

Gli anni di Eco. Riflessioni sull'uso di prove strutturate con risposte chiuse e aperte a margine di una ricerca sulla comprensione della lettura

Alessandra Intraversato - Pietro Lucisano

«Sapienza» Università di Roma, Dipartimento di Psicologia dei Processi di Sviluppo
e Socializzazione

doi: 10.7358/ecps-2013-007-intr

alessandra.intraversato@uniroma1.it

pietro.lucisano@uniroma1.it

ECO'S AGE. REFLECTIONS ON THE USE
OF STRUCTURED TESTS WITH OPEN ENDED
AND MULTIPLE CHOICE ITEMS LINKED
TO A STUDY ON READING COMPREHENSION

ABSTRACT

The present study stems from a broader work aiming to validate a reading comprehension test for final year high school students; the work stands as a stimulus for reflection on the use of structured tests to assess reading comprehension skills and, in particular, analyses the relationship between the different areas of assessment of reading comprehension processes and question types (in open-ended versus multiple choice items). By analysing an open-ended question, we discuss the possibility of assessing the mental processes involved in reading and their relations with students' encyclopaedic knowledge. A great many studies have demonstrated how prior knowledge is one of the main variables explaining reading skills (20-70% of variance in a test of reading comprehension) and that excluding it can distort measures of reading literacy. Overall, it appears that in addition to the methodological rigor needed to construct valid and reliable tests, there is need for an approach to more fully understand and to more deeply consider all the solutions that strategic and skilled readers implement and put in place in order to completely comprehend a text.

Keywords: Encyclopaedic knowledge, Multiple choice question, Open-ended question, Reading comprehension, Validity.

1. INTRODUZIONE AL PROBLEMA DI RICERCA

Lungi da noi voler mettere in questione le procedure di misurazione svolte per mezzo di prove strutturate. Siamo convinti che se le prove sono ben costruite, validate, tarate e analizzate con strumenti adeguati, possono contribuire a migliorare la conoscenza degli studenti e fornire delle utili indicazioni pratiche. Tuttavia, le indicazioni che possiamo trarne possono essere talvolta distorte, se le prove non sono ben costruite, o se non sappiamo usare i risultati in modo consapevole e con la necessaria comprensione dei limiti imposti da strumenti di questo tipo.

In questa sede vogliamo mettere a fuoco un problema che attiene al limite di questi strumenti, limite che si accentua ogni qual volta l'utilizzatore inesperto si serva delle informazioni per trarne dogmi per quella che Fedro definiva «Chiesa della ragione».

Corre l'obbligo di definire l'utilizzatore inesperto perché questa definizione comprende tutti coloro che ritengono di poter costruire domande, esaminare risposte e valutarne i risultati senza una adeguata preparazione metodologica, estendendo l'assunto gentiliano che chi conosce i contenuti può far tutto. Ci risulta che siano utilizzatori inesperti tutti coloro che si avvicinano al *testing* senza una adeguata preparazione. I test strutturati si definiscono oggettivi perché la modalità di correzione delle risposte è predefinita: non è dunque soggettivo – dipendente dal valutatore – il punteggio che se ne ricava; questo è il motivo per cui ne viene fatto un largo uso in ambito selettivo, dai test di ammissione all'università ai concorsi pubblici. Ciò che l'utilizzatore inesperto solitamente ignora, è che nella teoria classica dei test, si assume il punteggio osservato al test come somma delle risposte esatte più l'errore standard; inoltre, la costruzione di un test con risposte multiple, fatta da chiunque a qualunque titolo, non implica necessariamente la validità dello strumento, poiché non è detto che ciò che stiamo misurando sia proprio ciò che intendiamo misurare. In altri termini, se è oggettiva la misura, sulla valutazione del risultato è già necessario usare delle cautele, sia in ragione della validità dello strumento che dell'affidabilità statistica del risultato.

Il nostro scopo è sollecitare l'attenzione sul fatto che la comprensione è un obiettivo che deve precedere di gran lunga la tanto evocata valutazione e che la misura contiene in sé la necessità di cautele anche per evitare i rischi di un relativismo totale.

Il problema che qui proponiamo, è capire fino a che punto gli strumenti che la ricerca educativa ci mette a disposizione – segnatamente le prove strutturate – possano verificare l'esattezza dei ragionamenti che avvengono durante la lettura.

2. ANALISI DEL QUADRO DELLA RICERCA

Nell'ambito della comprensione della lettura esiste una certa contrapposizione tra i fautori dei test basati su domande con alternative di risposta chiusa (es. a risposta multipla) e fautori delle domande a risposta aperta (Pearson, 1984; Lucisano, 1989; Pressley, 1990; Pearson, 1999; Moncada, 2010; Warwick, 2010).

La modalità a risposta chiusa è stata largamente utilizzata nelle valutazioni di sistema negli USA e nelle indagini internazionali a partire dalle ricerche negli anni '70 (Postlethwaite *et al.*, 1977; Trivellato & Zuliani, 1977; Giannarelli, 1977), tuttavia, importanti indagini come Pisa (Topping, 2003; Cresswell, 2006; OECD, 2010) o le più recenti indagini IEA (Gorman *et al.*, 1988; Postlethwaite *et al.*, 1992; Purves, 1992; Campbell *et al.*, 2001; Mullis *et al.*, 2006 e 2009), hanno adottato un numero sempre crescente di domande aperte, che prevedono la presenza di codificatori delle risposte a partire da una griglia comune di correzione.

Il motivo per cui il formato delle risposte chiuse è il più diffuso è semplice: innanzitutto garantisce l'oggettività di correzione, inoltre è più economico in termini di tempo necessario a produrle, a somministrarle e a correggerle, e attraverso varie tecniche di item analisi, è possibile ottenere utili informazioni sulle caratteristiche dello strumento e sulle misure rilevate. Tale economicità si raggiunge solo se il test è stato adeguatamente costruito, altrimenti si ha un inutile sperpero di denaro, in genere pubblico, e informazioni distorte sugli studenti. Queste precisazioni sono sempre necessarie, stante la mai sopita polemica circa l'impossibilità di principio, per i test strutturati, di rivelare le conoscenze e le abilità degli studenti (Lucisano, 2003).

