





UNIVERSITÀ DEGLI STUDI DI ROMA "LA SAPIENZA" - FACOLTÀ DI FILOSOFIA  
DOTTORATO DI RICERCA CONSORTILE IN PEDAGOGIA SPERIMENTALE  
TESI DI DOTTORATO



DOTTORANDO  
Giorgio Asquini

TUTORS  
Prof.ssa Maria Lucia Giovannini  
Prof. Pietro Lucisano

## *ACCEDERE ALL'UNIVERSITÀ*

*Indagine empirica sul trattamento dei dati della prova  
di ingresso al Corso di laurea in Scienze dell'Educazione  
e della Formazione di Roma "La Sapienza"*

Anno Accademico  
2004 – 2005  
*XVIII Ciclo*

Composizione grafica a cura dell'Autore

## Indice

Premessa .....	IX
Introduzione .....	XIII
Capitolo primo – Il problema delle prove per l’accesso pro- grammato all’università.....	3
1.1. Osservazione preliminare.....	3
1.2. Situazione problematica: accedere all’Università .....	5
1.3. Situazione problematica: come sono fatte le prove per l’accesso.....	8
1.4. Un problema nel problema: la scala dei punteggi .....	15
1.5. Conclusioni per avviare il lavoro di indagine .....	22
Capitolo secondo – Ipotesi di ricerca .....	25
2.1. Dalla situazione problematica all’ipotesi di ricerca .....	25
2.2. Disegno di ricerca .....	26
2.3. Fasi della ricerca.....	27
2.4. Campione di ricerca .....	30
2.5. Strumenti e metodologie di ricerca .....	30
Capitolo terzo – Le diverse modalità di item analisi: confronto fra il modello classico e il modello IRT .....	33
3.1. Valutare la qualità dei quesiti di una prova .....	33
3.2. Dall’item analisi classica al modello IRT .....	36
3.3. Confronto fra item analisi classica e IRT .....	42
3.4. Dall’item analisi ai punteggi degli studenti.....	51
3.5. Il trattamento delle omissioni nel modello IRT .....	60
Capitolo quarto – Gli strumenti di rilevazione .....	65
4.1. La validità della prova .....	65
4.2. La selezione dei materiali .....	67
4.3. La realizzazione della prova pilota .....	70
4.4. Item analisi della prova pilota (TCL) .....	72
4.5. Allestimento degli strumenti per lo studio principale .....	77
4.5.1. Prova di comprensione della lettura con quesiti a scel- ta multipla TCL .....	78
4.5.2. Questionario Punti di vista sulla scelta universitaria .....	81
4.5.3. Prova di comprensione della lettura Cloze .....	82
4.5.4. Questionario nuovi iscritti .....	85
4.6. La scheda di rilevazione del percorso accademico .....	87

Capitolo quinto – Risultati della prova di ingresso .....	91
5.1. La somministrazione della prova di ingresso .....	91
5.2. Verifica del campione partecipante alla prova di ingresso .....	95
5.2.1. Caratteristiche anagrafiche .....	95
5.2.2. Esperienze di studio .....	96
5.2.3. Contesto familiare .....	100
5.2.4. Rapporto con l'università .....	103
5.2.5. Scelta del corso di studi .....	107
5.2.6. Autovalutazione in lingua straniera .....	108
5.3. Item analisi del TCL .....	110
5.4. Item analisi del Cloze .....	118
5.5. Analisi dei risultati delle prove degli studenti di Roma .....	128
5.5.1. Le graduatorie della prova di comprensione TCL .....	129
5.5.2. Le graduatorie della prova di comprensione Cloze .....	136
5.5.3. Le relazioni fra le prove di comprensione TCL e Cloze .....	143
5.5.4. Le relazioni fra le diverse graduatorie prodotte dalle analisi degli item .....	147
5.6. Graduatorie di accesso secondo i modelli classico e IRT .....	150
Capitolo sesto – Verifica dell'ipotesi .....	167
6.1. La raccolta dei dati riguardanti gli esiti accademici del pri- mo anno di corso .....	167
6.2. Verifica del campione sul quale sono stati rilevati gli esiti accademici .....	168
6.3. Descrizione del campione definitivo secondo le variabili della scheda di rilevazione degli esiti accademici .....	176
6.4. Relazioni esistenti fra i risultati della prova di ingresso e gli esiti accademici .....	185
6.5. Confronto fra i sottogruppi determinati dalla prova di in- gresso .....	189
6.6. Verifica dell'ipotesi di ricerca .....	193
Capitolo settimo – Conclusioni e prospettive di ricerca .....	195
7.1. Per una selezione consapevole, se proprio necessaria .....	195
7.2. Quando la prova è diventata veramente selettiva .....	198
Bibliografia .....	205

*Alla mia famiglia*

*che ha atteso con pazienza  
che questo lavoro finisse.*





## Premessa

Ho il privilegio di partire da un ricordo personale. Nel 1993, durante un incontro per un gruppo di insegnanti di scuola secondaria sul tema della dispersione scolastica, il prof. Visalberghi citava i risultati dell'indagine IEA-SAL, che si era da poco conclusa. Commentando le scale secondo cui venivano espressi i risultati Visalberghi accennava al metodo di calcolo dei punteggi, basato su una nuova modalità di analisi delle prove proposta da uno studioso danese, Georg Rasch. Non si trattava certo del contesto adatto per entrare nel dettaglio psicometrico, ma ricordo la sua annotazione circa l'opportunità di continuare a studiare la cosa per verificare se ci potessero essere sviluppi utili per la didattica quotidiana.

L'accento mi colpì per due motivi: in primo luogo per l'esempio metodologico, che solo un grande studioso che aveva attraversato già allora quasi 50 anni della ricerca pedagogica italiana poteva offrire, cioè la necessità costante di confrontarsi con le novità, senza accontentarsi di percorrere sempre strade conosciute; in seconda battuta, proiettando l'osservazione sul mio mestiere di allora (insegnante di scuola media), illuminò di una luce nuova il problema quotidiano di dare i punti alle prove dei miei studenti e di trasformare questi punti in voti.

D'altra parte il tema della misurazione in un processo valutativo è stata una delle costanti nel percorso pedagogico di Visalberghi<sup>1</sup>, che ha influenzato il lavoro di chi ha avuto la possibilità, come il sottoscritto, di collaborare con il CEDE negli anni in cui è stato da lui presieduto e con il Dipartimento di Ricerche Storico-Filosofiche e Pedagogiche dell'Università di Roma "La Sapienza", di cui resta ancora uno dei docenti più prestigiosi.

Da allora però il tema di questo nuovo modo di elaborare i punteggi di un prova è stato sollevato solo in occasione della presentazione dei

<sup>1</sup> Sempre a proposito delle indagini internazionali basta ricordare il capitolo di Visalberghi *La misurazione all'interno della valutazione scolastica*, che conclude il volume di presentazione delle indagini IEA-IPS e IEA-SAL (Corda Costa e Visalberghi 1995).

risultati delle indagini internazionali, dove ormai è diventato uno standard. Per cui quando si è trattato di definire, all'interno del Dottorato di Ricerca Consortile in Pedagogia Sperimentale, un percorso di indagine, in accordo con il Collegio, mi è sembrato di cogliere a distanza di dieci anni lo spunto involontariamente offertomi da Visalberghi, anche se calato in un contesto per me del tutto nuovo, quale quello della selezione universitaria in accesso.

In effetti uno dei motivi materiali che hanno suggerito il percorso di ricerca è la particolare situazione che si è creata nel Corso di laurea in Scienze dell'Educazione e della Formazione dell'Università di Roma "La Sapienza". Partito con molte perplessità nell'A.A. 2001-02, in particolare per i dubbi esistenti fra i docenti circa l'effettiva attrattiva presso gli studenti di un corso di laurea così specifico e vago allo stesso tempo, ma sfruttando le novità introdotte dalla riforma degli ordinamenti prevista dal DPR 509/99 (quindi senza la necessità di riconvertire schemi di laurea precedenti), il corso di laurea ha incontrato un favore crescente in termini di iscrizioni, passando dai 275 nuovi iscritti del primo anno ai 357 dell'A.A. 2002-03. Il trend di crescita, che ha sorpreso non poco docenti e collaboratori del corso di laurea, ha sollecitato una riflessione sull'opportunità di contenere negli anni a venire il numero delle iscrizioni, per mettersi nelle condizioni di offrire un servizio didattico valido, considerando la difficile situazione strutturale in cui si trovava, e si trova ancora, il corso di laurea.

La riflessione avviata non ha però permesso di introdurre l'accesso a numero programmato per il terzo anno di vita del corso di laurea, ma considerata la delicatezza del problema si è cercato di affrontare la scomoda necessità nel modo migliore. Di fronte alla probabile esplosione di nuovi iscritti per l'A.A. 2003-04, il Dottorato di Ricerca Consortile in Pedagogia sperimentale, afferente al Dipartimento di riferimento del corso di laurea, è stato sollecitato a studiare il problema dell'accesso programmato, per affrontare in modo corretto, dall'anno accademico successivo, l'introduzione della selezione in ingresso. La previsione si è rivelata esatta: gli iscritti nell'A.A. sono stati 516 e dall'A.A. 2004-05 è stato introdotto per il corso di laurea l'accesso a numero programmato, confermato anche l'anno successivo ed esteso anche al corso di laurea specialistica collegato.

La necessità di riflettere sulle prove di accesso si è concretizzata, all'interno del Dottorato di ricerca con diversi percorsi di approfondimento, che vedono impegnate anche le colleghe Truffelli e Di Donato. Il mio campo indagine specifico riguarda il modo in cui vengono elaborati

i punteggi delle prove per definire la graduatoria di accesso, in particolare quale modalità di trattamento dei punteggi risulti più efficace in relazione alla funzione selettiva della prova. Per far questo si rende necessario riflettere sui diversi modelli di analisi dei quesiti che compongono una prova, la cosiddetta *item analisi*<sup>2</sup>, considerando l'opportunità di utilizzare, per la definizione delle graduatorie di accesso, i punteggi scaturiti dall'analisi IRT (*Item Response Theory*).

Per lo svolgimento di questo lavoro mi sono avvalso delle ancora deboli strutture del nuovo Corso di laurea in scienze dell'Educazione e della Formazione dell'Università di Roma "La Sapienza", in particolare della fattiva collaborazione dei docenti, del personale amministrativo, degli studenti e dei miei colleghi di dottorato. Senza il loro contributo nelle diverse fasi della ricerca difficilmente questo lavoro sarebbe giunto al termine.

<sup>2</sup> Si è scelto di utilizzare sempre nel corso del rapporto l'espressione "item analisi", attestata in diverse pubblicazioni italiane (Lucisano 1989, Benvenuto, Lastrucci e Salerno 1995) in luogo dell'originale inglese "*Item Analysis*".



## Introduzione

Questo rapporto di ricerca si apre (capitolo primo) con una presentazione della normativa di riferimento che regola lo svolgimento delle prove di accesso ai corsi a numero programmato nelle università italiane e con una panoramica delle modalità di realizzazione delle prove di accesso attualmente in uso nel sistema universitario italiano, in particolare per quanto riguarda le diverse tipologie di prove strutturate a risposta chiusa utilizzate e i criteri di attribuzione dei punteggi per le risposte corrette.

Sulla base alla situazione problematica rilevata nella panoramica delle prove di accesso vengono definiti i parametri di svolgimento del lavoro di ricerca (capitolo secondo), con la definizione dell'ipotesi di ricerca e degli aspetti legati ai tempi di svolgimento, al campione di riferimento e agli strumenti utilizzati.

Poiché l'obiettivo principale dell'indagine è l'analisi delle differenze nelle graduatorie prodotte dai diversi tipi di item analisi, il terzo capitolo è dedicato all'approfondimento teorico delle diverse modalità di analisi dei quesiti delle prove, ripercorrendo l'evoluzione dei diversi modelli e approfondendo in particolare i criteri di formazione del punteggio ottenuto dai rispondenti. Una particolare attenzione è dedicata al tema delle omissioni di risposta, che risultano una delle variabili più delicate nel trattamento dei punteggi secondo l'analisi IRT.

Il quarto capitolo è dedicato alla presentazione degli strumenti utilizzati per l'indagine, con una prima parte dedicata allo svolgimento della prova pilota che ha permesso il collaudo di alcuni materiali, seguita da una presentazione specifica dei tre strumenti, due prove di comprensione della lettura e un questionario, utilizzati nella somministrazione principale. Infine viene presentata anche la scheda per la rilevazione del percorso accademico, destinata al controllo di alcune variabili legate al successo accademico e rilevate al termine del primo anno di corso.

L'ampio quinto capitolo dà spazio all'analisi dei risultati della prova di ingresso, approfondendo in particolare gli aspetti legati alla simulazione di una situazione di accesso alla frequenza universitaria. Dopo una preventiva verifica delle caratteristiche del campione raggiunto si

passa alle diverse procedure di item analisi svolte sui dati delle prove di comprensione. Infine vengono considerate le differenze fra i diversi tipi di graduatoria dei punteggi prodotti dalle item analisi, con particolare attenzione alle situazioni di alcuni studenti che presentano profili di risposta anomali.

Il sesto capitolo è dedicato alla verifica dell'ipotesi. Prima di tutto viene analizzato il gruppo di studenti in uscita, di cui si sono rilevati gli esiti alla fine del primo anno di corso, quindi si passa alla verifica delle correlazioni esistenti fra le diverse graduatorie derivate dalla prova di ingresso/accesso e gli esiti accademici. Infine vengono messi a confronto i risultati dei due gruppi di studenti (ammessi/esclusi) determinati dalla simulazione.

Nel settimo capitolo è proposta una riflessione parallela sulla necessità e sull'opportunità di costruire prove di selezione, indicando i possibili sviluppi del presente lavoro. Infine sono presentati i principali risultati relativi all'uso effettivo della prova di comprensione della lettura in un reale contesto selettivo.

Completa il volume una bibliografia aggiornata, suddivisa secondo i principali temi trattati nel lavoro di indagine.

*A frequentare il brigante, dunque, Cosimo aveva preso una smisurata passione per la lettura e per lo studio, che gli restò poi per la vita. L'atteggiamento abituale in cui lo si incontrava adesso, era con un libro aperto in mano, seduto a cavalcioni d'un ramo comodo, oppure appoggiato a una forcella come a un banco da scuola, un foglio posato su una tavoletta, il calamaio in un buco dell'albero, scrivendo con una lunga penna d'oca.*

Italo Calvino  
*Il barone rampante (1957)*





# Capitolo primo

## Il problema delle prove per l'accesso programmato all'università

### 1.1. Osservazione preliminare

Nel campo dell'istruzione e della formazione si è molto diffuso nell'ultimo decennio l'uso di prove strutturate a risposta chiusa. Hanno contribuito a questa diffusione le diverse indagini internazionali a cui ha partecipato l'Italia, basate soprattutto su quesiti a scelta multipla. Ricordiamo le principali indagini IEA<sup>3</sup> (*Civic*, sulle conoscenze in educazione civica, TIMSS sulle conoscenze in matematica e scienze, PIRLS sulla comprensione della lettura dei bambini di nove anni) ed OCSE<sup>4</sup> (IALS-SIALS sulle competenze linguistiche degli adulti, PISA sulle competenze degli studenti quindicenni): le istituzioni responsabili delle indagini, il largo numero di scuole coinvolte e la diffusione dei risultati di queste comparazioni hanno permesso di sollecitare il dibattito su test e questionari, ed in particolare sulle modalità di costruzione dei materiali.

Non vanno poi dimenticate le esperienze del Seris (Servizio Rilevazioni di Sistema, promosso dal CEDE<sup>5</sup>), che per tre anni ha realizzato indagini sulle principali aree di conoscenze in quasi 2000 scuole dei tre livelli scolastici, sempre utilizzando prove strutturate e comunicando alle

<sup>3</sup> La IEA (*International Association for the Evaluation of Educational Achievement*), è il principale organismo internazionale che ha realizzato, a partire dalla fine degli anni Cinquanta, indagini comparative internazionali sul profitto degli studenti.

<sup>4</sup> L'OCSE (Organizzazione per la Cooperazione e lo Sviluppo Economico), è un organismo intergovernativo che ha realizzato negli ultimi 10 anni alcune importanti indagini comparative internazionali, finalizzate a raccogliere dati per alimentare il raffronto periodico di indicatori dell'istruzione (OCSE 2005).

<sup>5</sup> Il CEDE (Centro Europeo Dell'Educazione), è stato l'organismo che ha realizzato a livello nazionale le principali indagini IEA e OCSE. Dal 2001 si è trasformato nell'INValSI (Istituto Nazionale per la Valutazione del Sistema dell'Istruzione), che ha conservato la tradizionale funzione di referente locale per le indagini internazionali.

single scuole i risultati, e soprattutto dell'istituzione della terza prova dell'Esame di Stato, la cui modalità più utilizzata è stata quella basata su domande a scelta multipla<sup>6</sup>.

In questo quadro di progressiva abitudine all'uso di prove strutturate va letta anche la crescente diffusione di prove d'ingresso per l'università, non sulla spinta di stringenti obblighi istituzionali (come per la terza prova dell'Esame di Stato), ma in genere con lo scopo di acquisire informazioni sulla preparazione delle aspiranti matricole a fini orientativi<sup>7</sup>. La raccolta di informazioni può essere finalizzata all'analisi dei prerequisiti, per poter colmare eventuali lacune di preparazione nel settore scelto, o all'orientamento, per raccogliere informazioni utili alla definizione di un piano di studio individuale.

Accanto a questo intento squisitamente didattico, finalizzato in modo esplicito al miglioramento dell'offerta formativa delle università, si è però diffuso un uso selettivo delle prove strutturate, finalizzato al controllo degli accessi in facoltà o corsi di laurea dove il numero dei posti disponibili, in relazione alle risorse strutturali esistenti, risulta essere necessariamente limitato.

E' evidente che il doppio uso delle prove strutturate, secondo le finalità appena indicate, comporta sostanziali differenze nella costruzione delle stesse e soprattutto nella gestione dei risultati (Andreani Dentici e Amoretti 2000), in particolare perché le prove selettive assumono aspetti legali che comportano forti elementi di riservatezza, riguardanti sia il contenuto delle prove sia i risultati dei rispondenti, che in base a quelle prove si sono visti negare, o permettere, l'accesso a un corso di studi. Questo aspetto, sicuramente essenziale, può nell'ambito di questo studio essere mantenuto sullo sfondo, considerando anche il lavoro parallelo svolto in un'altra ricerca di dottorato, dove vengono approfonditi anche gli aspetti giuridici delle prove d'ingresso all'università, confrontando la situazione italiana con esperienze di altri paesi<sup>8</sup>.

<sup>6</sup> Cfr. Bolletta (2000), che analizza anche le modalità di attribuzione dei punteggi utilizzate dalle commissioni degli esami di stato.

<sup>7</sup> L'art 3 del DM 245/97 prevede che "Le università, di norma prima dell'inizio dei corsi ufficiali e in relazione ad uno o più corsi di laurea, organizzano attività di orientamento e insegnamento, le quali comprendono i contenuti caratterizzanti, le conoscenze generali e propedeutiche, forme di tutorato e di assistenza agli studenti, nonché test autovalutativi. Tali attività si concludono con una valutazione finale, non condizionante l'iscrizione".

<sup>8</sup> Si tratta dell'indagine svolta dalla collega Di Donato che, svolgendosi nel XIX ciclo di dottorato, sarà completata l'anno successivo a questo lavoro.

## 1.2. Situazione problematica: accedere all'Università

La realizzazione di rilevazioni nazionali ed internazionali, realizzate secondo le più avanzate specifiche sulla costruzione di prove oggettive, dovrebbero costituire un naturale modello di riferimento per chi si occupa di istruzione e formazione e vuole utilizzare prove standardizzate di questo tipo.

Le prove utilizzate nelle università sono invece costruite singolarmente dagli atenei (se non da facoltà o dipartimenti specifici), in mancanza di uno o più modelli istituzionalmente riconosciuti<sup>9</sup>. Quindi ogni realtà universitaria ha mano libera nel gestire l'accesso.

