum, P., Grover, D., & Flake, S. (1981). Predicting test item difficulty with item response theory. *Journal of Educational Psychology*, 73(6), 848-854.
- Ferrara, S., Steedle, J. T., & Frantz, R. S. (2022). Response Demands of Reading Comprehension Test Items: A Review of Item Difficulty Modeling Studies. *Applied Measurement in Education*, 35(3), 237–253. <https://doi.org/10.1080/08957347.2022.2103135>
- Freedle, R., & Kostin, I. (1992). The relationship between reading ability and item difficulty in multiple-choice reading comprehension tests. *Language Testing*, 9(1), 1-12.
- Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.
- Gorin, J. S., & Svetina, M. (2008). The relationship between item characteristics and item difficulty in reading comprehension tests. *Applied Measurement in Education*, 21(1), 35-56.
- Graesser, A.C, McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223–234. <https://doi.org/10.3102/0013189X11413260>
- Guthrie, J.T., Klauda, S.L. & Ho, A.N. (2013). Modeling the Relationships Among Reading Instruction, Motivation, Engagement, and Achievement for Adolescents. *Reading Research Quarterly*, 48(1), 9–26. [doi: 10.1002/rrq.035](https://doi.org/10.1002/rrq.035)
- Hannon, B. (2012). Understanding the relative contributions of lower-level word processes, higher-level processes, and working memory to reading comprehension performance in proficient adult readers. *Reading Research Quarterly*, 47, 125–152.



INVALSI. (2024). *Rapporto nazionale 2024*. Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione. <https://www.invalsi.it>

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Boston: Allyn & Bacon.

Kendeou, P., & Van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, 35, 1567–1577. <http://dx.doi.org/10.3758/BF03193491>

Kincaid, J. P., Fishburne, L. R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. In *Research Branch Report* (pp. 8–75). Memphis: Chief of Naval Technical Training: Naval Air Station.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge university press.

Kirsch, I. S., & Mosenthal, P. B. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly*, 25(1), 5–30. doi:10.2307/747985

Klare, G. R. (1984). Readability. In P. D. Pearson, R. Barr, & M. L. Kamil (Eds.), *Handbook of Reading Research* (pp. 681–744). New York, NY: Longman.

Koslin, B. L., Zeno, S., & Koslin, S. (1987). *The DRP: An effective measure in reading*. New York, NY.

Kulesz, P. A. (2014). *The effects of reader characteristics, text features, and comprehension processes on reading comprehension*. University of Houston.

Kulesz, P. A., Francis, D. J., Barnes, M. A., & Fletcher, J. M. (2016). The influence of properties of the test and their interactions with reader characteristics on reading comprehension: An explanatory item response study. *Journal of Educational Psychology*, 108(8), 1078–1097.

Lennon, C., & Burdick, H. (2004). The lexile framework as an approach for reading measurement and success. *electronic publication on www.lexile.com*.

Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, *94*, 778–784. <http://dx.doi.org/10.1037/0022-0663.94.4.778>

Lucisano, P., & Piemontese, M. E. (1988). Gulpease: una formula per la predizione della leggibilità di testi in lingua italiana. *Scuola e città*, 110-124.

Lumley, T., Routitsky, A., Mendelovits, J., & Ramalingam, D. (2012). A framework for predicting item difficulty in reading tests. <https://research.acer.edu.au/pisa/5>

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2009). Linguistic Features of Writing Quality. *Written Communication*, *27*(1), 57-86. <https://doi.org/10.1177/0741088309351547> (Original work published 2010)

McNamara, D. S., Louwse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: *Capturing linguistic features of cohesion*. *Discourse Processes*, *47*(4), 292-330.

McNamara, D. S., Graesser, A., & Louwse, M. (2012). Sources of text difficulty: Across genres and grades. In J. Sabatini, E. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 89-116). R & L Education.

McNamara, D.S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, MA: Cambridge University Press.

McNamara, D. S., Ozuru, Y., & Floyd, R. G. (2011). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International Electronic Journal of Elementary Education*, *4*(1), 229-257.

Medina, A. L., & Pilonieta, P. (2006). Once upon a time: Comprehending narrative text. In J. S. Schumm (Eds.), *Reading assessment and instruction for all learners* (pp. 222–261). Guilford Press.



Mullis, I. V., & Martin, M. O. (2019). *PIRLS 2021 Assessment Frameworks*. International Association for the Evaluation of Educational Achievement. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

National Governors Association [NGA], & Council of Chief State School Officers[CCSSO]. (2010). *Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects*. Retrieved June 2, 2019 from <http://www.corestandards.org/>

OECD (2019), *PISA 2018 Assessment and Analytical Framework*, PISA, OECD Publishing, Paris, <https://doi.org/10.1787/b25efab8-en>.