Una delle critiche mosse a questo genere di prove risiede nella limitatezza e predeterminazione delle possibili opzioni di risposta: in altri termini, non è detto che i distrattori presentati intercettino effettivamente le possibili alternative di risposta elaborate dagli studenti (Ozuru *et al.*, 2007). Inoltre, proponendo delle alternative di risposta, si incanala, anche involontariamente, il ragionamento dello studente in una certa direzione, escludendo a priori soluzioni alternative che non sono state prese in considerazione da chi ha costruito il test; in questo caso si crea una specie di *feedback* a partire dalle alternative di risposta che facilita la risoluzione dell'esercizio. Nel caso di alternative di

risposta palesemente inverosimili (spesso accade quando chi costruisce il test non è un professionista), si ha l'effetto di ridurre agli occhi dello studente le risposte possibili, e quindi di aumentare le sue possibilità di successo nel caso gettasse a caso una risposta, aumentando quindi la facilità dell'item.

Si consideri, inoltre, che la stessa sezione delle domande di comprensione che seguono un testo – è di per sé – un testo: l'autosufficienza della sezione delle domande è stata dimostrata da alcune ricerche che hanno verificato come sia possibile rispondere a talune prove senza nemmeno leggere il testo cui si riferiscono le domande (Katz *et al.*, 1990).

È stato notato già da molto tempo (Frederiksen, 1984) che le prove strutturate a risposta chiusa sono maggiormente utilizzate per rilevare abilità di base e possesso di nozioni. Ad esempio nel caso di testi di comprensione della lettura, richiedono l'individuazione d'informazioni nel testo, mentre meno frequentemente vengono utilizzate per analizzare l'attività inferenziale del lettore, o per dare valutazioni del testo. È assai complesso ricavare informazioni sui processi di *problem solving* messi in atto dagli studenti domande a scelta multipla, mentre è più facile intercettare e valutare tali processi tramite quesiti a risposta aperta.

I fautori del formato delle risposte aperte le considerano più vicine all'esperienza scolastica quotidiana (Rauch & Hartig, 2010), inoltre attraverso tale formato, è più facile indagare competenze più complesse, non limitandosi al solo recupero d'informazioni nel testo. La correzione delle risposte aperte richiede molto più lavoro in termini di griglie di correzione comuni, codifica e trasformazione, ma permettono di valutare la ricchezza delle idee presentate, le soluzioni alternative e i ragionamenti seguiti dagli studenti, sono quindi sentite, anche dagli insegnanti, più vicine al lavoro svolto in classe. Tuttavia le risposte aperte hanno un forte limite: richiedono da parte dello studente un'attività di produzione scritta che può essere considerata una difficoltà ulteriore nella risposta, a svantaggio quindi di chi ha una scarsa abilità di produzione scritta (non a caso i correttori delle risposte aperte sono addestrati a ignorare errori di ortografia e grammatica e ad accettare le risposte fin quando esse sono comprensibili). Normalmente in un test che presenti sia risposte aperte che chiuse, le risposte aperte hanno una percentuale di non risposta notevolmente più elevato delle chiuse, proprio perché ritenute più impegnative.

Si deve infine notare che l'operazione di semplificazione che mette in atto il costruttore del test quando produce le domande chiuse, trova una sua corrispondenza nell'opera del codificatore delle risposte aperte, che è costretto a ridurre e semplificare quanto scritto dagli studenti per individuare le categorie di risposta e valutarle di conseguenza. Infatti, sebbene le risposte degli studenti «nascano» aperte, esse devono essere codificate per essere corrette, per poi essere analizzate con le stesse metodologie statistiche che si usano per le chiuse.

È importante sottolineare che la dicotomia esistente fra domande a risposta chiuse o aperte non risiede tanto nel formato della domanda, ma nelle abilità e conoscenze che queste intendono misurare. È stato più volte verificato che somministrando il medesimo test nelle due modalità (risposte chiuse e risposte aperte) gli indici di correlazione fra le stesse domande nelle due versioni è molto alto; tuttavia questo accade soprattutto quando l'adattamento avviene a partire da domande chiuse, infatti è più facile costruire domande strutturate che richiedano solo nozioni di base o il semplice recupero di informazioni, viceversa quando si devono adattare domande a risposta aperta che richiedano attività inferenziali complesse è più difficile creare dei veri omologhi nel formato chiuso; quando è stata testata questa soluzione la correlazione fra le due tipologie di test è diminuita fortemente (Frederiksen, 1984).

La tecnica del *thinking aloud*, utilizzata all'estero (Ericsson & Simon, 1993; Pearson & Fielding, 1991; Kucan & Beck, 1997) come in Italia (Lumbelli, 1989, 1996, 2009) soprattutto a scopo diagnostico, è uno degli strumenti più utili per esplodere i molti significati e ragionamenti del lettore; Ericsson e Simon (1993) hanno usato questo strumento per validare esternamente delle prove strutturate, ossia per verificare se le attività mentali messe in atto nel rispondere a quesiti aperti o chiusi differissero in modo sostanziale. La ricerca non è riuscita a dimostrare l'esistenza di attività cognitive realmente diverse in connessione con le due modalità di domanda, le differenze emergono solo quando il tipo di lavoro richiesto prevedeva un approfondito lavoro interpretativo del testo, in questo caso le risposte aperte sembrano essere meglio in grado di descrivere il lavoro compiuto dagli studenti.

Nell'ambito della comprensione della lettura, le aree di verifica sono riasumibili in tre categorie: (a) individuare di informazioni nel testo; (b) interpretare il testo integrando informazioni e concetti (inferenze semplici, inferenze complesse, collegamenti di idee, nessi sintattici); (c) riflettere e valutare la forma e i contenuti del testo (OECD, 2009; Invalsi, 2011). Perché il processo di comprensione si realizzi, non è però sufficiente che il lettore sia in grado di svolgere queste operazioni; per realizzare il modello situazionale di cui parla Kintsch è necessario inserire quanto è presente nel testo (*text base*) nella memoria a breve e lungo termine, affinché i significati si generino a partire dalle conoscenze già in possesso del lettore (*knowledge net*) (Kintsch, 1983 e 1998).