In origine il numero chiuso per l'accesso all'università era previsto, su scala nazionale, solo per quei corsi di laurea in cui la frequenza di attività di laboratorio limitasse forzatamente il numero dei partecipanti<sup>10</sup>, ma già la legge 341 del 1990, nell'ambito della autonomia didattica, permette ai singoli atenei, rispettando una serie di indicazioni nazionali del ministero che di fatto estendono il criterio interpretativo della disponibilità delle strutture previsto in precedenza per i soli laboratori, di limitare l'accesso ad alcune facoltà o corsi di laurea<sup>11</sup>. Tra i principi ispiratori stabiliti a livello centrale per definire, a livello locale, l'accesso ai corsi di laurea il ministero cita esplicitamente il possesso delle caratteristiche più adatte alla riuscita negli studi che conducono a una professione<sup>12</sup>.

Sullo svolgimento delle prove di accesso, il ministero rivendica a sé (DM 245/97 artt. 4 e 5) il compito di definire le modalità di espletamento delle procedure, ma la responsabilità delle prove, allestimento e defini-

<sup>9</sup> L'articolo 6 del DPR 519/99, circa l'ammissione ai corsi di laurea (per la quale resta obbligatorio il titolo di studio secondario superiore o equivalente), chiama in causa direttamente i singoli atenei, che "nei regolamenti didattici definiscono le conoscenze richieste per l'accesso e ne determinano, ove necessario, le modalità di verifica".

<sup>10</sup> E' il caso per esempio della Facoltà di Medicina, il cui accesso è regolato con il DPR 95 del 1986.

<sup>11</sup> La norma viene sfruttata da molte facoltà, soprattutto nell'area medica (Bini e Borsese 1995).

<sup>12</sup> In tal senso va interpretato uno dei criteri dell'art. 4 comma 1 del DM 245 /97 che attua le indicazioni della legge 341/90: "da valutarsi per le determinazioni di limitazione degli accessi all'istruzione universitaria... il carattere specialistico e direttamente professionalizzante di determinati corsi". Il Decreto è stato successivamente modificato dal DM 235/99 che estende esplicitamente la limitazione dell'accesso ai corsi di Scienze della Formazione primaria e a tutti i corsi di nuova istituzione.

zione della graduatoria, resta compito degli atenei, nel quadro delle attività di orientamento previste di norma per il periodo delle preiscrizioni e in vista della prossima attuazione dell'autonomia universitaria, come indicato esplicitamente nelle premesse del Decreto Ministeriale.

Un'eventuale specificazione del contenuto delle prove selettive alla conclusione delle attività di orientamento e insegnamento e di quelle relative alla ammissione ai corsi di diploma si porrebbero in contrasto con l'autonomia universitaria, dovendosi sviluppare su di esse una necessaria collaborazione e dialogo tra Ministero e atenei.

E' solo nell'agosto del 1999 che il tema dell'accesso programmato all'università viene affrontato in un provvedimento legislativo *ad hoc*, la legge 264/99 che razionalizza i diversi provvedimenti intercorsi negli anni Novanta, definendo gli ambiti di attuazione dell'accesso programmato ai corsi di laurea<sup>13</sup>, distinguendo due livelli di programmazione.

#### Art. 1

Sono programmati a livello nazionale gli accessi

a) ai corsi di laurea in medicina e chirurgia, in medicina veterinaria, in odontoiatria e protesi dentaria, in architettura ...

b) ai corsi di laurea in scienza della formazione primaria ...

...

e) ai corsi universitari di nuova istituzione o attivazione, su proposta delle università ...

#### Art.2

Sono programmati dalle università gli accessi

<sup>13</sup> Per praticità si riportano solo le indicazioni riferite ai corsi di laurea, ma le indicazioni della legge 264/99 si estendono anche ai diplomi universitari e alle scuole di specializzazione.

a) ai corsi di laurea per i quali l'ordinamento didattico preveda l'utilizzazione di laboratori ad alta specializzazione, dei sistemi informatici e tecnologici o comunque di posti-studio personalizzati;

...

c) ai corsi o alle scuole di specializzazione individuate dai decreti attuativi delle disposizioni di cui all'articolo 17, comma 95 della legge 15 maggio 1997, n. 127 e successive modificazioni.

[La legge 127/97 "Misure urgenti per lo snellimento dell'attività amministrativa e dei procedimenti di decisione e di controllo", nel comma indicato, permette agli atenei di disciplinare, per mezzo dei regolamenti di ateneo, l'ordinamento dei corsi di studio. N.d.A]

...

#### Art. 4

L'ammissione ai corsi di cui agli articoli 1 e 2 è disposta dagli atenei previo superamento di apposite prove di cultura generale, sulla base dei programmi della scuola secondaria superiore, e di accertamento della predisposizione per le discipline oggetto dei corsi medesimi, con pubblicazione del relativo bando almeno 60 giorni prima della loro effettuazione, garantendo altresì la comunicazione dei risultati entro i successivi 15 giorni. Per i corsi di cui all'articolo 1, comma 1, lettere a) e b), il Ministro dell'università e della ricerca scientifica e tecnologica determina con proprio decreto modalità e contenuti delle prove di ammissione, senza oneri aggiuntivi per il bilancio dello Stato.

Per capire il modo in cui il Ministro determini modalità e contenuti possiamo citare l'art.2 del contemporaneo Decreto MURST del 27 luglio 1999, per l'accesso ai corsi di laurea in Scienze della formazione primaria:

L'ammissione degli studenti avviene per concorso pubblico effettuato secondo criteri e modalità definiti dalle Università, che prevedono una graduatoria determinata per 80 punti su 100 da una o più prove di esame e per 20 punti dalla valutazione dei titoli indicati nel bando. Le prove di esame si effettueranno il 16 settembre 1999.

Dunque anche le modalità, a parte la data della prova, e i criteri vengono devoluti, per cui si conferma la mano libera per gli atenei sulle prove di accesso ai corsi di laurea<sup>14</sup>, richiamando il generico criterio di “accertamento della predisposizione per le discipline oggetto dei corsi”.

Negli anni successivi il Ministero conserva il controllo al centro delle prove solo per gli indirizzi di area medico-veterinaria, e anche se prevede delle tabelle di riferimento per i contenuti, conferma la predisposizione delle prove a cura delle singole università<sup>15</sup>. La disciplina dell'accesso programmato trova sistemazione definitiva negli articoli 35-36-37-38 del Testo Unico sull'Università (20 febbraio 2001), confermando il contenuto della legge 264/99.

Ma i pacchetti di prove utilizzati per l'accesso all'università sono coerenti agli scopi selettivi che gli atenei si propongono? Si tratta evidentemente da una parte di un problema di validità del contenuto (le domande della prova misurano effettivamente l'abilità oggetto dichiarata?), dall'altra di validità di costruito (i risultati del test sono coerenti con le finalità e le aspettative di chi lo ha costruito?), senza dimenticare poi il possibile collegamento con altre misure riguardanti gli aspiranti iscritti, come il voto conseguito nell'Esame di Stato, collegamento che pone problemi di validità di criterio<sup>16</sup>.

### 1.3. Situazione problematica: come sono fatte le prove per l'accesso

Una panoramica di alcune informazioni riguardanti le prove di accesso, tratte dai bandi di attuazione delle stesse, permette di capire quali sono attualmente gli orientamenti delle università che attuano prove di accesso, permettendo anche di capire quali siano i possibili problemi di validità della prova.

<sup>14</sup> Viene però confermato il controllo centralizzato per le facoltà indicate nel punto a) dell'Art. 1 della legge 264/99. Maggiore controllo anche per le scuole di specializzazione: per l'accesso alle SSIS il Decreto MURST 7 giugno 2000 prescrive il numero e la tipologia dei quesiti, la durata della prova, il punteggio massimo previsto. La costruzione della prova e l'attribuzione dei punteggi resta però di competenza dei singoli atenei.

<sup>15</sup> Decreto MURST 2 luglio 2001, che però, nell'art. 6, prevede anche i criteri di attribuzione del punteggio per le domande della prova.

<sup>16</sup> Per i diversi problemi di validità che si possono presentare per una prova vedi Lucisano e Salerno (2002).

Partiamo dalla prova di uno degli indirizzi sulle quali il Ministero ha competenza di indirizzo, pur lasciando alle università la responsabilità di predisporre la prova.

**Università di Bologna - Corso di laurea triennale in Attuazione e gestione del progetto in Architettura (2003-04)**

**5) MODALITA' DI SVOLGIMENTO DELLA PROVA DI AMMISSIONE E PROGRAMMA DI ESAME**

La prova di ammissione consiste nella soluzione di ottanta quesiti a risposta multipla, di cui una sola risposta esatta tra le cinque indicate, su argomenti di:

- logica e cultura generale - storia - disegno e rappresentazione - matematica e fisica

Sulla base dei programmi sottoriportati, vengono predisposti ventisei quesiti per l'argomento di logica e cultura generale e diciotto per ciascuno dei restanti argomenti.

La prova di ammissione ha inizio alle ore 11. Per lo svolgimento della prova è assegnato un tempo di due ore e quindici minuti.

Le informazioni sulla prova ricalcano le indicazioni ministeriali, ma bisogna osservare che al bando sono stati allegati dei sintetici programmi per le quattro aree indagate, con le indicazioni dei contenuti di riferimento. Riportiamo a titolo esemplificativo il programma della prima area.

**Logica – Cultura generale**

Le prove sono mirate ad accertare le capacità di analizzare un testo sul piano lessicale, sintattico e logico; interpretare, riformulare e connettere le informazioni fornite; elaborare correttamente inferenze, implicazioni, conclusioni, scartando procedure ed esiti errati, arbitrari o non giustificati rigorosamente. I quesiti verteranno su testi di saggistica scientifica o narrativa di autori classici o contemporanei, oppure su testi di attualità comparsi su quotidiani o su riviste generalistiche o specialistiche; verteranno altresì su casi o problemi, anche di natura astratta, la cui soluzione richiede l'adozione di forme diverse di ragionamento logico. Quesiti relativi alle conoscenze di cultura generale completano questo ambito valutativo.

Le indicazioni lasciano intuire che alcuni quesiti siano riferiti alla comprensione di testi, oppure ad alcune situazioni stimolo molto simili a quelle previste nelle indagini internazionali, mentre altri quesiti risultino sostanzialmente di tipo enciclopedico.

Passiamo a esempi di prove di accesso relative a Corsi di laurea gestite interamente dalle università, non rientrando negli indirizzi di studio che la legge 264/99 assegna come competenza al Ministero.

**Università di Milano "Bicocca" – Corso di laurea in Scienze della Comunicazione (2002-03)**

La prova di ammissione per l'accesso al Corso di laurea in Scienze della Comunicazione avrà una durata massima di 1 ora e 15 minuti.

*Materie d'esame:* la prova di selezione consisterà nella soluzione di quesiti a risposta multipla volti ad accertare le capacità logiche e numeriche, le competenze linguistiche e la comprensione di testi.

Il bando di Milano Bicocca fornisce pochissime informazioni sulla prova, con una semplice indicazione di durata massima e indicazioni generiche sugli argomenti della prova e il tipo di domande previste, per le quali non si specifica neanche il numero complessivo. Si tratta di un bando indicativo per le prove di accesso costruite nei primi anni di autonomia delle università sulle prove di accesso. Ma ancora oggi vengono pubblicati dei bandi molto generici sui contenuti della prova.

**Università di Roma "La Sapienza" – Corsi di laurea in Scienze e tecniche psicologiche (2005-06)**

7. Argomenti della prova e criteri di valutazione

La prova consiste nel rispondere a 90 domande. Le domande proposte sono di tre aree:

- area 1: Comprensione della lettura
- area 2: Ragionamento matematico
- area 3: Ragionamento logico



Sicuramente la cultura popolare diffusa fra gli studenti che aspirano all'accesso dei corsi di psicologia può sopperire alla mancanza di informazione ufficiale più specifica su cosa si intende per "Comprensione della lettura", ma certo per chi affronta la prova risulterebbe più confortante, e motivante, sapere su che cosa si sarà chiamati a dimostrare le proprie capacità, soprattutto se in altra parte dello stesso bando sono precisate, talvolta in modo pignolo, le regole di comportamento che si è tenuti a rispettare durante la prova.

Durante lo svolgimento della prova ciascun candidato:

- non può tenere con sé borse o zaini, libri o appunti, carta, telefoni cellulari ed altri strumenti elettronici
- non può scambiare informazioni o avere contatti verbali con altri candidati.

Peccato che accanto all'orario di inizio della prova, burocraticamente indicato, non sia indicato nemmeno il tempo concesso per rispondere ai 90 quesiti.

La mancanza dell'indicazione del tempo necessario per lo svolgimento della prova non esclude naturalmente che in sede di svolgimento della prova tale indicazione venga fornita, ma fa riflettere sulla sostanziale sottovalutazione di questo aspetto da parte di chi gestisce la prova: si tratta di un'altra componente che può influire sul risultato e incidere sulla validità di contenuto della prova stessa, in quanto non sarà possibile, in sede di analisi dei risultati, capire quali sono i rispondenti le cui prestazioni sono state condizionate dalle limitazioni di tempo previste per la prova (Lucisano 1989).

Negli ultimi anni comunque i bandi tendono a essere più precisi, almeno per gli elementi chiave della prova.

**Università degli Studi del Piemonte Orientale "Amedeo Avogadro" – Corso di laurea in Scienze della comunicazione (2005-06)**

**ART. 9**

L'ammissione al primo anno del Corso di laurea in Scienze della Comunicazione avviene previo superamento di prove che consisteranno nella soluzione di n° 140 quesiti a risposta multipla, con una sola risposta esatta tra le 4 indicate. La

prova è suddivisa in tre test a risposta multipla (4 risposte possibili, 1 sola esatta) sui seguenti argomenti:

- abilità logico-linguistiche (100 domande);
- conoscenze informatiche di livello elementare (20 domande);
- conoscenza della lingua inglese a livello elementare (20 domande).

...

Per lo svolgimento della prova è assegnato un tempo di 3 ore.

In questo caso l'articolazione dei quesiti che compongono "le" prove è chiara ed è definito il tempo complessivo concesso per la compilazione, inoltre l'alto numero dei quesiti suggerisce, ma non esplicitamente, l'idea che si tratti di batterie di domande secche, senza riferimenti a letture specifiche.

La piena libertà di organizzare le prove comporta anche scelte originali, che si distaccano del tutto dal modello suggerito dal ministero.

#### **Università di Roma "Tor Vergata" – Corso di laurea in Lettere (2005-06)**

La durata delle prove è fissata indistintamente, per tutti i CCL, in sessanta minuti.

La tipologia delle prove, stabilita dai singoli CCL, è descritta più oltre nelle Norme particolari.

...

#### **CCL in FILOSOFIA**

##### **Art. 2 – Calendario, tipologia delle prove e criteri di valutazione**

La prova consiste nella risposta aperta ad 1 quesito di commento di un testo filosofico (valutabile fino ad un massimo di 10 punti) e nella soluzione di 40 Quesiti a risposta Vero/Falso (valutabili fino ad un massimo di 40 punti), così suddivisi: 10 Quesiti di comprensione della lingua italiana, 20 Quesiti di comprensione di un semplice brano filosofico, 10 Quesiti di conoscenza della storia della filosofia.

In questo caso, accanto a una norma quadro sulla durata della prova, valida per l'accesso a tutti i corsi di laurea, si delega a ogni corso la responsabilità di decidere l'articolazione delle prove, e come vediamo nell'esempio del corso di Filosofia è prevista una puntuale distinzione di prove e punteggi relativi, con l'introduzione di una prova aperta (commento di quesito filosofico, per il quale però non sono fornite indicazioni circa le modalità di valutazione) e la presenza di 40 quesiti vero/falso, per i quali sarà interessante tornare quando affronteremo il problema dei punteggi attribuiti alle risposte.

In altri casi si cerca di venire maggiormente incontro alle esigenze di trasparenza attese dagli studenti, anche se in modo talvolta confuso.

### **Università di Urbino – Corso di laurea in Scienze psicologiche (2005-06)**

#### Argomenti della prova

La prova è composta da 80 domande di cultura generale (sulla base dei programmi della scuola secondaria superiore) e di accertamento della predisposizione per la disciplina dei corsi medesimi, come previsto dall'art. 4 della Legge 264/99.

Sembrano informazioni sostanzialmente generiche, ma leggendo con attenzione il bando si scopre una lunga nota con alcune precisazioni, tra le quali spiccano alcune notizie utili per chi dovrà affrontare la prova, oltre naturalmente le ferree regole di comportamento corretto, pena l'esclusione. Gli errori ortografici presenti sul testo del bando non sono stati corretti.

- la prova consisterà nella risoluzione di un questionario composto da 80 domande a risposta multipla; tutte le istruzioni per la corretta compilazione del foglio risposta saranno stampate nelle prime pagine del questionario insieme agli esempi che aiuteranno i candidati a capire le modalità di risoluzione dei quesiti.

- il tempo che sarà concesso sarà di 70 minuti dal momento in cui sarà dato il "VIA" comprendente anche la lettura delle istruzioni di cui al punto precedente.

- per la particolare struttura del foglio risposta NON sarà consentito correggere la risposta data né tantomeno sostituire il foglio risposta stesso.

Risulta sicuramente contraddittorio, dal punto di vista di chi deve compilare la prova, l'inserimento del tempo necessario per la lettura delle istruzioni nel tempo concesso per la prova: conviene capire bene come si risponde o avere qualche secondo in più per rispondere? Decisamente inquietante risulta essere anche il divieto assoluto di ripensamento sulle risposte date, che avvalorata la prospettiva quiz della prova: è la prima risposta quella che conta.

Bisogna comunque aggiungere che in una successiva nota il bando offre un'altra possibilità di comprendere qualcosa in più sulla prova, quando dice che "Sul sito [www.uniurb.it/psicologia](http://www.uniurb.it/psicologia) sono presenti alcuni esempi di item previsti dalla prova."

Concludiamo questa panoramica su contenuti e modalità delle prove di accesso con un esempio originale, almeno per quanto riguarda le modalità di svolgimento.

**Università di Siena – Corso di laurea in Scienze della comunicazione (2005-06)**

1. ...Ciascun candidato, digitando sul terminale il proprio nome e cognome accederà alla pagina del test. Il test, da concludere nel tempo limite di 20 (venti) minuti, si compone di 30 domande (i set di 30 domande variano per ogni scaglione di candidati), con 4 opzioni di risposta di cui solo una esatta. Il candidato seleziona con il mouse la risposta che ritiene corretta e può rettificare la propria scelta selezionando una risposta diversa. Il candidato "consegna" il test selezionando il bottone di conclusione posto a fine schermata. Il server registra il tempo impiegato e mostra a schermo il numero di risposte esatte prodotte dal candidato.

...

4. La prova di ammissione consisterà in domande riguardanti le seguenti materie: cultura generale, linguistica, logica, storia contemporanea, istituzioni italiane ed europee, elementi di economia, scienze sociali e informatica.

Oltre al totale abbandono di carta e penna, spiccano le 8 materie concentrate in 30 domande e il fatto che le 30 domande possano essere diverse per ogni gruppo di candidati, e anche in questo caso sarà interessante verificare le indicazioni fornite nel bando per il trattamento dei punteggi.

Gli esempi forniti non intendono certo rappresentare un quadro esaustivo delle modalità e dei contenuti delle prove di accesso utilizzate nel sistema universitario italiano, ma permettono di considerare ancora attuali e valide le perplessità espresse da Bini e Borsese (1995) sulle prove di accesso, soprattutto sull'arbitrarietà dei contenuti scelti, sulla base dei quali gli studenti sono chiamati a dimostrare la propria predisposizione agli studi, con l'aggravante che nei primi anni Novanta questa pratica riguardava una minoranza di corsi, di tipo strettamente professionalizzante, mentre oggi, come abbiamo visto negli esempi, si estende a indirizzi di studio più generali e paradossalmente meglio connessi agli indirizzi di studio della scuola secondaria superiore, i cui risultati vengono quindi implicitamente snobbati.