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98(3), 554–566. <https://doi.org/10.1037/0022-0663.98.3.554>

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and instruction*, 19(3), 228-242.

Padovani, R. (2006). Text comprehension by school-age children. A review on normal and atypical development. *Psicologia Clinica dello Sviluppo*, 10(3), 369-398. <https://doi.org/10.1449/23210>

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, 11(4), 357–383. <https://doi.org/10.1080/10888430701530730>

Perfetti, C., Landi, N., & Oakhill, J. (2005) – In "The acquisition of reading comprehension skill"

Dell’Orletta F., Montemagni S., Venturi G. (2011) READ-IT: Assessing Readability of Italian Texts with a View to Text Simplification, in *Proceedings of the Workshop on Speech and Language Processing for Assistive Technologies (SLPAT 2011)*, Edinburgh, July 30, pp. 73-83, <http://aclweb.org/anthology/W/W11/W11-2308.pdf>.

Rice, M., Wijekumar, K., Lambright, K. *et al.* Inferencing in Reading Comprehension: Examining Variations in Definition, Instruction, and Assessment. *Tech Know Learn* 29, 1169–1190 (2024). <https://doi.org/10.1007/s10758-023-09660-y>



Sano, M. (2016, April). Improvements in automated capturing of psycho-linguistic features in reading assessment text. In *The annual meeting of National Council on Measurement in Education* (pp. 1-27).

Schnick, T., & Knickelbine, M. (2007). *Using the lexile analyzer for educators and media specialists*. Durham, NC: MetaMetrics, Inc.

Sheehan, K. M. (2016). *A review of evidence presented in support of three key claims in the validity argument for the TextEvaluator® text analysis tool (Research Report No. RR-16-12)*. Parsippany, NJ: John Wiley & Sons, Ltd. <https://doi.org/10.1002/ets2.12100>

Sheehan, K. M., Kostin, I., Napolitano, D., & Flor, M. (2014). The TextEvaluator tool: Helping teachers and test developers select texts for use in instruction and assessment. *Elementary School Journal*, 115(2), 184–209. <https://doi.org/https://doi.org/10.1086/678294>

Sheehan, K., & Mislevy, R. J. (1990). Integrating cognitive and psychometric models to measure document literacy. *Journal of Educational Measurement*, 27(3), 255–272. doi:10.1111/j.1745-3984.1990.tb00747.x

Snow, C. (2002) Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. RAND Corporation.

Spencer, M., Gilmour, A. F., Miller, A. C., Emerson, A. M., Saha, N. M., & Cutting, L. E. (2019). Understanding the influence of text complexity and question type on reading outcomes. *Reading and writing*, 32, 603-637.

Tarchi, C. (2012). La comprensione del testo espositivo. L'interazione tra conoscenze precedenti del lettore e contenuto del testo. *Giornale italiano di psicologia*, 39(1), 91-114.

Toyama, Y. (2019). *What makes reading difficult? An investigation of the contribution of passage, task, and reader characteristics on item difficulty, using explanatory item response models*. University of California, Berkeley.



Toyama, Y. (2021). What makes reading difficult? An Investigation of the contributions of passage, task, and reader characteristics on comprehension performance. *Reading Research Quarterly*, 56(4), 633-642.

Uccelli, P., Dobbs, C. L., & Scott, J. (2012). Mastering Academic Language: Organization and Stance in the Persuasive Writing of High School Students. *Written Communication*, 30(1), 36-62. <https://doi.org/10.1177/0741088312469013> (Original work published 2013)

Van den Broek, P. (2012). Individual and developmental differences in reading comprehension: Assessing cognitive processes and outcomes. *Measuring up: Advances in how we assess reading ability*, 39-58.

Valencia, S. W., Wixson, K. K., Ackerman, T., & Sanders, E. (2017). Identifying text-task-reader interactions related to item and block difficulty in the national assessment for educational progress reading assessment. *San Mateo, CA: National Center for Education Statistics*.

Vellutino, F. R. (2003). Individual differences as sources of variability in reading comprehension in elementary school children. *Rethinking reading comprehension*, 51-81.

Wang, P. (2021). *Reader-Test Interactions: An Explanatory Item Response Study on Reading Comprehension* (Doctoral dissertation, Middle Tennessee State University).

Yousefpoori-Naeim, M., Bulut, O., & Tan, B. (2023). Predicting reading comprehension performance based on student characteristics and item properties. *Studies in Educational Evaluation*, 79, 101309.