La rete delle conoscenze del lettore è un ulteriore ambito di verifica per la comprensione della lettura (Dochy, 1994; Kintsch, 1988; Mc Namara & Kintsch, 1996; Schapiro, 2004):

[...] it appears that domain-specific knowledge can compensate for low intellectual ability, but a high intellectual ability cannot compensate for a low prior knowledge. (Dochy, Segers, & Buehl, 1999)

La variabile «conoscenza pregressa» arriva a spiegare, anche in correlazione con le altre variabili, dal 30% al 60% della varianza nei punteggi di test di comprensione della lettura (Dochy *et al.*, 1999), risultando quindi la variabile indipendente con maggior potere esplicativo in assoluto.

Stante la complessità appena descritta, ci si deve chiedere se per verificare adeguatamente i livelli di padronanza in lettura degli studenti sia possibile utilizzare indifferentemente test strutturati con risposte chiuse e aperte, o se invece sia da preferire una delle due modalità a seconda dell'area di verifica (recupero informazioni, inferenze, interpretazione, opinioni sul testo).

La nostra domanda di ricerca è dunque se sia possibile misurare le stesse abilità indifferentemente con entrambi gli strumenti, e quale validità relativa abbiano le due modalità. Cercheremo di capire se sia possibile tracciare l'attività interpretativa dello studente a partire da domande a risposta chiusa, e verificare l'esattezza dei ragionamenti seguiti per arrivare alla risposta, o se non sia invece possibile pervenire alla risposta esatta anche tramite percorsi alternativi, inizialmente non previsti dal costruttore dei test.

3. PRESENTAZIONE DELLA RICERCA

Le riflessioni che seguono nascono a margine di un'indagine del 2010 volta a misurare i livelli di abilità di comprensione del testo su un campione probabilistico di 723 studenti liceali della città di Roma. La ricerca, sviluppata all'interno di un lavoro di dottorato, ha misurato i livelli di abilità con prove coerenti con i percorsi di studio degli studenti liceali italiani, cercando di superare il limite delle tradizionali prove internazionali e delle imitazioni nazionali che tendono a escludere dai testi elementi di conoscenze enciclopediche (Bottani, 2011). Per questo agli studenti era chiesto di rispondere a quesiti a scelta multipla con quattro alternative di risposta su testi di un certo impegno e di discreta lunghezza, che richiedevano non solo la semplice decodifica del testo, ma l'integrazione delle informazioni fornite dal testo con elementi di conoscenza che sono comunemente richiesti a studenti di quel livello di scolarità. Veniamo al problema sul quale vi proponiamo di riflettere. Tra i testi su cui fare domande per la prova avevamo scelto un gustoso articolo di Umberto Eco dal titolo *C'era una volta Churchill* che ben si prestava agli obiettivi della nostra ricerca dato che presentava sia contenuti interni complessi sia un intreccio di rimandi che richiedevano una elevata cooperazione interpretativa da parte del lettore e la capacità di quest'ultimo di ricorrere a informazioni «che uno studente liceale dovrebbe possedere». Il testo commentava i risultati di un'indagine svolta in Inghilterra dalla qua-

le risultava che un terzo degli inglesi ritenesse Churchill un personaggio di fantasia. Nell'argomentare, Umberto Eco introduce un passo che contiene informazioni che consentono di risalire alla sua età. Quando abbiamo proposto la domanda, nel Maggio 2010, Eco aveva 78 anni. L'informazione è ricavabile dal terzo paragrafo del testo:

Però è vero, e ce ne accorgiamo quando si rivolgono domande ai nostri giovani (per non dire a quelli, che so, americani), che le idee sul passato anche prossimo sono molto vaghe. Si è letto di test da cui appariva che qualcuno credeva che Moro fosse un brigatista rosso, De Gasperi un capo fascista, Badoglio un partigiano eccetera. Uno dice: è passato tanto tempo, perché dei diciottenni devono sapere chi era al governo cinquant'anni prima che loro nascessero? Beh, sarà che la scuola fascista ce ne faceva una testa così, ma io a dieci anni sapevo che il primo ministro ai tempi della marcia su Roma (vent'anni prima) era Facta, e a diciott'anni sapevo anche chi erano stati Rattazzi o Crispi, ed era roba del secolo prima.

A partire da quanto riferito nel testo, e recuperando nelle proprie conoscenze la data della marcia su Roma (1922), sarebbe stato possibile risalire all'età esatta di Umberto Eco. Il nostro obiettivo era verificare in che misura gli studenti fossero in grado, attraverso un ragionamento complesso, di integrare nozioni da recuperare nelle conoscenze enciclopediche dello studente, inferenze complesse e calcoli matematici basilari. Questo compito esemplifica i meccanismi che sono alla base della comprensione della lettura e dell'apprendimento a partire da testi complessi come quelli presentati ai liceali. Su questo passo abbiamo pensato di porre la domanda in due modi, nel modo più tradizionale con un item a scelta multipla (Tabella 1).

**Dalle indicazioni sulla sua esperienza scolastica
quanti anni dovrebbe avere Umberto Eco?**

- a) 50 -60
- b) 60-70
- c) 70-80
- d) Più di 80

Poi, per verificare l'attivazione delle inferenze di cui sopra, abbiamo inserito una domanda aperta che chiedeva di esplicitare il percorso seguito per determinare la risposta alla domanda precedente:

**Puoi descrivere il ragionamento con cui ricavi dal testo
l'età di Umberto Eco?**

Parallelamente abbiamo proceduto a una valutazione delle risposte aperte di tipo dicotomico (Tabella 2).

Per analizzare la risposta aperta abbiamo elaborato una check list delle azioni che lo studente dovrebbe compiere, in termini di recupero di informazioni dalla memoria a lungo termine e processi mentali, al fine di rispondere correttamente alla domanda. La check list corrisponde alla somma delle azioni necessarie alla creazione di un modello situazionale completo per la corretta interpretazione del testo e alle attività di *problem solving* richieste per rispondere alla domanda.