#### **1.4. Un problema nel problema: la scala dei punteggi**

Se l'analisi delle informazioni fornite nei bandi e riguardanti contenuti e modalità di svolgimento della selezione permette di ricavare indicazioni sulla validità di contenuto della prova, il controllo un altro aspetto permette di capire in che modo sono costruite le scale di punteggio che conducono alla formazione delle graduatorie.

Anche in questo caso il punto di riferimento principale è costituito dalle indicazioni del ministero circa i corsi di laurea di sua competenza.

##### **Decreto MURST 20 maggio 2002 - Art. 6**

1. Nella valutazione delle prove di cui agli articoli 2, 3, 4 e 5 si tiene conto dei seguenti criteri:

a) 1 punto per ogni risposta esatta;

-0,2 punti per ogni risposta sbagliata;

0 punti per ogni risposta non data;

b) in caso di parità di voti, prevale il punteggio ottenuto dal candidato nella soluzione di quesiti relativi ai seguenti argomenti:

1) per i corsi di laurea in medicina e chirurgia, odontoiatria e protesi dentaria, medicina veterinaria e per i corsi di laurea delle professioni sanitarie prevale in

ordine decrescente il punteggio ottenuto dal candidato nella soluzione rispettivamente dei quesiti relativi agli argomenti di logica e cultura generale, biologia; chimica; fisica e matematica.

Il modello indicato dal ministero quindi prevede un'assegnazione standard di punteggio per ogni domanda, con penalizzazione delle risposte sbagliate e neutralità delle omissioni. Non è certo questo il luogo per entrare nel dibattito sull'opportunità di penalizzare le risposte sbagliate<sup>17</sup>, ma è evidente che lo scopo principale della penalizzazione è scoraggiare le risposte date a caso. Però questa indicazione esplicita, addirittura nel regolamento della prova, influenza fortemente il comportamento del rispondente durante la prova: da una parte può essere distratto da ragioni di convenienza legati a una valutazione rischi/benefici nel mettere o no una crocetta, dall'altro può essere condizionato da stati d'animo personali di insicurezza o autostima, in ogni caso tutti elementi di disturbo circa l'oggetto della valutazione. Inoltre, come dimostrato da Gattullo e Giovannini (1989) le strategie di risposta si snodano su un *continuum* che ha agli estremi chi è del tutto sicuro della risposta da indicare e all'opposto chi risponde completamente a caso. In mezzo ci può essere l'eliminazione certa di un'alternativa, il dubbio fra due sole alternative, o la quasi certezza a favore di una alternativa: quale sarà il comportamento effettivo del rispondente in queste situazioni intermedie? Come valuterà la convenienza a rispondere opposta al rischio di sbagliare? Sempre in Gattullo e Giovannini (1989) l'elemento casualità di risposta viene fortemente ridimensionato, in relazione al numero di quesiti che compongono la prova e al numero di alternative dei quesiti: giocare ai dadi con un test ben fatto non risulta molto conveniente.

Ma anche prescindendo da considerazioni di opportunità circa la penalizzazione degli errori, il modello ministeriale fa implicitamente riferimento a uno specifico, quanto bizzarro, tipo di quesito, cioè una scelta multipla a sei alternative di risposta<sup>18</sup>, per cui di fatto viene introdotto un nuovo elemento di disturbo per i rispondenti, perché chi ha qualche dimestichezza con i calcoli può ulteriormente affinare il rapporto rischi/benefici nel caso in cui la penalizzazione, fissa, si applichi a quesiti con meno alternative di risposta.

<sup>17</sup> Cfr. Gattullo (1968), Vertecchi (1984), Boncori (1993), che riprendono la formula suggerita da Guilford (1954) per la neutralizzazione delle risposte casuali.

<sup>18</sup> E quanto si ottiene applicando la formula  $P_i = E_i - S_i/(n-1)$ , dove  $n$  è il numero di alternative previste per la domanda e  $S_i$  il numero di risposte errate, per cui per una risposta errata la penalizzazione 0,2 (1/5) corrisponde a  $n=6$ .

Il ministero indica anche la strada per risolvere eventuali punteggi *ex aequo* nella graduatoria, attraverso un confronto gerarchico dei punteggi ottenuti nelle diverse aree di contenuto previste dalla prova. Questo aspetto introduce un ulteriore disturbo per i rispondenti, poiché un quesito di cultura generale vale 1 punto, quanto uno di chimica, ma potrebbe anche "contare" di più, per cui ... dove conviene impegnarsi di più?

Nei bandi che abbiamo già visto nel precedente paragrafo viene generalmente confermata l'impostazione ministeriale di penalizzazione degli errori, apparentemente con una penalizzazione fissa di 0,25 per ogni errore, ma con effetti diversi sui punteggi secondo il tipo di quesiti proposti.

**Università degli Studi del Piemonte Orientale "Amedeo Avogadro" – Corso di laurea in Scienze della comunicazione (2005-06)**

... La prova è suddivisa in tre test a risposta multipla (4 risposte possibili, 1 sola esatta) ...

Nella valutazione delle prove si tiene conto dei seguenti criteri:

- 1 punto per risposta esatta,
- -0,25 punti per risposta sbagliata,
- 0 punti per risposta non data;

...

In caso di parità di punteggio, prevale il candidato che abbia ottenuto il maggiore punteggio nei seguenti argomenti: competenze logico-linguistiche;

La penalizzazione prevista corregge le risposte casuali di scelte multiple a 5 alternative, ma i quesiti sono, esplicitamente, solo a 4 alternative, per cui in fin dei conti i candidati possono ragionevolmente pensare a una strategia di "risposta comunque". Anche in questo caso eventuali *ex aequo* vengono risolti privilegiando il punteggio ottenuto in una delle prove, per inciso quella con il maggior numero di domande (100 su 140), per cui può essere interpretata come un'indicazione esplicita di priorità nella preparazione alla prova.

**Università di Milano "Bicocca" – Corso di laurea in Scienze della Comunicazione (2002-03)**

... la prova di selezione consisterà nella soluzione di quesiti a risposta multipla  
...

... la graduatoria sarà compilata sulla base del punteggio ottenuto nella prova di selezione e del voto conseguito all'Esame di Stato (maturità). Alla prova di selezione verrà attribuito un peso del 75%, al voto dell'Esame di Stato verrà attribuito un peso del 25%. Per ogni risposta non data verranno attribuiti 0 punti. Verrà attribuito 1 punto per ogni risposta esatta e -0,25 punti per ogni risposta errata.

In caso di pari merito il candidato più giovane d'età precede in graduatoria (art. 3 comma 7 della Legge n. 127/1997)

Il riferimento ai quesiti a scelta multipla non prevede l'indicazione del numero di alternative, per cui i candidati dovranno considerare, al momento della prova, la convenienza del rischio. Per il calcolo della graduatoria di accesso al corso il punteggio della prova viene integrato dal voto dell'Esame di Stato, che rappresenta il secondo criterio più utilizzato dalle università nelle procedure di accesso<sup>19</sup>. Correttamente viene anche indicato il fattore di pesatura dei due punteggi, che risulta motivante a favore di un buon rendimento della prova. In questo caso eventuali problemi di parità vengono risolti da un fattore anagrafico, del tutto esterno alla prova.

Il bando dell'Università di Urbino per il Corso di laurea in Scienze psicologiche risulta uguale al precedente per l'attribuzione dei punteggi e l'indeterminatezza del numero di alternative previste per i quesiti, ma risolve in modo diverso il problema degli *ex aequo*.

**Università di Urbino – Corso di laurea in Scienze psicologiche (2005-06)**

In caso di parità di punteggio verrà data la precedenza al candidato che riporta il punteggio più elevato nel diploma di scuola media superiore. A parità di votazione dello stesso vengono inseriti tutti i nominativi.

<sup>19</sup> Pur considerato un efficace predittore di successo accademico, il voto dell'Esame di Stato ha visto ridurre il suo peso negli ultimi anni a favore delle prove di accesso (Guicciardi e Lostia 1998, Andreani Dentici e Amoretti 2000).



Questa volta il punteggio dell'Esame di Stato non entra direttamente nella formazione della graduatoria, ma viene utilizzato come fattore determinante dell'eventuale parità di punteggio. Colpisce la notazione finale che di fatto può modificare il numero totale di posti complessivi messi a concorso, ma rappresenta soprattutto un'esplicita rinuncia a selezionare utilizzando criteri (età, data di presentazione della domanda ...) non immediatamente collegabili alla predisposizione per il corso di studi.

Veniamo a un altro caso che lascia molto perplessi sull'attribuzione dei punteggi.

#### **Università di Roma "Tor Vergata" – Corso di laurea in Lettere (2005-06)**

Il punteggio conseguito nelle prove sarà ponderato con il voto ottenuto nell'esame di diploma di scuola superiore nella proporzione di 90 (per le prove) a 10 (per il voto di diploma); la graduatoria risulterà ponderata tra questi due valori; in caso di parità sarà data preferenza al candidato con il punteggio più alto nelle prove e, in caso di ulteriore parità, a quello con il voto di diploma più alto.

...

*CCL in FILOSOFIA*

...

La prova consiste nella risposta aperta ad 1 quesito di commento di un testo filosofico (valutabile fino ad un massimo di 10 punti) e nella soluzione di 40 Quesiti a risposta Vero/Falso ...

La valutazione del questionario prevede l'assegnazione dei seguenti punteggi:

1 punto per ogni risposta corretta

0 per ogni risposta non data

- 0,25 per ogni risposta errata

Avevamo già notato nel paragrafo precedente la genericità dei criteri di valutazione del quesito di commento (valutabile fino a 10 punti), ma in questo caso osserviamo anche che la penalizzazione standard di 0,25 si applica a quesiti vero/falso, quindi a due alternative di risposta, per

cui l'effetto di scoraggiamento di risposte date a caso scompare: conviene comunque provare a rispondere. Il risultato è che il valore della prova aperta risulta alquanto ridimensionato, viste le proporzioni sul punteggio complessivo. Anche in questo caso sulla graduatoria complessiva incide, ma in misura ridotta, il punteggio dell'Esame di Stato, e per la risoluzione degli *ex aequo* si ricorre esclusivamente a criteri interni alla graduatoria.

Esistono comunque anche dei casi in cui la penalizzazione dell'errore risulta coerente con la formula canonica.

#### **Università di Bologna - Corso di laurea triennale in Attuazione e gestione del progetto in Architettura (2003-04)**

La prova di ammissione consiste nella soluzione di ottanta quesiti a risposta multipla, di cui una sola risposta esatta tra le cinque indicate ...

...

Verrà compilata una graduatoria generale di merito per i candidati comunitari ed assimilati ed una per i candidati non comunitari residenti all'estero, basate sul punteggio conseguito nella prova.

Nella valutazione della prova la Commissione giudicatrice si atterrà ai seguenti criteri: 1 punto per ogni risposta esatta; meno 0,25 punti per ogni risposta sbagliata; 0 punti per ogni risposta non data;

In caso di parità di voti, la Commissione giudicatrice si atterrà al seguente criterio: prevalenza in ordine decrescente al punteggio ottenuto dal candidato nella soluzione rispettivamente dei quesiti relativi agli argomenti di logica e cultura generale; storia; disegno e rappresentazione; matematica e fisica.

In questo caso il trattamento di base dei punteggi risulta corretto anche rispetto all'originale dettato ministeriale (ricordando che pur nell'autonomia degli atenei i corsi di Architettura restano programmati a livello nazionale), prevedendo la consueta gerarchia di importanza fra le diverse aree di contenuto in caso di parità di punteggio.

Se la linea di condotta suggerita dal ministero è stata sostanzialmente accolta da molti atenei, non mancano interpretazioni particolari sull'attribuzione dei punteggi.

**Università di Roma "La Sapienza" – Corsi di laurea in Scienze e tecniche psicologiche (2005-06)**

... per ciascuna risposta esatta sarà attribuito un punteggio di 1,1, per ciascuna risposta errata sarà sottratto il punteggio di 0,1. Per ciascuna risposta lasciata in bianco non è prevista né attribuzione di punteggio né penalizzazione.

Il candidato sarà valutato separatamente per ogni area. La somma del punteggio conseguito in ciascuna delle tre aree determinerà il punteggio totale. La graduatoria terrà conto del punteggio totale e, a parità di punteggio totale (tra più candidati), dei punteggi parziali ottenuti nella 2° area, nella 3° area e nella 1° area.

Non essendo specificato il tipo di quesiti che compongono la prova (sappiamo solo che sono 90, distribuiti in modo indefinito su tre aree) non è possibile capire l'effetto scoraggiante della penalizzazione prevista per le risposte casuali, ma si può ragionevolmente pensare che sia conveniente comunque indicare una risposta anche in caso di forti dubbi. Rimane l'impatto "psicologico" sul rispondente ("Perché ogni risposta giusta vale 1,1?") che potrebbe risultare di disturbo durante la preparazione e lo svolgimento della prova. Anche in questo caso i problemi di pari merito sono risolti con la gerarchia delle diverse aree di contenuto della prova.

Concludiamo con un'altra interpretazione originale sull'attribuzione dei punteggi in una prova di accesso.

**Università di Siena – Corso di laurea in Scienze della comunicazione (2005-06)**

... la Commissione formerà una graduatoria di merito dei candidati secondo un punteggio decrescente risultante dall'esito del questionario di cui sopra; saranno considerati idonei i candidati che avranno risposto in modo esatto ad almeno 10 domande. Nella valutazione della prova si terrà conto dei seguenti criteri:

- a) 1 punto per ogni risposta esatta
- 0 punti per ogni risposta sbagliata
- 0 punti per ogni risposta non data

b) in caso di parità di punteggio prevale la votazione conseguita nell'esame di maturità, persistendo ancora la parità, il tempo di conclusione del test ed eventualmente il più giovane di età.

Ci sono alcune differenze rispetto ai bandi visti in precedenza, e d'altra parte l'originalità di questa prova era evidente fin dalle modalità di compilazione, esclusivamente informatiche. In primo luogo è stabilita una soglia minima di idoneità, cioè un criterio di sbarramento per essere ammessi in graduatoria. Si tratta di un criterio di qualità, che però deve essere supportato, per essere effettivamente tale, da quesiti effettivamente capaci di distinguere l'esistenza di requisiti di base che possano funzionare da sbarramento. Per questo sarebbe stata preferibile una articolazione della prova in due parti, una più funzionale allo sbarramento, l'altra di affinamento del punteggio raggiunto.

Da sottolineare poi l'assoluta non penalizzazione dell'errore, che funziona da incentivo nel cercare di rispondere comunque, visto anche il tempo previsto per la prova (20 minuti per 30 domande) che esalta sicuramente l'elemento velocità, non a caso utilizzato anche come fattore dirimente in caso di parità subito dopo il punteggio dell'Esame di Stato. Quindi meglio attrezzarsi per rispondere sempre e il più rapidamente possibile: si tratta di un elemento di "predisposizione" per gli studi legati alla comunicazione?

### **1.5. Conclusioni per avviare il lavoro di indagine**

Dal complesso della panoramica fornita sulle prove di accesso attualmente svolte nelle università italiane emerge un elemento di criticità che si aggiunge alle perplessità sull'uso di questo strumento, a cominciare dal diverso grado di familiarità che i candidati possono avere nei confronti di prove strutturate (Bini e Borsese 1995). Bisogna però considerare che nell'ultimo decennio le prove strutturate hanno trovato larga e crescente utilizzazione nella scuola secondaria, anche in situazioni istituzionali quali l'Esame di Stato<sup>20</sup>, per cui la particolare forma dello strumento di rilevazione non risulta così discriminante.

Molto forti però restano i dubbi su quella che dovrebbe essere la vera funzione delle prove di accesso, cioè la capacità di distinguere la predisposizione agli studi, ovvero le possibilità di frequentare con successo il

<sup>20</sup> La terza prova, che può essere basata su prove strutturate, è stata introdotta nella sessione di esami 1999-2000 (cfr. ONES 2000)

corso universitario. Si tratta quindi di una funzione predittiva finalizzata anche a rendere razionale ed economica l'offerta delle università. Ma prima Guicciardi e Lostia (1998) e poi Andreani Dentici e Amoretti (2000) hanno rilevato, con studi sulle carriere universitarie successive alle prove di accesso di diverse leve di studenti, la scarsa relazione esistente fra i risultati di queste ultime e gli esiti accademici successivi<sup>21</sup>. Questi studi però non sono entrati nel dettaglio specifico delle scale di misura utilizzate per la selezione, ovvero hanno dato per scontate le modalità di attribuzione secondo la logica una risposta, un punto. I problemi legati allo strumento di selezione sono stati segnalati da Santambrogio (1997), che pur partendo da un giudizio positivo sulla limitazione degli accessi all'università, ha manifestato molte perplessità sulle prove utilizzate per la selezione, in particolare per i contenuti di riferimento della prova. Gli esempi tratti dai bandi delle prove di accesso, più vicine a noi nel tempo, ravvivano quei dubbi e li estendono anche al modo in cui le risposte dei candidati si trasformano in graduatorie, ponendo un interrogativo: quanta parte della scarsa efficacia predittiva delle prove di accesso è dovuta a difetti di costruzione delle prove dal punto di vista psicometrico?

Inoltre le analisi svolte finora hanno approfondito i risultati di prove di accesso che si sono effettivamente svolte, controllando gli esiti accademici di coloro che poi erano stati effettivamente ammessi, sulla base dei risultati delle prove, a sostenere un determinato corso di studi. Ma quali sarebbero stati gli esiti dei "non ammessi" secondo la prova di accesso?

Il problema della predittività si incrocia quindi con la verifica della funzionalità dello strumento di selezione utilizzato, rispetto agli scopi per cui è viene utilizzato, la creazione di una scala di predisposizione agli studi.

Il presente lavoro di ricerca si propone di trovare elementi utili per la risoluzione di questi problemi.

<sup>21</sup> Andreani Dentici e Amoretti (2000) osservano però che i risultati della prova di accesso acquistano peso nel prosieguo degli studi, avvicinandosi alla laurea.



## Capitolo secondo

### Ipotesi di ricerca

#### 2.1. Dalla situazione problematica all'ipotesi di ricerca

Abbiamo visto nel primo capitolo che l'attribuzione dei punteggi nelle prove di accesso svolte nelle università italiane segue criteri diversi, ma sostanzialmente allineati alle recenti indicazioni ministeriali, nonostante la relativa libertà di azione concessa agli atenei nella gestione delle prove. Il punteggio di ogni quesito è predeterminato e risulta omogeneo, a prescindere dalle differenze di contenuto fra quesiti. Il diverso peso che dovrebbero assumere determinati argomenti o contenuti all'interno della prova dipende quindi esclusivamente dal numero di domande presenti nella prova rispetto al numero totale<sup>22</sup>.

Abbiamo anche constatato la relativa diffusione del criterio di penalizzazione delle risposte errate, spesso realizzato in modo incoerente con le indicazioni della ricerca docimologica. L'insieme di queste osservazioni hanno fatto sorgere il dubbio che la scadente predittività delle prove di accesso circa il successo accademico, rilevata da diverse indagini, fosse in parte attribuibile a difetti nella costruzione dello strumento circa l'attribuzione dei punteggi.

D'altra parte il modello di riferimento fornito dalle indagini internazionali indica una modalità alternativa di costruzione delle scale di misura, basate sull'analisi degli item che compongono la prova secondo l'*Item Response Theory* (IRT). Il vantaggio più immediato fornito da questa modalità di analisi è la considerazione, su una sola scala di misura, della difficoltà della prova e dell'abilità dello studente, ottenendo quindi punteggi che dovrebbero risultare più affidabili e meno dipendenti dalla prova stessa. All'approfondimento dell'IRT e al suo rapporto con l'item analisi classica è dedicato il terzo capitolo del presente lavoro di ricerca,

<sup>22</sup> Sull'argomento, in relazione alla terza prova dell'Esame di Stato, si può vedere Bolletta (2000), che affronta il problema dei punteggi in relazione ai gruppi di domande che compongono una prova.

ma fin d'ora possiamo dire che secondo il modello IRT i punteggi dei singoli quesiti sono automaticamente pesati e non richiedono quindi un'esplicita penalizzazione delle risposte errate<sup>23</sup>.

Alla luce di queste considerazioni è possibile definire la seguente ipotesi di ricerca.

L'attribuzione dei punteggi di una prova di ingresso svolta secondo il modello di item analisi IRT permette di ottenere una graduatoria degli ammessi a un corso di laurea più precisa nella scala dei punteggi e maggiormente predittiva del successo accademico rispetto a una graduatoria ottenuta con il modello di item analisi classica.

## 2.2. Disegno di ricerca

Dall'analisi della situazione problematica esposta nel primo capitolo è emersa la constatazione che la verifica della predittività della prova di accesso si è limitata finora al controllo degli esiti del gruppo che la prova stessa aveva selezionato, evitando qualsiasi confronto con gli esiti accademici dei candidati esclusi secondo i risultati della prova. Quindi per poter verificare l'ipotesi si renderebbe necessario disporre anche dei dati relativi al percorso accademico dei candidati che risulterebbero esclusi dai risultati della prova.

Si tratta naturalmente di un'eventualità impossibile per una prova di accesso reale, per cui si rende necessario il ricorso a un contesto simulato, considerando i risultati di una prova costruita con finalità di selezione per l'accesso, ma utilizzata effettivamente solo per l'orientamento. Questa necessità ha comportato la rinuncia immediata a raccogliere risultati di prove di accesso reali svolte nei corsi di laurea delle università italiane, considerando anche la comprensibile ritrosia degli stessi corsi a fornire materiali così delicati.

Il contesto simulato nel quale svolgere la sperimentazione è costituito dai nuovi iscritti per l'anno 2003-04 al Corso di laurea in Scienze dell'Educazione e della Formazione dell'Università di Roma "La Sapienza", per il quale non è prevista prova di accesso, ma per il quale è prevedibile l'iscrizione di diverse centinaia di studenti (cfr. Premessa).

<sup>23</sup> Nelle indagini internazionali le istruzioni di compilazione delle prove incoraggiano piuttosto gli studenti a rispondere il più possibile, evitando le omissioni (OECD 2002).



La leva di studenti così identificata verrà coinvolta in una prova di ingresso appositamente predisposta, di cui verranno analizzati i risultati secondo diversi modelli, con la simulazione di graduatorie di accesso. Successivamente verranno rilevati degli indicatori di successo accademico che permetteranno di controllare le relazioni esistenti fra risultati della prova ed esiti accademici, confrontando le diverse modalità di analisi sia in relazione alla funzione selettiva, sia per gli aspetti riguardanti la predittività.

Considerando il tempo di svolgimento del dottorato di ricerca verranno considerati gli esiti relativi al solo primo anno di corso. La durata del lavoro di ricerca comporta necessariamente (Lucisano e Salerno 2002) il disturbo di alcuni fattori esterni, in particolare le esperienze esterne di vita degli studenti, che possono incidere pesantemente sugli esiti accademici fino alla riduzione del gruppo di riferimento per abbandoni o trasferimenti. Si tratta di un rischio da considerare e monitorare nelle diverse fasi di lavoro, puntando a raccogliere comunque tutti i dati realisticamente disponibili.

Per la realizzazione del lavoro sperimentale non è previsto alcun intervento sulle variabili oggetto di osservazione, per cui lo studio si inquadra nel filone delle "ricerche basate sulle correlazioni" (Lucisano e Salerno 2002), essenzialmente di tipo "relazionale", indirizzata cioè a cogliere relazioni significative fra variabili, ma con aperture prospettiche di tipo "predittivo", nel momento in cui si considereranno i rapporti fra risultati della prova di accesso e il percorso accademico.

La sperimentazione si svolge parallelamente nel Corso di Laurea di Scienze della Formazione dell'Università di Bologna, a cura della collega Truffelli, al cui rapporto si rimanda per tutte le informazioni circa il contesto bolognese.

### **2.3. Fasi della ricerca**

La struttura di svolgimento dell'indagine deve adattarsi da una parte ai tempi di svolgimento del dottorato di ricerca, dall'altra ai tempi effettivi della didattica universitaria. Questo comporta una serie di limitazioni, che però risultano ineliminabili in qualsiasi contesto reale di ricerca (Lucisano e Salerno 2002).

In particolare i vincoli più forti riguardano lo svolgimento della prova di ingresso e la rilevazione dei risultati del primo anno accademico, con differenze sostanziali fra i due contesti sperimentali di Roma e Bolo-

gna. Infatti l'inizio dell'A.A. 2003-04 di Roma è fissato al 1 novembre 2003 per Roma, ma le iscrizioni sono possibili fino al 5 novembre e sono state ulteriormente prorogate fino al 20 dicembre 2003. Questa sfasatura di date (peraltro modificata in corso d'opera) ha reso complesso raggiungere tutti i nuovi iscritti, pur predisponendo lo svolgimento di una serie di sessioni per la prova di ingresso (fra il 29 settembre e il 15 novembre 2003). A Bologna il particolare calendario previsto per l'attività didattica ha comportato per la collega Truffelli la possibilità di svolgere una sola sessione il 24 settembre 2003.

Un altro vincolo forte per l'organizzazione dell'indagine è rappresentato dalla conclusione del primo anno accademico. A Roma per i nuovi iscritti dell'A.A. 2003-04 è possibile svolgere esami del primo anno fino al mese di dicembre 2004, mentre a Bologna gli esami del primo anno di corso corrispondono ai crediti acquisiti entro il mese di marzo 2005.

Queste differenze spiegano perché non è possibile un confronto puntuale fra le due popolazioni, per cui i due lavori di ricerca hanno proceduto su strade parallele, condividendo però i dati più importanti per la validazione degli strumenti<sup>24</sup>.

Considerando i due momenti vincolanti del periodo triennale di svolgimento del dottorato è stato possibile definire lo schema di svolgimento che ha compreso le altre fasi dell'indagine e in particolare:

- una fase esplorativa di impostazione dell'indagine;
- la costruzione degli strumenti sperimentali da utilizzare nel corso dell'indagine, con collaudo pilota di alcuni materiali;
- l'analisi dei dati, sia per quanto riguarda la prova di ingresso, sia per l'incrocio con gli esiti accademici.

Inoltre nel corso del triennio si sono resi necessari diversi momenti di approfondimento teorico, in relazione ai temi incontrati durante l'indagine.

Nel presente schema sono riepilogate e descritte le fasi di realizzazione dell'indagine.

<sup>24</sup> Ricordiamo che il lavoro di ricerca di Truffelli è incentrato sulla sperimentazione di uno strumento innovativo basato su domande aperte, i cui risultati sono stati analizzati anche in relazione alle prove strutturate utilizzate in comune dai due lavori di ricerca.



## 2.4. Campione di ricerca

La popolazione di riferimento è rappresentata dai nuovi iscritti al Corso di laurea in Scienze dell'educazione e della Formazione dell'Università di Roma "La Sapienza" dell'Anno Accademico 2003-04. Poiché la prova di ingresso non risulta formalmente selettiva, tutti i nuovi iscritti sono chiamati alla partecipazione volontaria. Si può ipotizzare la perdita di un certo numero di potenziali aspiranti, senza però predeterminarne la numerosità, per cui si rende necessaria una verifica a posteriori della partecipazione e la definizione effettiva del campione su cui si è svolta la somministrazione in ingresso<sup>25</sup>.

Al termine del primo anno accademico tutti i partecipanti alla prova di ingresso verranno interpellati per verificare una serie di informazioni utili per ricostruire l'esito accademico, e anche in questo caso è possibile ipotizzare la perdita di un certo numero di iscritti, per cui si rende necessaria un'ulteriore verifica sul campione definitivo di studenti che ha partecipato all'intera sperimentazione.

Un elemento caratterizzante dell'analisi del campione effettivo in uscita dalla sperimentazione sarà comunque la sua articolazione in due sottogruppi, costituiti dai potenziali ammessi e non ammessi al corso di laurea secondo i dati della prova di ingresso.

## 2.5. Strumenti e metodologie di ricerca

Naturalmente per poter avviare il lavoro di verifica dell'ipotesi bisogna chiarire in partenza alcuni elementi, quali la validità della prova di ingresso, la scelta del modello di analisi IRT da utilizzare per il confronto e la definizione di successo accademico.

Per quanto riguarda la prova di ingresso si è deciso, in pieno accordo con la collega Truffelli, di allestire una prova strutturata originale da somministrare ai nuovi iscritti dell'A.A. 2003-04 dei corsi di laurea di

<sup>25</sup> Si può però anticipare che alla data di scadenza prevista per le iscrizioni dell'A.A. 2003-04 risultavano iscritti, secondo i dati forniti dalla segreteria studenti, 448 studenti, ed i partecipanti alla prova di ingresso sono stati 419, quindi è stato raggiunto circa il 94% della popolazione di riferimento. Le iscrizioni però sono state successivamente prorogate al mese di dicembre ed il numero complessivo di nuovi iscritti per l'A.A. 2003-04 è stato di 516 studenti.

Scienze dell'educazione e della Formazione di Roma "La Sapienza" e di Scienze della Formazione di Bologna, corsi per i quali non era previsto il numero programmato. La prova, articolata in diversi strumenti di rilevazione, è accompagnata da un questionario per la rilevazione di una serie di variabili di sfondo utili per l'interpretazione dei risultati. Gli strumenti costruiti sono illustrati nel quarto capitolo del presente rapporto.

La scelta del modello di analisi IRT presuppone un approfondimento teorico e storico di questa modalità di analisi dei dati di una prova, e il conseguente confronto con la modalità di analisi classica dei test. Da una parte vanno considerati gli sviluppi teorici intervenuti negli ultimi 50 anni di questo filone di ricerca docimologica, proiettati sull'intero secolo scorso che ha visto la crescita e lo sviluppo degli studi psicometrici, dall'altra non bisogna trascurare gli spunti operativi offerti dal trattamento dei dati nelle indagini internazionali e nei software applicativi per l'analisi degli item. L'approfondimento teorico-operativo sull'analisi IRT è contenuto nel terzo capitolo del presente rapporto.

Per quanto riguarda la definizione del successo accademico, l'approfondimento risulta necessario e funzionale per la costruzione di uno strumento specifico per la raccolta di informazioni sui risultati del primo anno di corso, considerato che la particolare situazione delle segreterie dei corsi di laurea dell'Università di Roma "La Sapienza", a differenza dell'Università di Bologna, non permette di ottenere dati affidabili sul percorso di studio degli studenti nei tempi previsti per l'indagine. Le considerazioni circa il successo accademico sono contenute nell'ultimo paragrafo del quarto capitolo, per quanto attiene ai principi ispiratori dello strumento di rilevazione degli esiti, e nel sesto capitolo, in sede di analisi dei dati.



## Capitolo terzo

### Le diverse modalità di item analisi: confronto fra il modello classico e il modello IRT

#### 3.1. Valutare la qualità dei quesiti di una prova

La definizione delle procedure di item analisi può essere fatta risalire al lavoro svolto da Spearman nei primi anni del Novecento, arricchito e sistematizzato fino agli anni Sessanta con il contributo di Novick.<sup>26</sup>

Spearman (1904a, 1904b) ha avuto il merito di creare la prima teoria psicometria dell'intelligenza, riconoscendo il collegamento esistente fra due fattori: quello specifico misurato da una prova e quello generale rappresentato dall'intelligenza. Partendo da questo assunto Spearman ha utilizzato il calcolo matematico proprio per affinare la qualità delle misure rilevate nei test: la semplice formula

$$X = T + E$$

dove X è il punteggio osservato, T il punteggio vero e E l'errore di misura, è diventata la base per le applicazioni psicometriche elaborate nel corso del Novecento, e lo stesso Spearman ha indicato la riduzione dell'errore di misura come l'impegno più critico per il "misuratore" dell'intelligenza, introducendo di fatto il concetto di *reliability* nella considerazione critica di una prova (Bartholomew 1995).<sup>27</sup>

Nel corso del secolo la teoria classica dei test (CTT, *Classical Test Theory*) viene applicata e affinata nei particolari (Guilford 1936 – 1954, Guttman 1950) fino al saggio *The axioms and principal results of classical test*

<sup>26</sup> Una panoramica esauriente dell'evoluzione delle tecniche psicometriche nel corso del '900 si può trovare nel contributo di Crocker e Algina (1986).

<sup>27</sup> Da ricordare anche la definizione da parte di Spearman di un coefficiente di correlazione ( $\rho$ ) tra punteggi di rango, ancora largamente usato e la partecipazione alla definizione della misura di attendibilità Spearman-Brown per la lunghezza dei test.

*theory* di Novick (1966), anche se il contributo più importante alla sistematizzazione del trattamento statistico dei dati si ritrova nell'opera di Guilford (1950), *Fundamental Statistics in Psychology and Education*.

Nell'item analisi di tipo classico vengono considerate alcune caratteristiche dei quesiti (facilità, discriminatività, qualità dei distrattori) per poter giudicare l'efficienza dell'intera prova<sup>28</sup>. L'osservazione di queste caratteristiche permettono una descrizione del test (Calonghi 1978) che conduce a una serie di decisioni sulla qualità del test stesso e dei punteggi ad esso collegati, fornendo anche indicazioni esplicite per intervenire sui quesiti allo scopo di migliorare la qualità della prova.

L'indice di facilità (o di difficoltà) di ogni quesito viene definito calcolando la percentuale di risposte corrette date dai rispondenti per il quesito stesso, ovvero dividendo il numero di risposte corrette per il numero totale dei rispondenti. Si ottiene così un valore compreso fra 0 (nessuna risposta corretta) e 1 (tutte risposte corrette), considerando che i quesiti più efficienti si trovano nella fascia centrale, compresa fra 0,25 e 0,75 (Boncori 1993).

L'indice di discriminatività considera la capacità di ogni quesito di distinguere i rispondenti più bravi dai meno bravi, dove la bravura è definita dal punteggio complessivo ottenuto nella prova. Per fare questo bisogna considerare l'andamento delle risposte corrette del singolo quesito rispetto all'andamento della graduatoria complessiva, ed esistono due modalità alternative: è possibile rapportare il numero di risposte corrette del gruppo migliore di studenti con quello del gruppo peggiore, ipotizzando che il risultato debba essere positivo, cioè che gli studenti migliori rispondano meglio anche nel singolo item<sup>29</sup>; la modalità alternativa consiste nel calcolo del punto biseriale, che definisce la correlazione fra le due serie di punteggi (per il singolo quesito e per l'intera prova), anche in questo caso ipotizzando che l'andamento sia simile. Con entrambe le modalità di calcolo risultano critici i valori negativi, che atte-

<sup>28</sup> Il trattamento statistico dei dati di una prova permette di ricostruire decine di indici diversi (Boncori 1993), ma i tre aspetti indicati rappresentano la struttura base dell'item analisi classica.

<sup>29</sup> I due gruppi, detti anche estremi, devono essere numericamente uguali, ed essere costituiti da almeno il 25% dei rispondenti complessivi, ma considerando che tanto è più piccolo il gruppo tanto più è rischioso limitare la composizione del gruppo (Gattullo e Giovannini 1989), vista la maggiore incidenza del risultato del singolo soggetto. Per evitare questi rischi in genere gli estremi utilizzati per il confronto sono costituiti da circa il 33% dei soggetti complessivi.



stano un andamento anomalo dei punteggi del singolo item rispetto alla scala dei punteggi complessivi, ma anche i valori positivi inferiori a 0,20<sup>30</sup>, che testimoniano una insufficiente capacità del quesito di distinguere i bravi dai meno bravi.

Il controllo dei distrattori, cioè delle alternative non corrette di un quesito a scelta multipla, si realizza calcolando anche per loro gli indici di facilità e discriminatività. Naturalmente in questo caso i criteri di qualità si rovesciano rispetto a quelli considerati per l'alternativa corretta, per cui la percentuale di scelta di un distrattore non dovrebbe superare quella della risposta esatta, altrimenti testimonierebbe una eccessiva capacità distrattiva, ma anche non essere prossima allo 0, in tal caso il distrattore risulterebbe troppo poco plausibile; per la discriminatività il dato relativo al distrattore dovrebbe risultare negativo, cioè dovrebbe risultare più scelto fra gli studenti meno bravi.

Il controllo preventivo della qualità dei quesiti di una prova deve considerare l'insieme di questi elementi, ma il presupposto fondamentale dell'item analisi classica è l'attendibilità (*reliability*) della prova, cioè la sua capacità di misurare in modo preciso e costante, riducendo al minimo gli errori di misura.

Il concetto di attendibilità non deve pertanto essere limitato alla prova in sé, ma esteso a tutte le fasi e a tutti gli aspetti della sua utilizzazione: le condizioni di somministrazione, il comportamento dei rispondenti, i tempi assegnati, l'inserimento dei dati. In tutti questi momenti possono intervenire variabili diverse da quelle che la prova intende misurare (disturbi esterni, stanchezza, mancanza di tempo) e verificarsi veri e propri errori materiali (un inseritore di dati che digita una lettera per un'altra nella tabella di analisi) che aumentano l'errore (E) e riducono il valore cercato (T). La stima dell'attendibilità costituisce quindi un passaggio obbligato e preventivo dell'item analisi, e a questo scopo sono state elaborate diverse tecniche statistiche, da una parte basate sul concetto di duplicazione/ripetizione del test (forme parallele, re-test), dall'altra sul controllo della sua consistenza interna.<sup>31</sup>

<sup>30</sup> Se la discriminatività è calcolata con il metodo di confronto fra gli estremi è prudente considerare accettabili i valori superiori a 0,30 (Gattullo e Giovannini 1989), soprattutto se il numero complessivo dei rispondenti è limitato.

<sup>31</sup> La consistenza può essere verificata anche con il metodo split-half (divisione in due parti del test e relativo confronto), o attraverso il calcolo di coefficienti quali l'Alfa di Cronbach o il KR 20 di Kuder-Richardson (Andreani Dentici 1968, Boncori 1993)

Naturalmente una verifica negativa dell'attendibilità di una prova non permette immediatamente di capire quanta parte dell'errore sia dovuta alla prova stessa e quanta agli aspetti legati alla somministrazione e al trattamento dei dati. Proprio per questo la standardizzazione e il controllo delle procedure di utilizzazione di una prova risultano determinanti per limitare, o stimare ragionevolmente, il margine di errore che sarà rilevato dal controllo della attendibilità.

Viene considerata indispensabile per uno strumento di misura anche un'altra caratteristica: la validità. Il test, nel suo insieme, ma anche nelle singole domande, deve misurare esattamente quello che si vuole misurare. Solo in questo modo è possibile considerare il punteggio ottenuto da uno studente nel test una stima effettiva di un'altra cosa (il suo profitto, la sua preparazione in ingresso, la sua competenza.... in un determinato ambito).

Sempre sul versante della verifica preventiva della validità della prova, bisogna ricordare un ultimo presupposto per la realizzazione di un item analisi efficace, definito fin dalle prime intuizioni di Spearman: l'esistenza di un costrutto da analizzare di tipo unidimensionale, per cui i punteggi ricavati dalla prova possono far riferimento a una sola variabile indipendente. In situazioni in cui si può ipotizzare una multidimensionalità del costrutto, risulta opportuno svolgere un'analisi di tipo fattoriale, che distingua preventivamente le variabili secondo cui raggruppare i singoli item di un test<sup>32</sup>.

### 3.2. Dall'item analisi classica al modello IRT

Parallelamente alla calibrazione del modello classico di item analisi è cominciata però una riflessione sull'effettivo rapporto esistente fra il quesito e lo studente, cioè fra la facilità/difficoltà di risoluzione del primo e l'abilità del secondo, in definitiva il complesso rapporto esistente fra misura, strumento di misura e oggetto della misura.

I principali aspetti critici della teoria classica di analisi degli item riguardano in particolare: la forte sensibilità dei valori degli indici rispetto

<sup>32</sup> Il filone dell'analisi degli item collegata all'analisi fattoriale è stato esplorato nella definizione dell'*Item Factor Analysis* (IFA), per il quale si può ricordare il contributo di Christofferson (1975) e Bock R.D., Gibbons R., Muraki E. (1988). Bartholomew (1987) e Takane, De Leeuw (1987) hanno dimostrato che IFA e IRT risultano equivalenti laddove le funzioni di risposta si distribuiscono in modo normale.

al gruppo che svolge la prova; le difficoltà di comparazione fra soggetti sottoposti a prove diverse (nonostante le procedure di normalizzazione dei punteggi); un ripensamento sul concetto di attendibilità della prova (non facilmente verificabile se non, almeno inizialmente, con procedure complesse quali le forme parallele); la sfuggente definizione del rapporto esistente fra punteggi e caratteristiche degli item.