Gli elementi della nostra check list sono quattro:

1. Individua il paragrafo di riferimento.
2. Recupera l'informazione circa la data della Marcia su Roma (1922).
3. Compie l'inferenza: se Eco aveva 10 anni 20 anni dopo la Marcia su Roma, allora Eco è nato 10 anni dopo la Marcia su Roma.
4. Calcola l'età di Eco 2010 - $[(1922 + 20) - 10] = 78$.

Per ognuna di queste fasi lo studente poteva compiere l'azione in modo corretto, sbagliato oppure approssimativo e generico, ed infine poteva scegliere di seguire un percorso completamente diverso.

L'attribuzione del valore corretto alla domanda deriva dalla considerazione degli elementi espressi nella risposta aperta. Viene considerata corretta la risposta che muove dalla individuazione degli indizi presenti nel testo e dalla esplicitazione dei calcoli necessari.

Sono state considerate corrette anche le risposte che pur non esplicitando tutti i passaggi li sottintendevano in modo tale che il lettore ne potesse ricostruire il processo. Sono state considerate errate le risposte in cui lo studente non coglie gli indizi di riferimento, li coglie ma non è in grado di collocarli (la data della marcia su Roma), li coglie, li colloca correttamente, ma sbaglia nel calcolare l'età di Umberto Eco. Infatti, è frequente che il ragionamento corretto si concluda con un banale errore di calcolo, dovuto alla non considerazione di elementi, talvolta enunciati esplicitamente come termini del ragionamento, ma dimenticati al momento del calcolo materiale. È evidente che in questo caso non entrano in causa abilità di calcolo poiché si tratta di addizioni e sottrazioni che assumiamo come disponibili in studenti liceali, ma di problemi di attenzione o di rilevanza attribuita all'operazione in questione.

In altri casi l'impianto del ragionamento è corretto, ma i dati enciclopedici sono sbagliati. Dunque se l'anno della marcia su Roma viene spostato in avanti o indietro e i calcoli sono svolti correttamente il risultato è comunque sbagliato.

L'incrocio degli esiti della domanda chiusa e della correzione della domanda aperta mostra tuttavia come molti studenti raggiungano il risultato corretto nella domanda chiusa attraverso considerazioni diverse da quelle che

noi avevamo previsto come necessarie, o attraverso stime approssimative, o ancora attraverso una molteplicità di errori che si elidono a vicenda.

Veniamo agli esiti delle due domande: nella Tabella 1 sono presentate le risposte degli studenti alla domanda chiusa, e nella Tabella 2 sono presentate le correzioni della domanda aperta. Per l'item analisi classica la domanda chiusa ha un indice di correlazione punto biseriale 0,36 che indica una discreta discriminatività e una facilità di 0,53; a seguito della codifica, la domanda aperta è stata analizzata con la medesima metodologia della chiusa, e ha un punto biseriale di 0,40 ma risulta molto più difficile, l'indice di facilità è 0,17.

Tabella 1. – Frequenze delle risposte alla domanda chiusa.

	FREQUENZA	PERCENTUALE	PERCENTUALE VALIDA
a) 50-60	19	2,6	2,7
b) 60-70	107	14,8	15,4
c) 70-80	391	54,1	56,3
d) Più di 80	178	24,6	25,6
Totale	695	96,1	100,0
Mancanti	28	3,9	
Totale	723	100,0	

Tabella 2. – Risposte alla domanda aperta su età di Eco codificate in modo dicotomico.

	FREQUENZA	PERCENTUALE	PERCENTUALE VALIDA
Corretto	131	18,1	21,8
Errato	469	64,9	78,2
Totale validi	600	83,0	100,0
Mancanti	123	17,0	
Totale	723	100,0	

Salta agli occhi la differenza negli esiti delle due domande. Nella prima modalità di domanda le risposte corrette sono il 54% del campione e il 56,3% degli studenti che hanno risposto alla domanda. Nella domanda aperta le risposte corrette sono il 18% del campione, le omissioni aumentano notevolmente, ma la percentuale delle risposte corrette degli studenti che hanno

risposto si ferma al 21%. Ne consegue che c'è una rilevante percentuale di studenti che ha raggiunto la soluzione del problema che avevamo presentato secondo percorsi diversi da quelli canonici e, tuttavia, anticipiamo che questo non è dovuto solo al fenomeno del «tirare a indovinare», ma in diversi casi a percorsi in cui i ragazzi hanno utilizzato indizi diversi da quelli canonici e un po' di intuito.

3.1. *Analisi della risposta aperta*

Vediamo alcuni esempi di risposte corrette che individuano gli indizi e che esplicitano tutti gli elementi del ragionamento.

In questa categoria abbiamo compreso gli studenti che individuano il paragrafo in cui sono contenute le informazioni, conoscono la data della marcia su Roma, e calcolano correttamente la data di nascita di Umberto Eco. Si tratta del 20,5% del campione.

[18460759] Eco ci parla della marcia su Roma avvenuta nel 1922 che avvenne 20 anni prima dell'epoca in cui Eco frequenta la scuola 1942, anno in cui Eco aveva già 10 anni, ora siamo nel 2010, Eco dovrebbe avere 78 anni.

Ecco la risposta di uno studente molto sintetico:

[19480789] $2010 - (1922 + 20 - 10) = 78$.

Lo studente successivo rimane più vago nel riferirsi agli anni '30 e tuttavia raggiunge il risultato richiesto; risponde come lui il 18,7% degli studenti.

[06170253] Umberto Eco afferma che a dieci anni sapeva chi fosse il primo ministro ai tempi della marcia su Roma. La marcia su Roma è del 1922 ed Eco afferma di avere dieci anni vent'anni dopo tale evento. Di conseguenza egli deve essere nato negli anni '30 quindi ora dovrebbe essere alle soglie degli ottanta anni.

Il riferimento enciclopedico che gli studenti dovevano identificare per determinare l'età di Umberto Eco è la data della marcia su Roma.

Sulla data della marcia su Roma solo un numero ristretto di studenti non ha esitazioni (175 pari al 24,2%), mentre il 10,8% degli studenti esplicita una data sbagliata, il 36,9% individua il passo e sviluppa il ragionamento senza citare la data della marcia su Roma. Numerosi sono anche gli studenti che non individuano gli elementi da noi ritenuti necessari per risolvere il problema (11,1%) e gli studenti che non rispondono alla domanda aperta (17%).