La riflessione, avviata negli anni Quaranta da Lawley (1943) e da Tucker (1946), ha proprio come primo obiettivo quello di approfondire il rapporto esistente fra le risposte a una prova e una variabile indipendente, quale può essere l'abilità o l'intelligenza. L'intuizione base consiste nella possibilità di ottenere misure dai singoli quesiti di ogni prova (indipendenza locale degli item), da cui il nome definito da Lord (1952) di *Item Response Theory* (IRT).

Una prima importante svolta si ha con l'opera del matematico danese Rasch, che formula un modello innovativo di item analisi. Rasch (1960) parte da un'assunzione: la probabilità che uno studente risponda correttamente a un quesito dipende da due fattori, la difficoltà dell'item e l'abilità dello studente, fattori che quindi risultano misurabili con una sola variabile su una sola scala.

In pratica mentre con l'item analisi classica le misure relative a un item non avevano niente a che fare con la scala dei punteggi degli studenti, con l'analisi IRT si utilizzano le stesse misure per stimare la qualità dell'item e la qualità dello studente, stimando per entrambi la posizione sulla scala dell'abilità misurata dalla prova, detta tratto latente (l'IRT è anche definita *Latent Trait Theory*). Il tratto latente ( $\theta$ ) è l'elemento che tiene uniti i quesiti di una prova (viene confermata la necessità di analizzare un costrutto di tipo unidimensionale, come per l'item analisi classica), in pratica quello che la prova sta misurando (la validità); ovviamente con l'analisi IRT è anche possibile verificare quali quesiti non sono coerenti con il tratto latente della prova, come pure quali studenti non sono coerenti, perché hanno risposto in modo anomalo alle domande della prova.

Il classico modello Rasch dicotomico (ogni item ha due possibili risposte 0/1) suppone che la probabilità che il soggetto  $n$  risponda correttamente sull'item  $i$  ( $P_{ni}$ ) dipenda dall'abilità del soggetto ( $\beta_n$ ) e dalla difficoltà dell'item ( $\delta_i$ ), ma poiché la differenza tra abilità e difficoltà è potenzialmente infinita a fronte di  $P_{ni}$  variabile tra 0 e 1, viene specificato un legame logistico. Ecco la formula di riferimento definita da Rasch.

$$(1) \quad \ln (P_{ni}/(1 - P_{ni})) = \beta_n - \delta_i$$

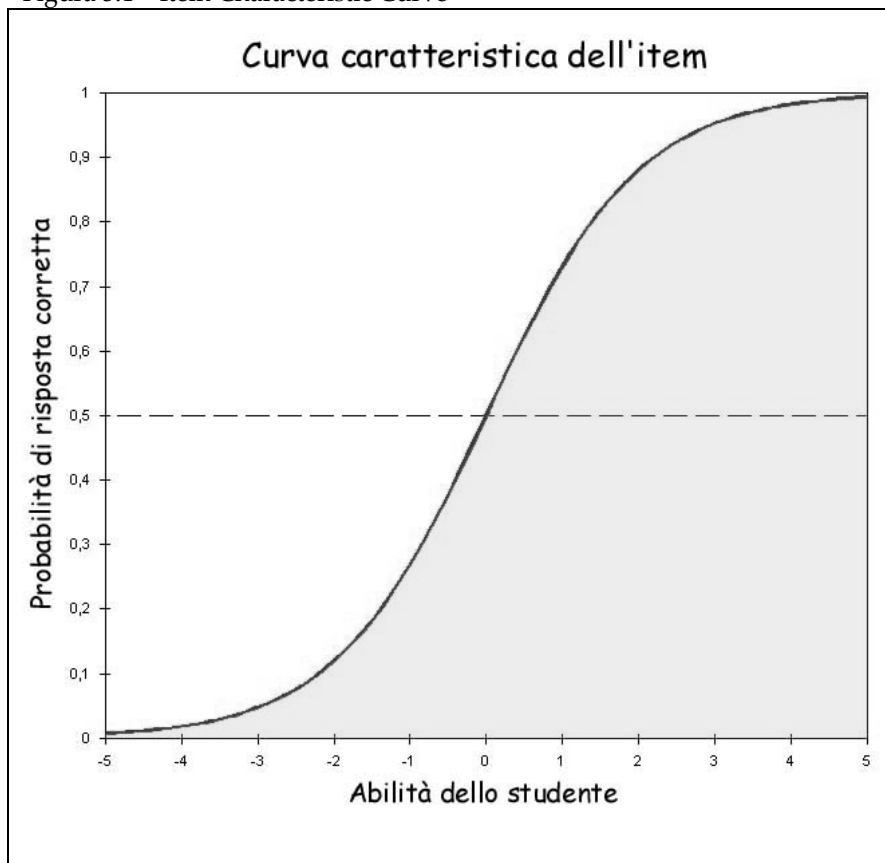
La misura ottenuta non risulta pertanto proporzionale alla probabilità di risposta, quanto alla trasformazione logit.<sup>33</sup> Nel modello di Rasch si nota che  $\beta_n$  dipende solo dall'abilità della persona  $n$  e che  $\delta_i$  dipende solo dalla difficoltà dell'item  $i$ . Si noti che  $P_{ni}$  aumenta all'aumentare di  $(\beta_n - \delta_i)$ , ma non in modo lineare, e che vale 0,5 se l'abilità del soggetto è uguale alla difficoltà dell'item.

Per comprendere la relazione esistente fra difficoltà dell'item e abilità dello studente si può osservare la curva caratteristica dell'item (ICC, *Item Characteristic Curve*).

Come si può notare con il crescere dell'abilità dello studente aumenta anche la probabilità di risposta corretta (area bianca del grafico). Se l'abilità è 0 (corrispondente all'abilità media del gruppo cui appartiene lo studente) la probabilità è del 50% (0,5), se l'abilità scende le probabilità diminuiscono, mentre per uno studente più abile le probabilità di risposta corretta crescono. La forma della curva è data dal particolare tipo di relazione fra le variabili, poiché come visto nella formula (1) l'abilità si misura in logit, cioè l'unità di misura dell'abilità dello studente è il logaritmo naturale del rapporto tra risposte corrette ed errate rispetto ad item che si trovano all'origine della scala, cioè che hanno valore 0.

<sup>33</sup> L'intuizione del modello logistico deve essere attribuita a Lord (1952), che come vedremo più avanti ha fornito la base per l'elaborazione di modelli di analisi IRT a più parametri (Birnbaum 1968). Tuttavia Rasch, pur partendo da un approccio legato alla teoria della probabilità, di fatto sviluppa un modello di tipo logistico (Baker 2001).

Figura 3.1 – Item Characteristic Curve



Allo stesso modo si può “leggere” il grafico come la curva caratteristica dello studente che risponde all’item, dove le probabilità di risposta sono legate alla difficoltà dell’item, esattamente con lo stesso andamento appena visto, ma riferito alla area grigia del grafico: difficoltà 0 con probabilità 50%, superiore a 0 (più difficile) probabilità crescenti e quindi raggiungibili dagli studenti più abili, inferiore a 0 probabilità discendenti, alla portata anche degli studenti meno abili. Anche dal punto di vista visivo il grafico dimostra come difficoltà e abilità rappresentino due facce della stessa medaglia: al crescere della difficoltà dell’item, deve crescere l’abilità necessaria per dare una risposta corretta.

La strada aperta da Rasch è stata naturalmente proseguita anche da altri ricercatori, con la messa a punto di altre metodologie che considerano diversi parametri di analisi. Se il modello originario di Rasch considera unicamente la facilità/difficoltà dell’item, dalla fine degli anni

Cinquanta Birnbaum approfondisce dal punto di vista matematico il modello logistico di Lord, con la messa a punto (Birnbaum 1968) di modelli di analisi che considerano anche la discriminatività (modello a due parametri) o anche le probabilità di risposta casuale all'item (modello a tre parametri).

Ecco l'equazione per il modello a due parametri, quello più largamente utilizzato nell'analisi IRT (Bove 2000).

$$(2) \quad P_{ni} = \frac{\exp \{a_i(\beta_n - \delta_i)\}}{1 + \exp \{a_i(\beta_n - \delta_i)\}}$$

la probabilità  $P$  di rispondere correttamente all'item  $i$  da parte del soggetto  $n$  è funzione della difficoltà dell'item  $i$ ,  $\delta_i$ , della capacità di discriminatività dell'item  $a_i$  e dalla abilità del soggetto  $n$ ,  $\beta_n$ .

Da notare che nel caso  $a_i=1$ , l'equazione coincide con il modello originale di Rasch. Rispetto alla curva caratteristica dell'item, la discriminatività si può osservare nel grado di pendenza della curva stessa, risultando più discriminativa dove la pendenza è maggiore. Il modello di analisi elaborato da Birnbaum risulta particolarmente efficace per i quesiti a scelta multipla.

Il cammino compiuto per la definizione di un modello alternativo di analisi vede due tappe fondamentali nella pubblicazione di *Statistical theories of mental test scores* di Lord e Novick (1968) che sistematizza la riflessione statistica sui diversi modelli di analisi (e dei 5 capitoli dedicati all'IRT, 4 sono scritti da Birnbaum) e di *Application of item response theory to practical testing problems*, in cui Lord (1980) chiude un trentennio di autorevole studio personale, e impegno operativo nell'ambito dell'*Educational Testing Service*, sull'analisi degli item.

Rispetto agli aspetti critici che avevano avviato la riflessione sulla teoria classica, l'analisi IRT ha definito un modello matematico che definisce la performance di un soggetto in funzione di una variabile indipendente (il tratto latente), riuscendo d'altra parte a ottenere misure immediatamente comparabili fra loro perché indipendenti dal gruppo che ha svolto la prova.

Il problema più delicato di questi modelli riguarda la stima dei parametri, da effettuare basandosi sui risultati effettivi della prova, stima tanto più precisa quanto più consistente è il numero degli item o dei rispondenti. Un quadro riepilogativo del cammino verso la definizione di

una stima efficace dei parametri è stato fornito da Baker (1987), ma forse il passo decisivo è stato realizzato a partire dal lavoro di Bock e Lieberman (1970) con la definizione della procedura MML (*Marginal Maximum Likelihood*), sviluppata da Bock e Aitkin (1981) con l'inserimento dell'algoritmo EM (*Expected Maximum*), che permette una stima più precisa dei parametri. Questo ha condotto a un netto miglioramento nella qualità dell'analisi anche per prove composte da un numero ridotto di item e svolte da pochi soggetti, nonché di rendere più affidabile l'analisi IRT a più parametri.<sup>34</sup>

Nella pratica i software costruiti per l'analisi IRT a più parametri lavorano su diverse fasi, o cicli (*loops*), che partono da una stima fondata sugli indici dell'item analisi classica, per poi reiterare la stima, secondo la procedura MML, ripulendola per ogni *loop* dai valori anomali (quindi migliorando la stima dei singoli item e dei singoli rispondenti). Per far questo si presuppone che la distribuzione del tratto latente ( $\theta$ ) sia tendenzialmente normale; in caso contrario la procedura di analisi può incepparsi e non riuscire a produrre stime accettabili anche dopo molti *loops*. L'affinamento delle procedure di calcolo per stimare i parametri permette di risolvere un altro limite tecnico dell'analisi classica, cioè la necessità di un collaudo preventivo della prova: l'analisi IRT permette di ridurre l'effetto degli item che risultano incoerenti, (non "fittano", dall'inglese *to fit*) con il tratto latente della prova (allo stesso modo di quelli non discriminativi rilevati dall'item analisi classica), modificando di conseguenza anche il punteggio raggiunto dagli studenti che rispondono a questi item critici.

Bisogna doverosamente aggiungere che anche le nuove procedure IRT non hanno risolto del tutto il problema della standardizzazione dei singoli item: il tratto latente della prova dipende ancora in una certa misura (dipendente dalla normalità della distribuzione) dalla combinazione degli item e dalla loro interazione con i rispondenti, per cui lo stesso item in un contesto diverso (altra prova o altro gruppo) può riportare valori diversi.

<sup>34</sup> Bisogna però ricordare che le indagini internazionali più importanti degli ultimi anni svolte dall'OCSE e dalla IEA utilizzano per l'analisi dei dati ancora il modello originale di Rasch a un solo parametro (OECD 2002 e 2005; Martin, Mullis, Kennedy 2003)

### 3.3. Confronto fra item analisi classica e IRT

Quali sono i vantaggi dell'analisi degli item con il metodo IRT? Abbiamo già detto che è prevista una sola scala di misura per item e studenti, e questo può risultare molto utile per chi deve al tempo stesso considerare la qualità della prova e i risultati degli studenti. Per comprendere meglio le differenze con il metodo classico ora applicheremo l'analisi IRT alla stessa prova utilizzata nelle unità precedenti, ma è necessario fare una premessa tecnica.

Il modello di analisi di Rasch e quello a due parametri di Birnbaum risultano di non semplice applicazione a livello matematico, raggiungendo livelli di complessità non implementabili in un comune foglio elettronico. Per cui mentre l'item analisi classica può essere svolta con un normale foglio elettronico, secondo le indicazioni sistematizzate da Guilford (1950 e 1954), per poter realizzare l'analisi IRT è necessario disporre di un software specifico.<sup>35</sup> Attualmente sono disponibili pacchetti per l'analisi IRT solo in lingua inglese, hanno costi significativi e essendo costruiti per addetti ai lavori non raggiungono quel livello di familiarità tipico dei programmi più utilizzati nei personal computer. Bisogna però anche aggiungere che non si tratta di programmi che richiedono enormi potenze di calcolo, anche perché si fondano su una grafica essenziale.

Vincenti e Calvani (1987) e Lucisano e Siniscalco (1992) hanno già proposto un confronto fra le due modalità di analisi, rilevandone vantaggi e limiti, ma anche il diffuso livello di sovrapposizione dei risultati, però il loro confronto si è svolto solo considerando il metodo di Rasch. Riteniamo utile, prima di passare ai capitoli sulla realizzazione operativa della ricerca, realizzare un confronto anche con il metodo a due parametri, considerato oggi adeguatamente affidabile e conveniente anche rispetto allo stesso metodo di Rasch<sup>36</sup>.

Per far questo utilizzeremo i risultati ipotetici di una prova già trattata con l'item analisi classica (Asquini e Piria 1998) nel contesto di un corso di formazione. Si tratta di una prova costruita a tavolino per evidenziare la maggior parte degli aspetti osservabili attraverso la procedura di

<sup>35</sup> Diversi manuali sull'IRT usciti negli ultimi anni fanno esplicito riferimento a pacchetti software, come Baker (2001), Thissen, D. (1991) e Hambleton, Swaminathan e Rogers (1991) che confrontano diversi applicativi per l'analisi IRT.

<sup>36</sup> Baker (2001) nel suo efficace manuale illustra il modello a due parametri prima di quello di Rasch, considerandolo più aderente alla struttura logistica dell'analisi IRT.



item analisi tradizionale, per cui risulta utile svolgere l'item analisi IRT (sia Rasch che a due parametri) per verificare le differenze, sia in termini di descrizione della prova, sia di distribuzione dei punteggi. La prova esempio è costituita da 10 item e risulta svolta da 19 rispondenti.

Il calcolo della discriminatività è stato effettuato attraverso il confronto dei risultati ottenuti, per ogni item, dagli studenti migliori e da quelli peggiori, così definiti secondo il punteggio complessivo riportato nella prova. Seguendo le indicazioni di Wood (1985) la ridotta ampiezza del gruppo considerato conduce all'ampliamento al 33% circa delle fasce estreme da confrontare (nel caso specifico 6 studenti su 19), laddove per gruppi molto più estesi la percentuale di rispondenti da considerare per il confronto può scendere<sup>37</sup> anche fino al 25%.

Parallelamente all'analisi delle percentuali di scelta per tutte le alternative di risposta, il calcolo della discriminatività è stato svolto anche per i distrattori, che naturalmente dovrebbero mostrare di norma valori negativi (Nunnally 1978). Ricordiamo che l'indice di discriminatività può essere ottenuto anche attraverso il calcolo della correlazione punto biseriale, in pratica la correlazione di Pearson tra il singolo item e il punteggio dell'intero test, o della sua versione probabilistica (correlazione biseriale); i risultati ottenuti con i due tipi di correlazione sono molto simili (Bove 2000) e gli indici ottenuti funzionano in modo simile a quello calcolato con il confronto degli estremi (Boncori 1993).

Per la realizzazione dell'item analisi fondata sulla teoria classica è stato utilizzato il foglio di calcolo Excel, secondo le indicazioni di Asquini (1997).

Nella Tabella 3.1 sono evidenziati i valori critici per quanto riguarda facilità (<0,25 e >0,75), discriminatività (<0,30) e i distrattori che risultano poco plausibili.

<sup>37</sup> Cfr. nota 29

Tabella 3.1– Item analisi prova esempio: modello classico

	item 1	item 2	item 3	item 4	item 5	item 6	item 7	item 8	item 9	item 10	Totale
stud.03	B	C	C	A	B	B	C	D	B	n	1
stud.01	D	C	B	A	B	D	D	B	C	D	3
stud.04	D	A	A	C	B	A	B	D	C	D	4
stud.15	C	B	C	C	B	A	D	B	B	n	4
stud.18	B	A	B	C	B	D	C	A	A	n	4
stud.02	B	A	B	C	B	B	D	B	B	D	5
stud.08	A	A	B	C	B	D	D	B	A	n	5
stud.10	C	A	C	D	B	A	D	B	C	D	5
stud.13	C	C	B	D	B	C	D	B	D	D	5
stud.14	A	A	B	A	B	C	B	A	A	D	5
stud.05	B	A	C	C	B	C	B	A	A	D	6
stud.07	C	A	D	A	B	B	D	A	C	D	6
stud.16	B	A	B	C	B	C	D	B	A	n	6
stud.09	C	A	B	B	B	C	D	B	A	D	7
stud.12	D	A	D	B	B	C	D	A	A	D	7
stud.06	C	A	D	B	B	C	D	A	A	D	8
stud.17	C	A	B	C	B	C	D	B	A	D	8
stud.19	C	A	D	C	B	C	D	B	A	D	9
stud.11	C	A	D	C	B	C	D	C	A	D	10
chiave	C	A	D	C	B	C	D	C	A	D	
risp. A	11	79	5	21	0	16	0	32	58	0	
risp. B	26	5	47	16	100	16	16	53	16	0	
risp. C	47	16	21	53	0	53	11	5	21	0	
risp. D	16	0	26	11	0	16	74	11	5	74	
omiss.	0	0	0	0	0	0	0	0	0	26	
errate	0	0	0	0	0	0	0	0	0	0	
facilità	0,47	0,79	0,26	0,53	1,00	0,53	0,74	0,05	0,58	0,74	
discrim.	0,67	0,50	0,67	-0,17	0,00	1,00	0,50	0,17	0,83	0,50	
discrim. distrattori											
A	0,00	0,50	-0,17	-0,33	0,00	-0,33	0,00	0,17	0,83	0,00	
B	-0,50	-0,17	-0,17	0,50	0,00	-0,33	-0,17	0,00	-0,50	0,00	
C	0,67	-0,33	-0,33	-0,17	0,00	1,00	-0,33	0,17	-0,33	0,00	
D	-0,17	0,00	0,67	0,00	0,00	-0,33	0,50	-0,33	0,00	0,50	

E' sufficiente un rapido commento dei dati relativi alla prova esempio per evidenziare la criticità degli item 4 (discriminatività negativa), 5 (eccessiva facilità), 8 (eccessiva difficoltà e bassa discriminatività) e 10 (eccessiva facilità, escludendo l'alto numero di omissioni). Naturalmente si tratta di una prova esempio, costruita proprio per evidenziare in poco spazio una grande varietà dei difetti nella costruzione degli item, per cui

a rigore la prova dovrebbe essere rigettata nel suo complesso, nonostante alcuni item (1, 6, 9) presentino valori più che accettabili e altri (2, 3, 7) quasi accettabili, con piccoli difetti relativi alla facilità o con alcuni distrattori che risultano poco efficaci. Nell'ultima colonna sono riportati i punteggi totali ottenuti dagli studenti, che richiameremo tra poco per il confronto con l'analisi IRT.