Tabella 3. – Domanda aperta sulla Marcia su Roma; risposte in relazione all'individuazione della data.

MARCIA SU ROMA	FREQUENZA	PERCENTUALE	PERCENTUALE VALIDA
1922	175	24,2	29,2
Sbaglia la data	78	10,8	13,0
Non cita la data	267	36,9	44,5
Non individua il paragrafo	80	11,1	13,3
Totale validi	600	83,0	100,0
Mancanti	123	17,0	
Totale	723	100,0	

La data della Marcia su Roma, oscilla vertiginosamente dal 1870 agli anni '40 del 900 (Tabella 4).

Tabella 4. – Oscillazioni della data della Marcia su Roma nelle risposte.

DATE DELLA MARCIA SU ROMA ERRATE	NUMERO DELLE RISPOSTE
Dal 1870 al 1920	24
1921 e 1923	15
Dal 1924 al 1940	39
Totale	78

[10240387] Umberto Eco rivela che a dieci anni sapeva chi era Facta, che vent'anni prima era il I ministro ai tempi della Marcia su Roma. La Marcia su Roma avvenne nel 1915, quindi Eco dovrebbe avere all'incirca 85 anni.

[19480801] La marcia su Roma avviene alla fine del 1800, quindi Eco è nato attorno al 1910, e ora ha circa 80 anni.

In uno dei casi seguenti lo studente sbaglia la data, ma compie il ragionamento e i calcoli correttamente (4% del campione); lo studente enuclea tutti i termini del ragionamento, ma prudentemente nelle conclusioni evita la risposta precisa.

[04130194] Dal testo si ricava che Umberto Eco aveva dieci anni vent'anni dopo la marcia su Roma di Mussolini. Poiché la marcia su Roma era avvenuta nel 1923, Eco nel 1943 aveva dieci anni. Di conseguenza oggi ha un'età che si aggira tra i settanta e gli ottant'anni.

Pur trattandosi di calcoli elementari, gli errori sono rilevanti. In generale si ha l'impressione che gli studenti, rilevati i dati del problema, operino a spanne dando al nostro autore di volta in volta età diverse. Il primo studente individua la data della marcia su Roma, poi ne ricava che Eco avesse dieci anni nel 1925.

[04130183] Umberto Eco dice di avere dieci anni quando andava a scuola sotto il regime fascista. Mussolini forma il primo governo nel 1922 da quel momento si costituisce il regime. Ipotizzo, quindi, che Eco aveva dieci anni intorno al 1925 e deduco la risposta.

I tre esempi che seguono muovono tutti dalla stessa data, ma i calcoli portano ad attribuire al nostro autore rispettivamente più di ottant'anni, 88 anni, 102 anni; nell'ultimo caso si somma al ragionamento sbagliato anche una cattiva interpretazione del testo.

[13340566] Umberto Eco afferma che a 10 anni sapeva chi era il primo ministro ai tempi della marcia su Roma. La marcia su Roma è avvenuta nel 1922, quindi Eco doveva essere nato circa 10 anni dopo, negli anni '30, facendo un rapido calcolo dovrebbe avere più di 80 anni.

[13340557] Eco afferma che all'età di 10 anni sapeva chi fosse il primo ministro 20 anni prima ovvero durante la marcia su Roma; se la marcia su Roma è avvenuta nel 1922 allora Eco dovrebbe avere 88 anni.

[14360609] La marcia su Roma fu nel 1922, lui dice che vent'anni prima sapeva chi era il primo ministro, quindi nel 1942 lui ha 30 anni. Infatti ora ne ha più di 80. Ora ha 102 anni.

Sono numerosi i casi in cui una semplice inversione fra addizione e sottrazione fa ringiovanire Eco di 20 anni, ecco un esempio in cui nonostante i riferimenti enciclopedici corretti il ragionamento inciampa sulla sottrazione.

[20490806] Quando Eco aveva dieci anni la marcia su Roma (1922) c'era stata vent'anni prima. $1922 + 20 = 1942 + 10 = 1952$.

Ecco degli esempi di ragionamento approssimativo che dimostrano chiaramente come gli studenti in questi casi non siano riusciti a compiere le inferenze necessarie per la piena comprensione del testo; il mancato compiersi del modello situazionale porta gli studenti a inserire nei ragionamenti elementi non pertinenti o del tutto inventati per giustificare le modalità con cui hanno risposto:

[20500825] Date le sue conoscenze molto ampie mi portano a pensare una vita nella quale ha potuto studiare in modo prolungato nel tempo.

[20500833] Eco dice che venti anni dopo la marcia su Roma aveva dieci anni. Dal testo emerge che al tempo in cui scrive esiste già internet.

[18520854] Umberto Eco specifica la propria età al momento dell'apprendimento di determinate informazioni a scuola (10 anni) e sottolinea il fatto che da allora siano trascorsi 50 anni.

Sono infine estremamente interessanti i casi in cui gli studenti, attraverso ragionamenti diversi rispetto a quanto da noi preventivato, riescono comunque a dare la risposta esatta. Gli studenti che rispondono con un ragionamento «diverso» sono l'11,7% dei casi, di questi, nel 43% dei casi (pari a 30 casi su 70) gli studenti danno una risposta esatta della domanda chiusa. Questi studenti non si riferiscono agli indizi testuali canonici, ne assumono altri meno rilevanti, a nostro parere, efficaci tuttavia ai fini della risposta all'item chiuso.

[22601001] Il fatto che faccia riferimento a Bufalo Bill mi fa pensare ad un periodo relativo di 50-60 anni fa, momento di grande risalto per il circo. Inoltre in quel periodo la tecnologia non era così avanzata.

[17430714] Dal testo fuoriesce una preparazione che appartiene ad un uomo che ha fatto esperienze.

[10240379] Se andava a scuola durante il regime fascista vuol dire che appartiene a due generazioni fa, quindi come i miei nonni, così mi sono ricavata una possibile età.

[1100146] Dalle poche volte che lo si vede dal vivo nelle trasmissioni televisive ne ricavo che è un uomo anziano, anche dal suo modo di scrivere, dandosi un tono di vecchio filosofo della polis greca, con più frasi retoriche ciceroniane, cerca di farsi dare più anni di quanti in effetti ne abbia.

[18460755] Sinceramente ho dedotto l'età di Umberto Eco non secondo un criterio ma ad occhio.

3.2. I profili di risposta e l'abilità di lettura

Confrontiamo ora i diversi stili utilizzati per rispondere alla domanda aperta con gli esiti alla domanda chiusa. Nella Tabella 5 è possibile osservare come si distribuiscono le diverse modalità di risposte rispetto alla categorizzazione delle risposte aperte da noi elaborata; a partire da sinistra possiamo seguire i diversi elementi della nostra check list:

- a-b) Individua il paragrafo e cita la data della Marcia su Roma.
- c) Compie il ragionamento necessario a risolvere l'inferenza sull'età di Eco.
- d) Calcola l'età di Eco.

Confrontando le modalità di risposta con gli esiti della risposta chiusa (**Quanti anni ha secondo te Umberto Eco?**) verifichiamo che la soluzione

del quesito chiuso viene raggiunta in misura maggiore da chi ha compiuto il ragionamento da noi previsto come completo e corretto, ma anche da studenti che seguono percorsi di tipo alternativo, sia in forma di ragionamento approssimativo, sia in forma di ragionamento diverso. In sostanza c'è una certa discrasia fra come lo studente ha risposto alla domanda chiusa sull'età di Eco e come invece ha giustificato il suo ragionamento, che spesso lo ha portato a esiti diversi. Nella parte destra della tabella abbiamo confrontato le diverse modalità di risposta con i punteggi ottenuti dagli studenti nella prova complessiva da cui sono tratte le due domande sull'età di Eco. La prova di lettura elaborata per la ricerca originaria era composta da tre tipologie di prove: quella da cui è tratta la domanda prevedeva sia domande di comprensione della lettura sia domande relative alle conoscenze enciclopediche: tali conoscenze non erano ricavabili dal testo ma erano necessarie a comprenderlo. La ricerca ha dimostrato come la disponibilità di tali conoscenze enciclopediche fosse la variabile con maggior potere esplicativo del punteggio al test, anche superiore alle variabili socioeconomiche di provenienza dello studente.

Nella nostra tabella possiamo confrontare i diversi profili di risposta con i punteggi relativi alla comprensione della lettura e al punteggio relativo alle conoscenze enciclopediche. Abbiamo visto come anche studenti che abbiano seguito ragionamenti diversi arrivano ugualmente alla risposta giusta, verificiamo quindi l'andamento di questo gruppo rispetto al punteggio ottenuto nel resto della prova; nel caso del punteggio ottenuto nelle conoscenze enciclopediche il loro punteggio è decisamente inferiore alla media, come a dimostrare l'esistenza di due modalità di risposta: o puntuale e basata sul possesso di nozioni enciclopediche, o di tipo diverso basata su elementi scollegati dal paragrafo da cui si dovevano trarre le informazioni. Nel caso del punteggio relativo alla comprensione della lettura, vediamo invece che gli esiti più alti sono piuttosto distribuiti fra i vari profili di risposta, senza che emerga una direzione evidente. Tale risultato è molto interessante perché ci fa riflettere sulle molteplici strade che il lettore compie basandosi su strategie affatto prevedibili e scontate. Filtrare il risultato della prova di comprensione della lettura attraverso l'analisi di questa domanda aperta fa emergere prepotentemente le molteplici vie che il lettore percorre per arrivare alla comprensione del testo.

Emerge dunque un doppio profilo di lettore, il lettore abile – colui che segue il ragionamento tracciato dal testo facilmente – e il lettore strategico – che coglie il senso finale adattando percorsi alternativi. Come affermano Afflerbach, Pearson e Paris (2008):

To characterize students as skilled readers is to recognize that they can orchestrate a wide array of processes to make reading work effortlessly. To characterize students as strategic readers is to recognize that they are flexible and adaptable to

particular circumstances, and, when the situation calls for it, they can select just the right strategy to overcome any temporary roadblock they might encounter.

Se osserviamo le correlazioni fra il risultato della correzione della risposta aperta con i due punteggi al test, possiamo verificare come le correlazioni – tutte significative – siano più alte nel caso del punteggio enciclopedico; la modalità di correzione della risposta aperta «premiava» la conoscenza dei riferimenti storici come la Marcia su Roma; la correlazione è molto meno forte nel caso del punteggio di comprensione della lettura; ciò evidenzia come un punteggio alto in comprensione della lettura non corrisponda necessariamente al ragionamento ordinato da noi ipotizzato, proprio del lettore abile, ma possa essere associato anche a strategie di lettura che permettono di risolvere la comprensione complessiva del testo tramite altri indizi testuali.

Tabella 5. – Correlazioni fra il risultato della risposta aperta e i punteggi alla prova.

		ECO CHIUSA	ENCICLOPEDICHE	COMPRENSIONE LETTURA
ECO APERTA	Tau_b di Kendall	,393**	,429**	,137**
	Sig. (2-code)	,000	,000	,000
	N	587	600	600

** La correlazione è significativa al livello 0,01 (2-code).

Il confronto fra risposte chiuse e aperte dimostra che anche attraverso i raffinati metodi statistici dell'item analisi e della statistica non è possibile determinare con esattezza la validità di una singola domanda, perché questi strumenti rivelano la loro validità solo su grandi numeri. A livello macro dunque la modalità a scelta multipla può essere usata con una certa confidenza su campioni cospicui, consapevoli che si tratta di una riduzione a fattore comune di una grande varietà di ragionamenti, che possono essere compresi appieno solo attraverso l'esplosione di questi stessi ragionamenti tramite la modalità aperta. Che vi sia una eterogeneità di esiti, delle contraddizioni o delle modalità alternative nella risoluzione delle risposte non è che una chiara trasposizione di quanto normalmente avviene al di fuori del mondo dei test carta e matita. Non staremo qui a dire l'importanza del pensiero creativo o del *lateral thinking*, ci interessa soltanto sottolineare che un test tenderà a considerare corretto quanto il creatore del test stesso ha considerato tale, pertanto questo ci riporta alla necessità di una competenza specifica nella costruzione di test, temperata dalla necessaria apertura e volontà di comprendere i processi di apprendimento piuttosto che dalla smania di valutazione.