Utilizziamo gli stessi dati per svolgere l'item analisi utilizzando il programma XCalibre, prodotto da *Assessment Systems Corporation*<sup>38</sup> da cui è possibile anche scaricare una versione demo del programma. XCalibre permette di scegliere tra una serie di impostazioni per il trattamento dei dati. In particolare, per l'analisi della prova esempio, è stato impostato il modello logistico a due parametri, considerando quindi, come già detto, non solo la facilità dell'item, ma anche la sua discriminatività. Inoltre il programma permette di scegliere diverse modalità di stima per i parametri, e secondo quanto accennato in precedenza è stata selezionata la procedura MML (*Marginal Maximum Likelihood*).

In questa fase è possibile sottolineare l'evidente parallelismo di questo modello con i due principali indici dell'item analisi tradizionale.

Ecco di seguito la tabella di item analisi prodotta da XCalibre per la prova esempio.

<sup>38</sup> Il programma si può visionare ed acquistare sul sito della società [www.assess.com/index.htm](http://www.assess.com/index.htm). Si tratta del programma in dotazione al Dipartimento di Ricerche Storico-filosofiche e Pedagogiche di Roma con cui sono state svolte le analisi IRT della ricerca.

Tabella 3.2 – Item analisi prova esempio: modello IRT a due parametri

```

XCALIBRE Analysis from Data File: C:\ESERDATI.TXT
Date: Jan 13, 2005   Time: 9.21

ITEM ANALYSIS

      Endorsement Rate          Item-Theta Corr.
Item  1  2  3  4  Oth  1  2  3  4  Oth
-----
  1   11 26 47~ 16   -7 -31 52~ -27
  2   79~ 5 16  0   59~ -12 -58  0
  3    5 47 21 26~ -22 -14 -40 65~
  4   21 16 53~ 11  -51  30 29~ -15
  5   -- Deleted --
  6   16 16 53~ 16  -27 -38 68~ -28
  7    0 16 11 74~  0 -19 -44 47~
  8   32 53  5~ 11   9  0 54~ -53
  9   58~ 16 21  5   68~ -45 -36 -10
 10   -- Deleted --

```

Nelle colonne “*Endorsement rate*” sono riportate le percentuali di scelta per ogni distrattore (il programma utilizza i numeri al posto delle lettere, per cui “A” è diventata “1”) valori che naturalmente coincidono con i dati della facilità dell’item analisi classica. La risposta corretta, e di conseguenza l’indice di facilità, è segnalata dal simbolo “~” accanto al valore relativo, per cui la risposta corretta dell’item 1 è la 3 (“C”) con 47 (corrispondente al valore indice 0,47 indicato nella tabella di item analisi classica). Notiamo anche di passaggio che la colonna Oth (che comprende eventuali omissioni ed errori di risposta) non è compilata, in quanto, per gli item considerati, non si sono verificate tali condizioni.

Notiamo subito che due item sono stati cancellati e non sottoposti a analisi per mancanza di varianza. Si tratta dell’item 5, a cui hanno risposto correttamente tutti gli studenti; il dato dell’eccessiva facilità era già stato osservato con il metodo tradizionale, ma in questo caso il programma ha escluso l’item ancora prima di elaborare la tabella riassuntiva, e lo stesso ha fatto per l’item 10, perché, anche in questo caso, tutti gli studenti che hanno risposto a questo item (esclusi quindi coloro che non hanno risposto) hanno dato la risposta corretta, per cui la varianza è stata azzerata.

La seconda parte della tabella (a destra) considera la correlazione esistente fra ogni alternativa di risposta e il tratto latente della prova (theta), cioè l'abilità misurata dalla prova. La correlazione naturalmente dovrebbe essere positiva per la risposta corretta e negativa per i distrattori. Di fatto si tratta di un'estensione del concetto di discriminatività, poiché si ipotizza che le risposte corrette siano date dagli studenti più abili, secondo il modello probabilistico della curva caratteristica dell'item.

Per l'item 1 vediamo che l'unico valore positivo si ha per l'alternativa 3 ("C"), che corrisponde alla risposta corretta. Viene pertanto confermato il giudizio positivo sull'item già visto con il modello classico. Se consideriamo l'analisi relativa all'item 4, notiamo la presenza di due valori positivi: il distrattore 2 ("B") con 30 e la risposta corretta 3 ("C") con 29. Il distrattore sembra quindi misurare l'abilità indagata dalla prova in modo più efficace della risposta corretta. Ricordiamo che l'item analisi classica aveva evidenziato forti problemi di discriminatività per questo item, quindi piena coincidenza di analisi fra i due modelli. In un altro caso nella Tabella 3.2 osserviamo una correlazione positiva inattesa per l'item 8 (distrattore A), già segnalato come item critico nell'analisi classica.

Sembra quindi esserci una sostanziale coincidenza fra i due tipi di analisi, e questo di fatto permette di considerare efficace il modello di item analisi classica, almeno per quanto riguarda il principale scopo dell'analisi, il controllo della qualità degli item.<sup>39</sup>

Come per l'ultima colonna della Tabella 3.1, secondo i punteggi legati all'item analisi classica, XCalibre produce una lista con i punteggi degli studenti calcolati secondo il modello di analisi, punteggi che a differenza dell'item analisi classica vengono ponderati proprio in relazione alla correlazione item-theta, per cui ogni studente si colloca ad un livello corrispondente agli item a cui ha risposto bene, corretto rispetto agli item a cui ha risposto male. Vedremo tra poco questo tipo di punteggi.

Proseguiamo il confronto applicando il modello classico di Rasch a un solo parametro ai dati della prova esempio.

<sup>39</sup> X Calibre permette di ottenere anche una serie di osservazioni sull'andamento complessivo della prova, per esempio la verifica della attendibilità, che in questo caso conferma una certa criticità della prova esempio (coefficiente KR20 = 0,63)

In questo caso è stato utilizzato il programma Ministep<sup>40</sup>.  
Ecco la tabella di item analisi prodotta.