4. CONCLUSIONI

Come abbiamo visto una sola domanda può aprire un complesso terreno di analisi alla domanda chiusa: infatti abbiamo rilevato il 56,3% di risposte esatte nella chiusa contro il 21,8% nella aperta. Certo la correzione della risposta aperta pretendeva che si seguisse un percorso. Un risultato, potremmo dire con Dewey, non è separabile dal percorso seguito per ottenerlo (Dewey, 1939), ma rimane viva anche la considerazione di Feyerabend che un risultato è comunque un risultato e che legarsi rigidamente a un metodo può limitare la costruzione di nuove interpretazioni dei fatti. Abbiamo visto che la soluzione corretta può venire da una somma di errori o da un approccio olistico al problema posto. Abbiamo anche visto che in generale esiste una relazione significativa tra ortodossia metodologica e risultati complessivi alla prova, ma anche che questa ortodossia spiega solo una parte minima della varianza tra le risposte.

Merita fermarsi sul fatto che escludere la componente enciclopedica quando si procede a verificare livelli di comprensione della lettura rappresenta una riduzione utile forse al confronto internazionale, ma non in grado di rendere conto degli effettivi problemi di comprensione soprattutto quando si considera l'interazione con testi complessi dove la lettura *top-down* può consentire salti nell'esame del testo il cui esito è protetto dalle competenze enciclopediche, modalità che probabilmente corrisponde alla lettura veloce di lettori esperti. Lo studente che ha risposto correttamente alla risposta chiusa e ha motivato la sua risposta affermando «Effettivamente perché conosco l'autore» e che ha avuto punteggi alti in tutte le prove è un esempio di questo modo di procedere.

La ricerca da cui abbiamo tratto lo spunto per questo lavoro aveva l'obiettivo di costruire prove valide per rilevare le abilità di lettura in uscita dai licei italiani, i testi proposti sono dunque significativamente diversi da quelli utilizzati nelle ricerche internazionali. In questa sede merita ricordare che il tema della validità di una prova rimane forse l'aspetto più serio sul quale procedere a confronto poiché stiamo trattando di un processo i cui contorni non sono riducibili al possesso di sole abilità procedurali.

La misura nelle scienze psicologiche e sociali e la costruzione dunque di prove valide è un esercizio concettualmente complesso; a fronte di strumenti di misura scientificamente definibili, e con un alto potere esplicativo da un punto di vista statistico, si riscontra che talvolta l'oggetto della misurazione, più che definito in sé, è definito dallo strumento stesso – profilato potremmo dire – a partire dal modo in cui lo stiamo esaminando. D'altra parte non esiste altro modo in cui possiamo verificare l'esistenza di abilità, come ad esempio la comprensione della lettura, se non attraverso prove che indirettamente ce ne

rivelino i tratti latenti. Se esista prima il fenomeno o lo strumento che lo rileva non deve quindi sembrare un problema ozioso a chi voglia trarre utili indicazioni dall'uso degli strumenti diagnostici sviluppati in ambito educativo.

In sostanza emerge la necessità di evitare conclusioni e di aprire nuovi percorsi di approfondimento, e dunque di tornare a una cautela scientifica che ci porti a anteporre lo sforzo di comprensione alla smania di trarre conclusioni, cioè di esprimere giudizi su risultati. Abbiamo premesso che questo non deve portarci a limitare le grandi ricerche, né a limitare l'uso delle prove a risposte chiuse, ma a comprendere che questi approcci devono sempre integrarsi con percorsi più articolati. L'uso stesso di domande aperte da poi ricodificare in chiuse in questo senso è solo una soluzione parziale. In sostanza per lavorare sulla comprensione è necessario un approccio comprensivo e non valutativo. Per concludere con Feyeraabend (1975):

L'unanimità di opinione può essere adatta per una chiesa, per le vittime bra-mose o atterrite di qualche mito (antico e moderno) e per i seguaci pronti di qualche tiranno. Per una conoscenza obiettiva è necessaria la varietà di opi-nione. E un metodo che incoraggi la varietà è anche l'unico metodo che sia compatibile con una visione umanitaria.

RIFERIMENTI BIBLIOGRAFICI

- Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher*, 61(5), 364-373.
- Bottani, N. (2011). L'ossessionante ripetizione dei test sulla comprensione della lettura. *Education 2.0*. <http://www.educationduepuntozero.it/didattica-e-apprendimento/ossessionante-ripetizione-test-comprensione-lettura-404092826.shtml>.
- Campbell, J. R., Kelly, D. K., Mullis, I. V. S., Martin, M. O., & Sainsbury, M. (2001). *PIRLS 2001 Assessment framework and specifications* (2nd ed., March).
- Cresswel, J., & Vayssettes, S. (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. OECD Pub.
- De Luca, M., & Lucisano, P. (2011). Item analisi fra modello e realtà. *Giornale Italiano della Ricerca Educativa*, 4, 85-96.
- Dewey, J. (1939). *Teoria della valutazione* (1ª ed.). Firenze: La Nuova Italia.
- Dochy, F. J. R. C. (1994). Prior knowledge and learning. In T. Husen & T. N. Postlethwaite (Eds.), *International encyclopedia of education* (2nd ed., pp. 4698-4702). Oxford - New York: Pergamon Press.
- Dochy, F. J. R. C., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowl-edge. *Review of Educational Research*, 69, Summer, 145-186.

- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge: EUA Massachusetts Institute of Technology.
- Feyerabend, P. K. (1975). *Contro il metodo. Abbozzo di una teoria anarchica della conoscenza*. Milano: Feltrinelli.
- Frederiksen, N. (1984). The real Test Bias. Influences of testing on teaching and learning. *American Psychologist*, 39(3), 193-202.
- Gallina, V. (2011). Il piacere di leggere. *Education 2.0*, 4 Marzo.
- Giannarelli, R. (1977). Indagine IEA e mastery learning. *Annali della Pubblica Istruzione*, 23, Marzo-Aprile.
- Gorman, T. P., Purves, A., & Degenhart, R. E. (Eds.). (1988). *The IEA study of written composition I: The international writing tasks and scoring scales*. Oxford: Pergamon Press.
- Invalsi (2011). *Quadro di riferimento della prova di italiano Sistema Nazionale di Valutazione*. http://www.invalsi.it/snv2012/documenti/QDR/QdR_Italiano.pdf.
- Katz, S., Lautenschlager, G. J., Blackburn, A. B., & Harris, F. H. (1990). Answering reading comprehension items without passages on SAT. *Psychological Science*, 1, 122-127.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163-182.
- Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. Cambridge: Cambridge University Press.
- Kintsch, W., & Dijck, T. A. van (1983). *Strategies of discourse comprehension*. New York: New York Academy Press.
- Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, 67(3), 271-299.
- Lucisano, P. (1989). *Lettura e comprensione*. Torino: Loescher.
- Lucisano, P. (2003). Validità e affidabilità delle pratiche valutative: a proposito del progetto pilota 2. *Cadmo*, 2, 37-57.
- Lumbelli, L. (1989). *Fenomenologia dello scrivere chiaro*. Roma: Editori Riuniti.
- Lumbelli, L. (1996). Focusing on text comprehension as a problem-solving task. A fostering project for culturally deprived children. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties. Processes and intervention* (pp. 301-330). Mahwah, NJ: Erlbaum.
- Lumbelli, L. (2009). *La comprensione come problema*. Roma - Bari: Laterza.
- McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, 22, 247-288.
- Moncada, S. M., & Moncada, T. P. (2010). Assessing student learning with conventional multiple-choice exams: Design and implementation considerations for business faculty. *International Journal of Education Research*, 5(2), 15-29.

- Mullis, I. V. S., Kennedy, A. M., Martin, M. O., & Sainsbury, M. (2006). *PIRLS 2006 Assessment framework and specifications* (2nd ed., February).
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., Trong, K. L., & Sainsbury, M. (2009). *PIRLS 2011 Assessment framework*. Chestnut Hill, MA.
- Organisation for Economic Co-operation and Development (2010). *PISA 2009 Assessment framework: Key competencies in reading, mathematics and science*. OECD Pub.
- Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, 25, 399-438.
- Pearson, P. D., Barr, R., Kamil, M. L., & Mosenthal, P. (Eds.). (1984). *Handbook of reading research*. New York: Longman.
- Pearson, P. D., & Fielding, L. (1991). Comprehension instruction. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research*, Vol. 2 (pp. 815-860). New York: Longman.
- Pearson, P. D., Garavaglia, D., Lycke, K., Roberts, E., Danridge, J., & Hamm, D. (1999). *The impact of item format on the depth of students' cognitive engagement*. Technical Report. Washington, DC: American Institute for Research.
- Pirsig, R. M. (1974). *Lo Zen e l'arte della manutenzione della motocicletta*. Milano: Adelphi.
- Postlethwaite, T. N., Husén, T., Laeng, M., Bellacicco, A., & Visalberghi, A. (1977). *Misurazione del rendimento scolastico: indagine IEA e situazione italiana*. Roma: Le Monnier.
- Postlethwaite, T. N., & Ross, K. N. (1992). *Effective schools in reading: Implications for educational planners. An exploratory study. The IEA Study of Reading Literacy II*. Hambourg: IEA.
- Pressley, M., Ghatala, E. S., Woloshyn, V., & Pirie, J. (1990). Sometimes adults miss the main ideas and do not realize it: Confidence in responses to short-answer and multiple-choice comprehension questions. *Reading Research Quarterly*, 25(3), Summer, 232-249.
- Purves, A. C. (Ed.). (1992). *The IEA study of written composition II: Education and performance in fourteen countries*. Oxford: Pergamon Press.
- Rauch, D. P., & Hartig, J. (2010). Multiple-choice versus open-ended response formats of reading test items: A two-dimensional IRT analysis. *Psychological Test and Assessment Modeling*, 52(4), 354-379.
- Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal*, 41(1), 159-189.
- Topping, K., Valtin, R., Roller, C., Brozo, W., & Lourdes Dionosos, M. (2003). *Policy and practice implications of the Program for International Student Assessment (PISA) 2000*. Report of the International Reading Association PISA Task Force.

- Trivellato, U., & Zuliani, A. (1977). *Le determinanti della riuscita scolastica in Italia: alcuni primi risultati dell'indagine IEA, con particolare riferimento all'influenza dell'ambiente familiare e dei fattori scolastici*. Firenze: Le Monnier - Cede.
- Warwick, J., Bush, M., & Jennings, S. (2010). Analysis and evaluation of liberal (free-choice) multiple-choice tests. *Innovations in Teaching & Learning in Information & Computer Sciences*, 9(2), 51-62.

RIASSUNTO

La ricerca prende le mosse da un lavoro più ampio che mira a validare e tarare una prova di comprensione della lettura che ha come popolazione di riferimento gli studenti dell'ultimo anno di liceo; il lavoro si pone come stimolo di riflessione sull'uso delle prove strutturate per la verifica della comprensione della lettura, in particolare analizza il rapporto esistente fra diverse modalità di risposta ad una stessa domanda (risposta aperta versus risposta multipla). Attraverso l'esame di una domanda aperta, confrontata con i risultati del test complessivo, si discute della possibilità di verificare i processi mentali sottesi alla lettura e i loro rapporti con le conoscenze enciclopediche degli studenti. Moltissimi studi hanno dimostrato che le conoscenze enciclopediche sono uno dei fattori che maggiormente spiega la padronanza in lettura (dal 20% al 70% della varianza nei test), e non considerarlo può portare a forti effetti distorsivi nelle misure. Emerge che oltre al rigore metodologico necessario a costruire test validi e affidabili, è necessario un approccio che cerchi di comprendere e di considerare più a fondo tutti i percorsi che il lettore strategico e quello abile mettono in atto per afferrare compiutamente un testo.

Parole chiave: Conoscenze enciclopediche, Comprensione della lettura, Prove di verifica con risposta aperta, Prove di verifica con risposta multipla, Validità.