Tabella 3.3 – Item analisi prova esempio: modello di Rasch a un parametro

```

C:\eserdati.txt Jan 14 17:31 2005
INPUT: 19 persons, 10 items  MEASURED: 19 persons, 10 items, 2 CATS
3.55.0
-----
                item STATISTICS:  ENTRY ORDER
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|ENTRY  RAW          MODEL|  INFIT  |  OUTFIT  |PTMEA|  |
|NUMBER SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.| item |
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|   1    7    16   54.99   5.64|1.12   .6|1.01   .1| .53| 1  |
|   2   13    16   34.13   6.93|.84  -.3|.61  -.3| .59| 2  |
|   3    3    16   69.89   6.98|.96   .0|.72  -.2| .63| 3  |
|   4    8    16   51.87   5.59|1.51  2.2|1.66  1.8| .33| 4  |
|   5   16    16    7.14  18.44| MINIMUM ESTIMATED MEASURE | 5  |
|   6    8    16   51.87   5.59|.77 -1.1|.67 -1.0| .67| 6  |
|   7   12    16   38.48   6.31|1.15   .6|.95   .1| .48| 7  |
|   8    0    16   96.75  18.25| MAXIMUM ESTIMATED MEASURE | 8  |
|   9    9    16   48.77   5.61|.73 -1.4|.63 -1.1| .67| 9  |
|  10   12    12    9.13  18.64| MINIMUM ESTIMATED MEASURE | 10 |
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| MEAN    8.8   15.6   46.30   9.80|1.01   .1|.89  -.1|   |   |
| S.D.    4.5    1.2   25.24   5.68|.25   1.1|.34   .9|   |   |
-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

In questo caso la tabella prodotta non presenta analogie con l'item analisi classica, ma pur non entrando nel dettaglio di tutte le voci della tabella, notiamo subito che Ministep ha considerato non analizzabili, oltre ai soliti item 5 e 10, anche l'item 8 per l'eccessiva difficoltà (da notare i valori della colonna "MEASURE" di questi item, molto vicini agli estremi 0-1).

<sup>40</sup> Si tratta della versione demo gratuita del programma Winsteps, visionabile ed acquistabile sul sito [www.winsteps.com](http://www.winsteps.com). Ministep utilizza il modello tradizionale di Rasch a un solo parametro.

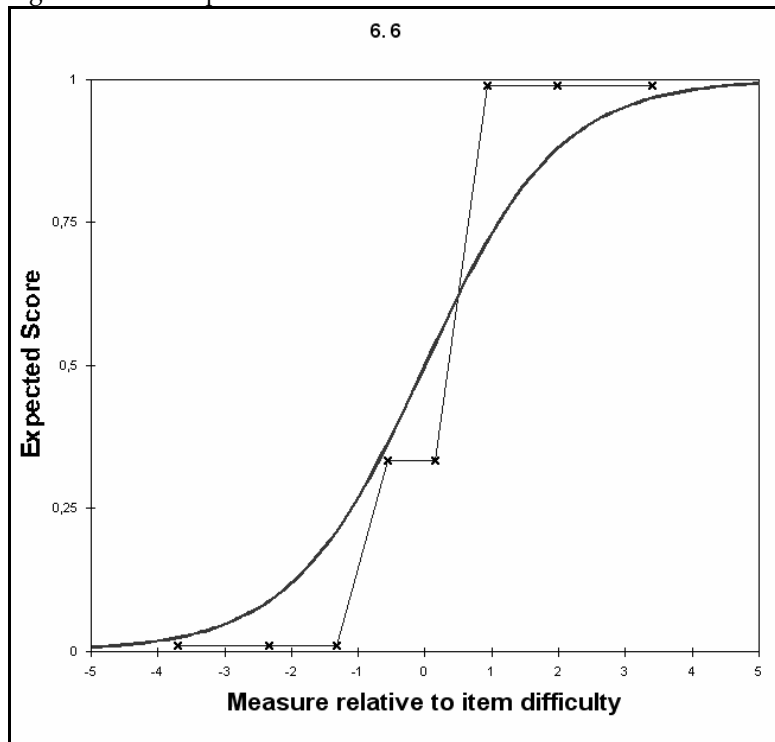
Notiamo anche che l'analisi degli item è stata fatta basandosi sulle risposte di 16 allievi (colonna "COUNT"), perché sono stati esclusi 3 studenti "estremi", cioè quelli con i punteggi più alti e più bassi (sono gli studenti 3, 11 e 19, vedi tabella dell'item analisi. Queste esclusioni hanno lo scopo di evitare effetti distorcenti di punteggi anomali rispetto alla media. Da notare anche che l'analisi dell'item 10 ha considerato solo 12 studenti rispondenti, escludendo anche i non rispondenti come già era avvenuto per XCalibre.

Ricordando quanto già notato a proposito dell'item, vediamo che anche Ministep ci segnala valori critici nelle colonne "INFIT" e "OUTFIT", che considerano rispettivamente le risposte anomale degli studenti che si trovano almeno sullo stesso livello dell'item (che quindi dovrebbero aver risposto bene) e le risposte anomale degli studenti che si trovano ad un livello inferiore (che quindi non dovrebbero rispondere correttamente). I valori sono considerati accettabili se oscillano fra 0,5 e 1,5, e notiamo che gli unici valori non accettabili sono proprio quelli relativi all'item 4 (1,51 di *INFIT* e 1,66 di *OUTFIT*), segno che sull'item le risposte rilevate non seguono l'andamento dell'abilità degli studenti, quindi non c'è coerenza con gli altri item, cioè il tratto latente della prova

La traduzione del termine inglese "to fit" circa il rapporto che si crea tra il singolo item e il tratto latente della prova (ciò che la prova misura) non risulta semplice, possiamo con una certa approssimazione parlare di coerenza dell'item, e forse risulta più chiara una verifica grafica di come un item "fitta", di come è coerente. Ministep produce dei grafici con l'andamento di ogni item rispetto alla curva caratteristica dell'item che abbiamo visto in apertura dell'unità.

Ecco la curva relativa a un item pienamente efficace, l'item 6, confrontata con l'*Item Characteristic Curve*.

Figura 3.2 – Esempio di curva di un item efficace

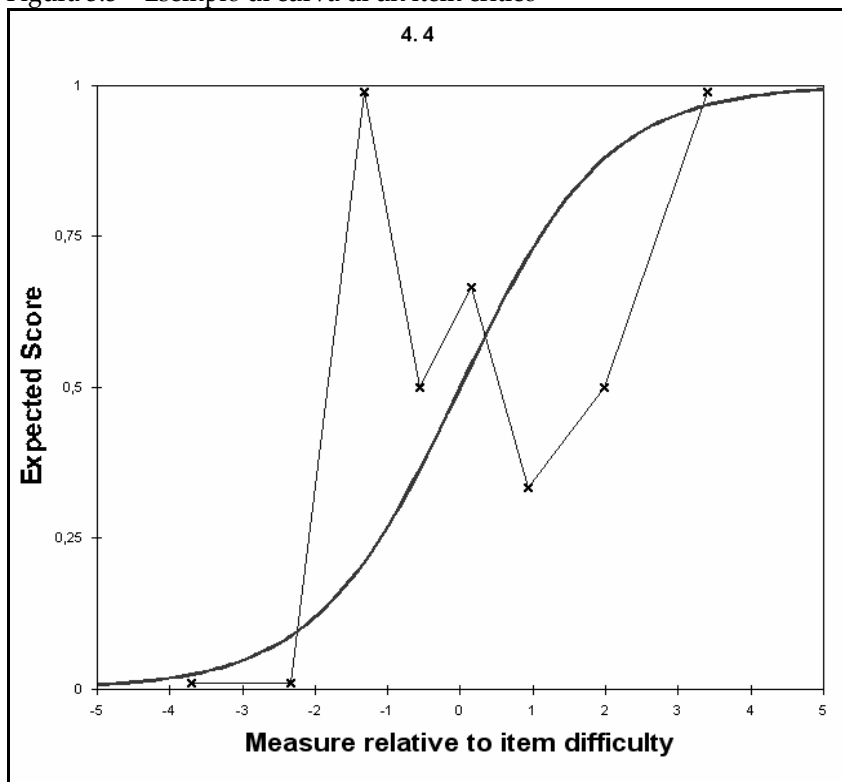


Lo scarso numero dei punteggi disponibili (solo 16 studenti) produce una curva più simile a una linea spezzata, ma è evidente che l'andamento risulta abbastanza vicino alla curva caratteristica dell'item. Il relativo abbassamento nella parte bassa della curva, e la rapida salita, sono dovuti alla forte discriminatività dell'item (pendenza della curva) già rilevata sia con l'item analisi tradizionale che con il modello a due parametri; vediamo, infatti che le probabilità di risposta restano a 0 anche per studenti che si avvicinano alla difficoltà media 0, salendo bruscamente a probabilità 1 per studenti che si trovano solo un logit sopra la difficoltà media. Da notare che, pur non considerando esplicitamente il parametro della discriminatività, proprio attraverso l'osservazione della pendenza della curva anche il modello di Rasch permette di effettuare una stima.

Vediamo ora lo stesso grafico riferito all'item 4, per il quale tutte le item analisi hanno rilevato criticità, sempre confrontato con l'*Item Characteristic Curve*.



Figura 3.3 – Esempio di curva di un item critico



Risulta estremamente chiara l'irregolarità della curva rispetto alla curva caratteristica, con diversi studenti di abilità inferiore alla media che rispondono correttamente, in misura nettamente superiore alla probabilità attesa, e di contro altri studenti, anche di due logit superiori alla media, che sono ben sotto la probabilità di risposta corretta per il loro livello di abilità.

L'item è chiaramente incoerente (non "fitta") rispetto al tratto latente della prova, e pertanto potrebbe essere escluso dalla prova.

### 3.4. Dall'item analisi ai punteggi degli studenti

Abbiamo visto che il principale vantaggio dell'analisi IRT è quello di utilizzare la stessa scala di misura per la difficoltà degli item e per l'abilità degli studenti. Questo vuol dire che ogni item fa riferimento a un certo gruppo di studenti che dovrebbero essere in grado, di rispon-

dere correttamente. Col crescere della difficoltà diminuiscono gli studenti che sono in grado di rispondere correttamente. È possibile visualizzare questa corrispondenza tra item e studenti, e lo vediamo con un esempio pratico sempre con i dati della “prova esempio” trattata con Ministep.

La “mappa” mostra un asse centrale con le relative misure riportate all'estrema sinistra, comprese fra 3 e -3 logit. Sulla parte destra trovano posto gli item, scalati a partire dall'item 8, il più difficile con un valore di 3 logit, fino agli item 10 e 5 che presentano la minore difficoltà con un valore di -3 logit. Allo stesso modo, sulla parte sinistra, trovano posto gli studenti, compresi fra i numeri 11 e 19, i più abili con un punteggio di 3 logit, e il numero 03 che si trova a -3 logit.



misura, permettendo di trarre delle conclusioni incrociate: l'item 8 ha una difficoltà compatibile con l'abilità dei soli studenti 11 e 19, mentre l'item 3 è alla portata di tutti gli studenti che si trovano più in alto nella scala, quindi oltre a 11 e 19 anche 06 e 17; gli studenti 09, 12 e 16 raggiungono un livello di abilità che dovrebbe permettergli di rispondere ai quesiti di difficoltà inferiore, cioè a tutti i quesiti esclusi il 3 e l'8, mentre lo studente 04 ha un'abilità che dovrebbe permettergli di risolvere correttamente solo i quesiti (in ordine di difficoltà decrescente) 7, 2, 10 e 5.

Da notare che questa distribuzione appaiata non sempre coincide con le effettive risposte degli studenti, ma permette di assorbire le eventuali incoerenze (INFIT e OUTFIT) delle risposte, come per esempio la risposta corretta dello studente 01 all'item 7, che si trova ad un livello di difficoltà nettamente più alto rispetto alla sua abilità.

Bisogna doverosamente dire che l'item analisi IRT viene in genere utilizzata su gruppi più ampi e su prove con un più alto numero di item rispetto alla "prova esempio", con l'ovvia riduzione dei fenomeni anomali che possono colpire distribuzioni di punteggi in cui ogni soggetto rappresenta il 5% del gruppo.

A questo proposito vediamo ora un altro esempio relativo ad un'altra prova, composta da 25 item e somministrata a 75 studenti. La prova è tratta dalla documentazione illustrativa di Ministep.

I rispondenti (KIDS) sono identificati con la prima parte del cognome. Si può notare subito che 7 di loro risultano di abilità superiore all'item più difficile (4), segno che per loro il test non era adeguato: dovrebbero aver risposto correttamente a tutte le domande e eventuali errori devono essere considerati come anomalie. All'opposto 3 item (3, 15 e 21) presentano una difficoltà inferiore all'abilità del rispondente più in basso nella scala: si tratta di quesiti troppo facili e poco utili per stimare l'abilità del gruppo.

L'osservazione che si tratta di una prova sbilanciata si ricava anche dal fatto che gli estremi variano fra 5 e -4 logit, una gamma decisamente ampia per un test.

Figura 3.5 – Misure prova modello di Ministep: modello di Rasch a un parametro

TABLE 1.0 SCIENCE ZOU710ws.txt Jan 14 17:40 2005  
 INPUT: 75 KIDS, 25 ACTS MEASURED: 75 KIDS, 25 ACTS, 3 CATS  
 MINISTEP 3.55.0

	KIDS	-MAP-	ITEMS
			<more>   <rare>
5			ROSS +
			FATA
4			PAST +
			SCHA
			T
			DOEP
3			MCLO +
			T
			WRIG
			BUFF CHAZ   4
			CLAP SEIL
			ERNS SQUR S   10
2			CLAP EAST VROO +
			BADE KENT   19
			MOOS ROSS   14 23
			MAN, PATR
			CORL SQUI   S
			BECK CLAP HOGA HSIE REIS ROSS   5
1			BLOF LAMB STOL +
			FONT LEAD TOZE M
			BABB CORL PINH ROSS SABI   2 7
			ALLE MALA SQUI   13 18 25
			AIRE BEIS CIAN DENN DYSO EISE HSIE RINZ ROSS SAND STUL   22
			ANGU MULL PAUL ROSS   9 16
0			AMIR DRIS NEIM +M
			HWA, LIEB ROSS SCHU VAN +M
			S   1 20
			BAUD   24
			BOND ROSS   6
			11
-1			LIEB NORD + 17
			JACK LAND   S 8
			T 12
			SABO
-2			+ 3
			15
			T
-3			+ 21
-4			+ <less>   <frequ>

Da notare che la distribuzione dei nomi dei rispondenti ricorda il classico grafico di distribuzione dei punteggi (ruotato di  $90^\circ$  a destra), con un addensamento intorno alla media e progressive diminuzioni verso gli estremi inferiore e superiore.

La scala in logit presenta il vantaggio di essere normalizzata, quindi è possibile confrontare direttamente item provenienti da prove diverse. Lo stesso discorso si può applicare naturalmente ai punteggi dei rispondenti, ma in questo caso con il disagio di utilizzare punteggi negativi per stimare l'abilità.<sup>41</sup>

Possiamo ora passare al confronto fra le graduatorie prodotte con i tre modelli di item analisi.

La Tabella 3.4 riepiloga i punteggi ottenuti dagli studenti nella prova esempio, ordinati secondo il punteggio base dell'item analisi classica.

<sup>41</sup> In questo caso si può procedere con una trasformazione aritmetica simile a quella utilizzata per passare dai punti  $z$  ai punti  $T$  (Boncori 1993). Anche nel caso della scala basata sui logit si può intervenire con lo stesso artificio, considerando sempre i valori di media e deviazione standard. In genere le misure ottenute con il metodo IRT vengono normalizzate con la media a 500 punti e con deviazione standard uguale a 100. Questa convenzione viene utilizzata nelle indagini internazionali (OECD 2002 e 2005) e viene considerata efficace perché impedisce di fatto di ottenere punteggi negativi e permette una ampia scalatura dei punteggi, utile per gestire graduatorie complesse.

Tabella 3.4 – Confronto punteggi prova esempio

	Punteggi		
	Modello classico	Modello IRT a due parametri	Modello IRT a un parametro (Rasch)
stud.11	10	6,6419	85,94
stud.19	9	1,9263	85,94
stud.17	8	1,0681	71,66
stud.06	8	1,2067	71,66
stud.12	7	0,5966	61,14
stud.09	7	0,5534	61,14
stud.16	6	0,4866	61,14
stud.07	6	0,0115	53,4
stud.05	6	-0,0066	53,4
stud.14	5	-0,4199	46,31
stud.13	5	-0,4963	46,31
stud.10	5	-0,5003	46,31
stud.08	5	-0,0508	53,4
stud.02	5	-0,5612	46,31
stud.18	4	-0,5226	46,31
stud.15	4	-0,6004	46,31
stud.04	4	-1,0837	38,67
stud.01	3	-1,7618	28,5
stud.03	1	-6,7326	14,78

Osserviamo le differenze esistenti fra le graduatorie. Il primo aspetto che colpisce riguarda l'esistenza dei punteggi *ex aequo* per il modello classico e il modello di Rasch, mentre per i punteggi del modello a due parametri non ricorre mai lo stesso valore. Se per il modello classico si tratta di un limite connesso al fatto che i punteggi grezzi non vengono pesati in relazione a nessun parametro, per il modello di Rasch si tratta di una conseguenza della considerazione del solo parametro della facilità, per cui gli studenti che hanno risposto allo stesso numero di domande ottengono lo stesso punteggio<sup>42</sup>. Il fenomeno si poteva già osservare nella Figura 3.4, dove item da una parte e rispondenti dall'altra risultano raggruppati su 8 livelli (corrispondenti alla gamma di punteggi per le 7 domande considerate valide dal modello di Rasch).

<sup>42</sup> Ricordiamo (Tabella 3.3) che dall'analisi erano stati esclusi i tre rispondenti e i tre item con valori estremi, per cui il confronto si svolge su 7 item, ma naturalmente vengono considerati i risultati di tutti i rispondenti. Lo scarso numero di item naturalmente rafforza la presenza di punteggi *ex aequo*.

Il modello a due parametri invece, proprio perché considera la discriminatività insieme alla facilità, permette la ponderazione di ogni item, per cui ad ogni modello di risposta corrisponde un modello specifico, e sono possibili *ex aequo* solo nel caso di rispondenti con lo stesso schema di risposte, errori compresi. E' evidente che si tratta di una eventualità abbastanza remota, anche se più probabile in caso di basso numero di item e alto numero di rispondenti. Nel nostro caso sono bastati 8 item (XCalibre ne aveva comunque esclusi due dall'analisi) per ottenere misure diverse per ogni rispondente.

Se trasformiamo la Tabella 3.5 in una equivalente tabella di graduatoria si possono osservare gli effetti laddove la prova dovesse essere utilizzata per dividere il gruppo in due parti (come nel caso di una prova di accesso). Il criterio di ordine della tabella questa volta è stato quello più specifico del modello a due parametri.

Tabella 3.5 – Confronto graduatorie prova esempio

	Graduatorie		
	Modello classico	Modello IRT a due parametri	Modello IRT a un parametro (Rasch)
stud.11	1	1	1
stud.19	2	2	1
stud.06	3	3	3
stud.17	3	4	3
stud.12	5	5	5
stud.09	5	6	5
stud.16	7	7	5
stud.07	7	8	8
stud.05	7	9	8
stud.08	10	10	8
stud.14	10	11	11
stud.13	10	12	11
stud.10	10	13	11
stud.18	15	14	11
stud.02	10	15	11
stud.15	15	16	11
stud.04	15	17	17
stud.01	18	18	18
stud.03	19	19	19



Pur non trattandosi di differenze sostanziali, si può notare almeno un caso di inversione nella graduatoria classica (stud. 02 e 18) rispetto alle graduatorie IRT, ma anche diverse “scalature” fra le due graduatorie che presentano *ex aequo* (due primi pari merito per Rasch, lo stud.16 che sale al 5° posto per Rasch, lo stud.04 che scende di due posti nella graduatoria classica).

La sostanziale coerenza dei punteggi delle graduatorie è confermata dalle forti correlazioni fra i modelli.

Tabella 3.6 – Correlazioni punteggi prova esempio

		Classica	2 Param	Rasch
Classica	Correlazione di Pearson	1	,898(**)	,979(**)
	Sig. (2-code)		,000	,000
	N	19	19	19
2 Param	Correlazione di Pearson	,898(**)	1	,891(**)
	Sig. (2-code)	,000		,000
	N	19	19	19
Rasch	Correlazione di Pearson	,979(**)	,891(**)	1
	Sig. (2-code)	,000	,000	
	N	19	19	19

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

Si nota comunque una quasi assoluta sovrapposizione fra il modello di Rasch e quello classico, in accordo con quanto già rilevato da Vincenti e Calvani (1987), ma in termini di praticità d’uso per la creazione di una graduatoria, la capacità di distinguere i punteggi dei rispondenti dimostrata dal modello a due parametri risulta più funzionale, pur confermandosi molto simile, nella distribuzione dei punteggi agli altri due modelli di analisi.

Alla luce di quanto esposto viene confermata l’opportunità, secondo quanto esposto nell’ipotesi di ricerca, di confrontare il funzionamento dell’item analisi classica con il modello IRT, segnatamente quello a due parametri, per l’attribuzione dei punteggi necessari alla formazione di una graduatoria di accesso a un corso di studi.

### 3.5. Il trattamento delle omissioni nel modello IRT

Il trattamento delle omissioni di risposta nella definizione dei punteggi ricavati per mezzo dell'item analisi IRT comporta una serie di problemi teorici. Le omissioni possono in primo luogo essere distinte per la loro posizione. Se si trovano nella parte finale della prova vengono solitamente definite domande non raggiunte (*not-reached*)<sup>43</sup> e in quanto tali non vengono considerate, per cui la competenza di chi risponde viene stimata basandosi su un numero ridotto di domande, in pratica considerando utili le risposte fino all'ultima effettivamente fornita dal rispondente.

Il presupposto di questa considerazione è che il tempo assegnato per la prova non sia stato sufficiente per arrivare alle ultime domande, assumendo che la strategia di risposta di chi svolge la prova segua sempre un andamento lineare "passo per passo", anche se non si può a rigore escludere la possibilità che le risposte procedano per salti: questo potrebbe succedere anche per situazioni accidentali, quale può essere quella in cui lo studente sfoglia la prova e trova una domanda nella parte finale che considera facile e a cui risponde di getto; se nel prosieguo della prova non raggiungerà quella domanda per mancanza di tempo, tutte le omissioni compresa fra quella domanda e l'ultima effettivamente raggiunta non saranno considerate come domande non raggiunte. Si tratta comunque di casi estremamente particolari che non possono condizionare la modalità standard di trattamento di questo particolare tipo di omissioni<sup>44</sup>.

Diversa è la situazione se le omissioni si trovano all'interno della prova, o per meglio dire tra la prima domanda e l'ultima a cui il rispondente ha fornito una risposta. In questo caso il presupposto è che non sia mancato il tempo per rispondere, ma che l'omissione sia frutto di una scelta precisa da parte del rispondente<sup>45</sup>.

<sup>43</sup> Lord (1980), che sostiene come tali domande possano essere ignorate per la stima del tratto latente della prova, in quanto non contengono informazioni esplicite sulla competenza del rispondente.

<sup>44</sup> Mitlevy e Wu (1996), dopo aver considerato le diverse condizioni che conducono al non completamento di una prova confermano l'opportunità di non tener conto delle omissioni nella parte finale di una prova per stimare il tratto latente.

<sup>45</sup> Escludiamo subito l'ipotesi che l'omissione sia dovuta alla natura stessa della prova, cioè che alcune domande non siano state effettivamente affrontate dal rispondente. È il caso di alcuni tipi di test adattativi (Little e Rubin 1987) o di test in

Sono diverse le modalità nella scelta di non rispondere a una domanda: se consideriamo il livello del rispondente, quello di buona competenza dovrebbe avere dubbi di fronte a un numero ridotto di domande, quindi avere un minor tasso di omissioni rispetto a un rispondente di scarsa competenza. Ma d'altra parte la maggiore competenza può essere associata anche a una maggiore consapevolezza circa la non conoscenza di certi argomenti, e questo può condurre a non rispondere intenzionalmente ad alcune domande, laddove il rispondente di scarso livello non avverte particolari dubbi e finisce per rispondere con maggiore scioltezza<sup>46</sup>. A tutto ciò naturalmente devono aggiungersi tutte le diverse componenti legate ai motivi affettivi che guidano chi risponde a una domanda<sup>47</sup>.

Un altro fattore che può influire sulla decisione di non rispondere è la consapevolezza da parte del rispondente circa le modalità di trattamento dei punteggi: Lord (1980) sottolinea che, pur in mancanza di istruzioni esplicite, l'eventuale conoscenza del modello IRT potrebbe condurre a una strategia di risposta che esclude le domande di cui non ci si sente più che sicuri, confidando sul buon risultato delle (poche) risposte date, mentre il meccanismo standard di risposta si basa non su una serie di alternative secche del tipo giusto-sbagliato, ma su un *continuum* di scelte, caratterizzate da un livello maggiore o minore di sicurezza, da parte del rispondente; è proprio su questo *continuum* in realtà che si distingue il livello di competenza<sup>48</sup>.

cui alcuni sottogruppi di domande vengono presentate a gruppi diversi di rispondenti, pur facendo parte dello stesso disegno di ricerca. Un chiaro esempio di quest'ultimo caso è rappresentato dall'indagine OCSE-PISA (OECD 2002).

<sup>46</sup> Stocking, Eignor e Cook (1988) hanno comunque accertato che il numero delle omissioni cresce al diminuire della competenza del rispondente, ma Mislevy e Wu (1988) hanno mostrato come la tendenza a omettere le risposte sia associabile anche a caratteristiche personali, quali l'esperienza maturata nella risoluzione dei test, la disponibilità al rischio, l'insicurezza, l'ansia di prestazione.

<sup>47</sup> Wainer e Thissen (1994) prendono in esame proprio le strategie di comportamento di uno studente di fronte a un test di valutazione, considerando da una parte le variabili di tipo cognitivo, dall'altra quelle squisitamente affettive.

<sup>48</sup> Per esempio in un quesito a quattro alternative di risposta lo studente più competente può escludere immediatamente due alternative di cui riconosce immediatamente la non plausibilità, per cui il suo margine di dubbio si riduce a due alternative: il maggior livello di competenza favorisce quindi anche gli effetti positivi di una scelta in parte casuale.

De Ayala, Plake e Impara (2001) hanno svolto uno studio per verificare gli effetti del diverso modo di considerare le omissioni secondo il modello IRT<sup>49</sup>, rilevando che l'accuratezza della stima del tratto latente resta solida fino a quando le omissioni non superano il 10% delle domande, mentre emergono diversi problemi al crescere del tasso di omissioni, soprattutto per i rispondenti che si trovano, considerando le risposte effettivamente date, nella parte più alta della scala di competenza. Nel complesso sembrerebbe preferibile ignorare le omissioni rispetto al considerarle come risposte sbagliate, trattandole quindi alla stregua delle domande non raggiunte. In alternativa gli autori ipotizzano la maggiore efficacia di una stima grezza di correttezza per ogni risposta omessa dal rispondente<sup>50</sup>

Si tratta però di indicazioni che mal si conciliano con un uso selettivo della prova, dove lo scopo non è tanto la precisione della capacità di stima del test, ma la necessità di stilare una graduatoria; di conseguenza diventa essenziale la strategia di lavoro dei rispondenti, in strettissima relazione con le istruzioni assegnate, in modo esplicito, circa le modalità di risposta e il valore delle risposte corrette, sbagliate e omesse.

Considerando la maggior parte delle istruzioni relative alle prove di accesso analizzate nel primo capitolo, che prevedono una penalizzazione delle risposte errate, è evidente che la scelta di non rispondere si basa su sottili calcoli fatti dal rispondente in relazione al rischio di rispondere male (e quindi di guadagnarsi una penalizzazione), per cui l'omissione può in diversi casi apparire il male minore, ovvero alla convenienza di azzardare comunque una risposta. Ma è evidente che con queste strategie i fattori emotivi e personali crescono di importanza rispetto alla pura e semplice competenza nell'ambito indagato dalla prova.

D'altra parte, anche eliminando la penalizzazione delle omissioni è proprio il contesto selettivo di una prova di accesso che finisce per attenuare la componente di competenza a favore delle caratteristiche personali: la motivazione legata al superamento della prova, o più corretta-

<sup>49</sup> Sono state considerate le differenze esistenti fra il metodo standard (considerare le omissioni come errori) e la non considerazione delle omissioni nella stima dei punteggi, utilizzando diverse procedure IRT.

<sup>50</sup> Lord (1980) propone di assegnare per ogni omissione un valore positivo corrispondente al punteggio previsto per la risposta corretta moltiplicato per l'inverso (frazione) del numero di alternative previsto dalla domanda. De Ayala, Plake e Impara (2001) suggeriscono invece di assegnare comunque un valore 0,5 a prescindere dal numero di alternative previste nella domanda.

mente al superamento degli altri concorrenti, guida il rispondente fin dall'ingresso nell'aula in cui si svolge la selezione, orientandolo a scegliere un posto dove sedersi piuttosto che un altro, concludendosi con l'approfittare dei momenti di naturale confusione che accompagnano la riconsegna del foglio di risposte per modificare o aggiungere qualche risposta, magari dopo una rapida occhiata al foglio di un concorrente vicino.

Negli ultimi anni è stato approfondito il tema delle competenze trasversali che s'incrociano con le specifiche competenze di ambito, influenzando, talvolta in maniera determinante, sia l'attività di apprendimento, sia le prestazioni durante le prove di verifica<sup>51</sup>. Fin dalla metà degli anni Novanta, nel quadro della costruzione di un sistema internazionale di indicatori di qualità dell'educazione, è stata avviata una riflessione sull'opportunità di considerare anche gli aspetti tipicamente legati alla psicologia cognitiva per tracciare un quadro più articolato sulla misurazione dei risultati degli studenti<sup>52</sup>. L'interesse per questo filone di ricerca ha portato alla realizzazione, nell'ambito del progetto OCSE PISA<sup>53</sup>, di uno specifico studio parallelo a quello relativo alle competenze di comprensione della lettura, di matematica e di scienze, dedicato alle competenze cross-curricolari, con lo scopo di indagare sul complesso rapporto fra le competenze e sugli effetti di alcuni fattori di tipo emozionale ed affettivo (concetto di sé, autoefficacia, ansia) sulle prestazioni degli studenti.

Non è compito specifico di questo lavoro di indagine approfondire la rete di rapporti esistente fra le diverse competenze, ma da quanto esposto risulta inevitabile considerare il risultato di una prova di accesso come un prodotto di competenze di ambito e di competenze trasversali, anche di tipo affettivo, difficilmente scomponibili, o meglio graduate in misura diversa per ogni individuo. Per questo, in accordo con i responsabili del Corso di laurea in Scienze dell'educazione dell'Università di

<sup>51</sup> Sul tema di questo complesso rapporto e sulle ricadute che può avere sul processo di apprendimento e insegnamento si possono segnalare gli importanti contributi di Ausubel (1968) e Gardner (1983).

<sup>52</sup> Ricordiamo il contributo di Walberg e Haertel (1994) nel volume dell'OCSE dedicato proprio alla definizione degli indicatori internazionali dell'educazione, raccolti periodicamente in *Education at a Glance* (OECD 2005).

<sup>53</sup> Si tratta dello studio OCSE PISA-CCC (*Cross Curricular Competencies*), realizzato parallelamente al primo ciclo di PISA (2000), con lo scopo di verificare l'autoregolazione dell'apprendimento come repertorio integrato di competenze cross-curricolari (Scalera 2002).

Roma "La Sapienza", si è deciso di analizzare i risultati della prova di accesso simulata oggetto di questa ricerca applicando alle domande non raggiunte la stessa modalità di confronto utilizzata per le omissioni nello studio di De Ayala, Plake e Impara (2001), considerandole quindi anche come omissioni da penalizzare. In tal modo la graduatoria dei punteggi ottenuta grazie all'analisi IRT si dovrebbe riavvicinare a quella basata sui punteggi grezzi, oggetto dell'item analisi classica, in cui tutte le omissioni sono considerate come risposte errate.

Questa decisione si basa anche sugli esiti della prova pilota<sup>54</sup>, in cui si è verificato che il tempo assegnato per lo svolgimento della prova è sufficiente e non dovrebbe porre particolari problemi ai nuovi iscritti, anche fuori dal contesto specifico di una prova di selezione, quindi normalmente motivati a mostrare le proprie capacità di comprensione in una prova senza alcun valore accademico.

Poiché il software utilizzato per svolgere l'analisi IRT (XCalibre) non riconosce automaticamente l'ultima risposta data dal rispondente, ma richiede codifiche distinte di inserimento dati per omissioni e domande non raggiunte, è sufficiente costruire un secondo foglio dati per l'analisi alternativa in cui i due tipi di omissioni non sono distinti. Nell'analisi dei risultati della prova di ingresso questo secondo tipo di item analisi secondo il modello IRT verrà definito IRT Om, per distinguerlo dal modello IRT standard.

Risulterà interessante confrontare i risultati delle due modalità di analisi IRT, soprattutto per quanto riguarda gli effetti sulle graduatorie dei punteggi. Naturalmente gli effetti dell'omogeneizzazione delle omissioni saranno tanto più marcati quanto più sarà alto il numero di studenti con domande non raggiunte e quante più saranno queste ultime, ricordando che i problemi di stima della competenza del rispondente cominciano a manifestarsi quando le omissioni superano il 10% dei quesiti della prova.

<sup>54</sup> Cfr paragrafo 4.3. Nella prova pilota 70 minuti sono stati sufficienti per la maggior parte degli studenti che hanno svolto la prova, per cui i 60 minuti previsti per la prova di ingresso (con la diminuzione da 39 a 30 domande) risultano adeguati.

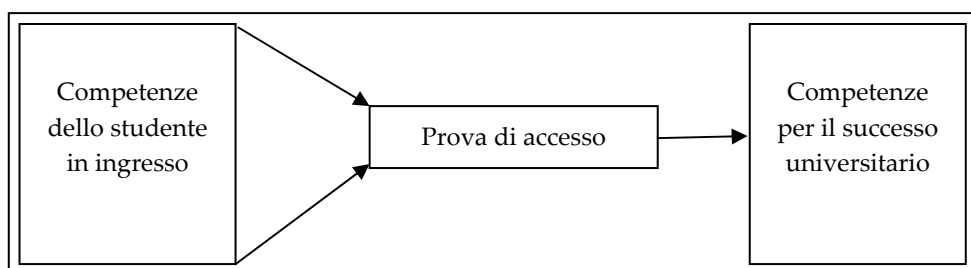
## Capitolo quarto

### Gli strumenti di rilevazione

#### 4.1. La validità della prova

Nel capitolo relativo all'item analisi sono stati approfonditi soprattutto gli aspetti legati all'attendibilità dello strumento, ma è evidente che nella definizione di una prova di accesso agli studi universitari la validità costituisce un aspetto estremamente delicato. Si tratta di capire il collegamento esistente tra le competenze dello studente, i quesiti della prova e il successo negli studi universitari. Per quest'ultimo aspetto si assume che il criterio esplicito di selezione consista nell'escludere dall'esperienza universitaria quegli studenti che forniscano meno garanzie di completamento, poiché ritenuti meno in grado di affrontare gli studi con successo. Più problematico definire le competenze in ingresso dello studente, competenze che dovrebbero essere filtrate dallo strumento di selezione secondo il criterio esplicito appena visto.

Figura 4.1 – Modello per la selezione in ingresso



Abbiamo visto, negli esempi del primo capitolo, che di solito le prove di selezione per l'accesso universitario interpretano le competenze in ingresso soprattutto come patrimonio di conoscenze possedute, per cui la base dei quesiti presenta un carattere enciclopedico, magari distinto per aree tematiche opportunamente miscelate. I temi proposti vengono talvolta anche esplicitati nei bandi che istituiscono la prova. Questa scelta comporta un primo problema di validità, poiché è possibile "prepararsi"

per sostenere la prova, ma si tratta di una preparazione, appunto, enciclopedica, che non può essere paragonata né allo studio per affrontare l'Esame di Stato, l'esperienza di selezione più prossima allo studente, né agli esami universitari, che soprattutto nella nuova versione introdotta dalla riforma, risultano sempre più specifici nei contenuti.

Di conseguenza viene considerata, fra le competenze in ingresso dello studente, solo l'abilità generale di studio e memorizzazione per sostenere un esame, che risulterà sicuramente utile nel corso degli studi universitari, ma non si può dire che ne costituisca il presupposto fondamentale. Di fatto le prove comunemente utilizzate per l'accesso finiscono per rilevare in modo indistinto da una parte le competenze complessive degli studenti (anche chi non si è preparato può ottenere discreti risultati in virtù di un solido *background* individuale), dall'altra la sua preparazione specifica negli argomenti trattati nella prova, ma i due aspetti non possono essere considerati sovrapponibili, per cui non risulta chiaro l'oggetto della valutazione. Si tratta evidentemente di un problema di validità del costrutto (Lucisano e Salerni 2002)

Dovendo affrontare l'esigenza di costruire una prova di accesso nell'ambito del Corso di laurea in Scienze dell'Educazione e della Formazione di Roma "La Sapienza", rivolta potenzialmente a grandi numeri di candidati, la riflessione dei docenti che compongono il corso di laurea è partita dall'identificazione di una competenza che risultasse strategica per affrontare l'esperienza di studio, rispetto alla semplice memorizzazione conseguente allo studio concentrato nella fase che precede immediatamente la prova.

L'esperienza delle indagini internazionali e nazionali ha condotto a identificare la competenza di lettura come l'indicatore più affidabile per stimare la preparazione complessiva degli studenti. Fin dalle prime rilevazioni svolte dalla IEA<sup>55</sup> è apparso chiaro come lo strumento migliore per comparare la preparazione complessiva degli studenti fosse costituito dalle prove di comprensione della lettura<sup>56</sup>. Nell'ultimo decennio la

<sup>55</sup> L' *International Association for the Evaluation of Educational Achievement*, realizza comparazioni internazionali dai primi anni '60 e in diverse indagini ha affrontato proprio la rilevazione delle competenze di lettura (IEA-RLS 1991-93, IEA-PIRLS 2001). Una panoramica degli studi svolti dalla IEA è presentata in Lucisano e Siniscalco (1994), e INValSI (2001).

<sup>56</sup> In tutte le ricerche internazionali in cui accanto a strumenti per misurare la comprensione della lettura sono stati usati anche strumenti per rilevare competenze più specifiche, si è rilevata una costante correlazione fra le competenze (Lucisano e



realizzazione delle comparazioni internazionali realizzate dall'OCSE ha confermato la comprensione della lettura in lingua madre come competenza strategica per stimare il profitto<sup>57</sup>: la preparazione di uno studente trova un efficace punto di sintesi nella sua capacità di comprendere un testo, capacità che richiede una complessa interazione di elementi linguistici, di conoscenze pregresse, di organizzazione del pensiero (Benvenuto, Lastrucci e Salerno 1995).

Risulta pertanto ampiamente fondata la decisione di considerare la competenza di lettura come criterio di base per la costruzione di una prova di accesso all'esperienza universitaria, considerato anche che una parte importante dello studio individuale, nel corso dell'esperienza universitaria, si realizza proprio nella lettura di testi, e che una buona capacità di comprensione può risultare motivante, ovvero che i problemi di comprensione possono risultare fonte di scoraggiamento per il completamento degli studi, assunto su cui si fonda la selezione in ingresso.

#### 4.2. La selezione dei materiali

Il rapporto con l'esperienza di studio universitario introduce il problema della validità di contenuto (Lucisano e Salerno 2002), cioè della scelta dei testi sui quali gli studenti devono essere misurati per stimare la loro competenza di lettura.

In accordo con la collega Truffelli, impegnata nella stessa rilevazione nel contesto bolognese, si è deciso di selezionare una serie di prove di comprensione della lettura già sperimentate in contesti limitati, ma inedite, all'interno di corsi di laurea di Scienze dell'Educazione, fondate su brani tratti da testi adottati per alcuni insegnamenti. Si tratta pertanto di brani che gli studenti possono effettivamente incontrare nella loro esperienza di studio. Naturalmente il problema tecnico principale riguarda l'estrazione da testi di una certa ampiezza di brani che si autosostengono (Benvenuto, Lastrucci e Salerno 1995, Osterlind 2001), cioè che permetto-

Siniscalco (1994), in particolare per l'indagine IEA *Written Composition* e per l'indagine IEA *Civic Education*.

<sup>57</sup> Ci riferiamo in particolare all'indagine IALS SIALS (Gallina 2000) e ai diversi cicli di PISA (OECD 2001 e OECD 2004), che hanno confermato le correlazioni esistenti fra competenza di lettura e altre competenze, come quelle di tipo logico e matematico.

no di costruire una serie di domande senza rimandi esterni al brano (ma interni al testo originale) troppo vincolanti.

Al termine dello spoglio sono stati selezionati sei brani, per un totale di 39 domande a scelta multipla con 4 alternative di risposta, ed è stato pertanto composto il primo strumento per la rilevazione, il TCL (Test di Comprensione della Lettura). Considerando che si trattava di prove già utilizzate in contesti diversi, con misure di validazione spesso limitate a pochi soggetti, si è ritenuto opportuno sottoporre lo strumento a verifica in una prova pilota specifica, per verificarne l'efficacia prima della somministrazione definitiva.

Nello schema seguente sono riepilogati i brani utilizzati, considerando anche la tipologia testuale di riferimento (Benvenuto, Lastrucci e Salerno 1995)

Tabella 4.1 – Riepilogo dello strumento TCL per la prova pilota

Testo	Tipologia	N. item
Legge sul prestito d'onore	Informativo	4
Assunzioni previste nelle imprese	Pragmatico	8
La rappresentazione nella professione	Pragmatico - Scientifico	3
Dallo Statuto dei lavoratori	Informativo	7
La Francia e le trentacinque ore	Informativo	5
Introduzione al saggio di John Dewey	Informativo - Scientifico	12

Il fascicolo TCL era introdotto da una pagina di istruzioni costruite *ad hoc* per illustrare le modalità della prova, le modalità di risposta, i tempi previsti per il completamento del lavoro.

Sempre in un'ottica di validità del contenuto, si è ritenuto utile affiancare allo strumento TCL basato su domande a scelta multipla uno strumento di comprensione basato sul Cloze, cioè sul riempimento di un testo "bucato" (Lucisano 1989), ricostruendo dal contesto i termini mancanti.

Questa particolare tecnica per misurare la comprensione della lettura è stata definita da Taylor (1953), ed è caratterizzata da una estrema facilità

tà di realizzazione, poiché consiste nell'eliminare da un testo una serie di parole che il lettore deve ricostruire. La tecnica di bucatatura adottata è stata quella più semplice (Marello 1984, Nuccorini 2001), con l'eliminazione sistematica di una parola ogni cinque, lasciando intatta però la prima parte del brano per permettere al rispondente di avere un contesto di riferimento anche per la ricostruzione dei primi termini mancanti.

Il fatto che la competenza di lettura risulti misurabile anche da una prova Cloze permette di avere due stime diverse sullo stesso oggetto, per cui la misurazione può essere considerata sia più valida che più affidabile<sup>58</sup>. Purtroppo la decisione di utilizzare anche il Cloze è stata presa troppo a ridosso della realizzazione della prova pilota, per cui non erano ancora stati scelti i testi di riferimento e il collaudo è stato svolto su un pacchetto di quattro testi rapidamente concordati con gli insegnanti della scuola di Roma in cui si è svolta la prova. Anche se i testi utilizzati nella somministrazione definitiva sono stati diversi (tranne uno recuperato proprio dal pacchetto pilota), si è comunque trattato di un utile verifica sulle modalità e i tempi di somministrazione di una prova basta sulla tecnica Cloze.

Ecco il quadro riassuntivo delle prove Cloze utilizzate a Roma durante la prova pilota. Si tratta di quattro testi tratti da manuali, quindi classificabili tutti come testi informativi scientifici, ma su quattro tematiche ben distinte che coprivano le aree di contenuto degli indirizzi di studio della scuola in cui si è svolta la prova pilota.

Tabella 4.2 – Riepilogo dello strumento Cloze per la prova pilota

Testo	Tematica	N. buchi
Incremento demografico nell'800	Storica	32
I Sepolcri di Foscolo	Letteraria	32
Lettera di Einstein a Freud	Scientifica	30
Epistola sulla tolleranza di Locke	Filosofica	31

Accanto agli strumenti dedicati alla rilevazione della competenza, sono stati predisposti due questionari per la rilevazione delle variabili di

<sup>58</sup> Lucisano (1989) segnala diverse ricerche che, a partire dagli anni Cinquanta, hanno accertato una forte correlazione tra i risultati delle prove Cloze e di prove di comprensione della lettura basate su domande a scelta multipla, somministrate agli stessi rispondenti.

sfondo secondo cui interpretare i risultati. Il primo è stato elaborato sulla base del "Questionario nuovi iscritti" in uso presso il Corso di laurea in scienze dell'Educazione e della Formazione di Roma "La Sapienza" (Benvenuto 2003). Il questionario è stato modificato con la semplificazione di alcune domande relative alle motivazioni della scelta del corso di studi e l'inserimento di alcune informazioni aggiuntive riguardanti la famiglia; tali modifiche hanno tenuto conto anche dell'utilizzo parallelo del secondo strumento messo a punto da Truffelli per la sua ricerca, il fascicolo "Punti di vista sulla scelta universitaria", composto essenzialmente da domande aperte (i dati provenienti da questo secondo questionario non sono stati utilizzati nel presente lavoro di ricerca).

Considerando la snellezza del questionario e la sua precedente utilizzazione, si è ritenuta inutile la sua verifica all'interno della prova pilota, anche perché molte delle domande potevano risultare poco adatte ai rispondenti (ancora iscritti nella scuola secondaria).

### 4.3. La realizzazione della prova pilota

Alcuni dei materiali selezionati per l'indagine, il TCL e il questionario "Punti di vista sulla scelta universitaria", sono stati collaudati preliminarmente in una prova pilota che si è svolta nel mese di maggio 2003: le prove di comprensione a scelta multipla e la prova scritta aperta sono stati somministrati in 9 classi di ultimo anno della scuola secondaria superiore di Roma e Bologna, (tutte in istituti a indirizzo psicopedagogico). Questa popolazione di riferimento è stata scelta in quanto rappresenta quella più vicina a quella effettivamente coinvolta nella somministrazione principale<sup>59</sup>, oltre ad essere quella praticamente raggiungibile nei tempi di lavoro previsti.

Nel complesso gli studenti coinvolti sono stati 168 (83 a Roma e 85 a Bologna). Nell'ambito di un'altra iniziativa di collaborazione fra l'Università di Roma "La Sapienza" e la scuola romana in cui si è svolta la prova pilota, è stata somministrata anche una prova di comprensione Cloze in 4 classi, per un totale di 72 studenti<sup>60</sup>. Come già detto al mo-

<sup>59</sup> Anticipando un dato presente nella Tabella 5.5, le matricole provenienti dall'indirizzo di studi pedagogici sono il 29% del totale, e rappresentano il gruppo più consistente.

<sup>60</sup> La realizzazione della prova pilota a Roma è stata possibile grazie alla partecipazione dell'IIS "Carducci", che collabora da tempo con il Corso di laurea in Scienze dell'Educazione e della Formazione. Per la realizzazione della prova pilota svolta a Bologna si rimanda al lavoro di Truffelli.

mento dello svolgimento della prova pilota non era ancora stata stabilita l'effettiva utilizzazione di una seconda prova di comprensione della lettura (cfr. 4.5.3), ma anche le informazioni raccolte durante le somministrazioni della prova Cloze sono risultate utili per la definizione del pacchetto complessivo di indagine.

Le somministrazioni svolte a Roma sono state realizzate con l'aiuto di un gruppo di laureandi collaboratori del Corso di laurea in Scienze dell'Educazione e della Formazione<sup>61</sup>: a questo proposito sono state predisposte alcune schede di lavoro per guidare le somministrazioni e garantire la qualità dei dati raccolti. Inoltre sono state realizzate due riunioni successive alle somministrazioni per verificarne l'andamento ed evidenziare gli aspetti critici emersi nel corso del lavoro. Il collaudo ha permesso quindi di gettare le basi per la struttura operativa delle somministrazione principale, che come vedremo in seguito è risultata particolarmente impegnativa a Roma per l'alto numero di candidati coinvolti.

Le somministrazioni di Bologna sono state curate dalla collega Truffelli, che ha successivamente raccolto tutte le prove scritte aperte realizzate dagli studenti (comprese quindi quelle di Roma) per il suo studio di dottorato.

Particolare importanza è stata data ai dubbi posti dagli studenti durante la somministrazione, che hanno permesso di identificare alcuni nodi da sciogliere nei testi utilizzati per la rilevazione e nelle istruzioni per la compilazione. Estremamente importante è stata poi la rilevazione dei tempi effettivi di svolgimento delle diverse prove, verificando lo stato di attenzione e impegno da parte degli studenti. In particolare si è verificato che la quasi totalità degli studenti ha completato la prova TCL entro 70 minuti, a fronte di un tempo massimo previsto di 60 minuti, ma con la possibilità, visto che lo scopo principale era il collaudo della prova, di utilizzare qualche minuto in più; invece per la prova Cloze, nelle classi in cui è stata realizzata, i 60 minuti assegnati si sono rivelati sufficienti per la maggior parte degli studenti per il completamento dei quattro testi.

Le risposte degli studenti alla prova di comprensione della lettura TCL sono state inserite in un database Excel (per le prove compilate a Roma è stato utilizzato uno scanner per la lettura ottica dei moduli di ri-

<sup>61</sup> Un ringraziamento particolare va a Serena Cacciamani, Agnese Marinucci, Monia Montana, Francesca Pensa, Stefano Scippo e Sara Sonnino che hanno collaborato nel corso delle somministrazioni e dell'inserimento dati.

sposta, che segnala anche le imprecisioni di lettura rilevate sul modulo stesso) e si è successivamente proceduto alla pulizia dei dati, verificando le incongruenze di risposta del database direttamente sui fogli di risposta originali. In tal modo è stato possibile praticamente azzerare qualsiasi tipo di distorsione dovuta all'inserimento dei dati.

Le risposte degli studenti nella prova Cloze sono state ugualmente inserite in un database Excel, ma distinto in quattro fogli di lavoro (uno per ogni testo), sfruttando la funzione di riempimento automatico offerta dal programma che permette di accelerare l'inserimento. La fase di pulizia dei dati inseriti in questo caso è stata svolta procedendo a successive fasi di riordinamento automatico dei database, che hanno permesso di eliminare errori di inserimento dovuti da difettosa digitazione dei termini. Trattandosi di una procedura molto più lenta rispetto all'inserimento con lettore ottico, l'esperienza accumulata in questa fase si è rivelata preziosissima per impostare il lavoro di inserimento dei dati relativi alla somministrazione principale.

Tutte le prove del fascicolo "Punti di vista sulla scelta universitaria", comprese quelle collaudate a Roma, sono state invece inserite dalla collega Truffelli in un database specifico, strutturato per essere utilizzato con uno specifico programma di analisi del testo<sup>62</sup>.

Al termine della fase di inserimento e pulizia dei dati si è passati all'analisi degli item che compongono il TCL, realizzata secondo il modello classico e con il modello IRT.

#### 4.4. Item analisi della prova pilota (TCL)

L'elaborazione secondo il modello classico è stata effettuata con il foglio elettronico Excel, che, come abbiamo visto nel precedente capitolo, grazie alle formule matrici può effettuare il conteggio delle alternative di risposta, nonché i calcoli necessari per gli indici di facilità e di discriminatività (calcolato con il confronto degli estremi).

Per l'analisi IRT è stato utilizzato XCalibre, impostato per applicare il modello a due parametri, stimati attraverso la procedura di stima MML.

Le due modalità di analisi sono messe a confronto nella Tabella 4.3.

<sup>62</sup> Per inciso osserviamo che è stato utilizzato il programma TALTAC, Trattamento Automatico Lessico-Testuale per l'Analisi del Contenuto.

Tabella 4.3 - Item analisi TCL (Prova pilota)

Item*	Modello classico							Modello IRT (2 parametri)						num. Item#
	A	B	C	D	Omis.	Fac.	Dis.	A	B	C	D	Omis.	PBt	
1 C	28,7	10,2	35,3	22,8	3,0	0,35	0,44	-11	-3	40	-23	-19	0,40	1
2 A	13,8	29,9	29,9	25,7	0,6	0,14	0,07	7	14	-13	-1	-22	0,07	Mod.2
3 D	3,6	11,4	18,6	63,5	3,0	0,63	0,33	-10	-20	-9	32	-21	0,32	3
4 B	39,5	24,0	31,7	2,4	2,4	0,24	0,38	-14	39	-9	-11	-22	0,39	4
5 B	2,4	42,5	0,6	53,9	0,6	0,42	0,53	-6	42	2	-35	-25	0,42	5
6 B	6,6	81,4	5,4	2,4	4,2	0,81	0,18	-10	22	-15	-7	-9	0,22	Elim.
7 A	76,0	1,2	5,4	16,8	0,6	0,76	0,20	21	6	-5	-15	-26	0,21	6
8 A	76,0	4,2	6,6	11,4	1,8	0,76	0,24	30	-4	-19	-12	-24	0,30	Mod.7
9 D	7,8	0,6	1,8	85,6	4,2	0,86	0,31	-18	-11	-22	41	-28	0,41	8
10 D	25,7	1,8	5,4	64,1	3,0	0,64	0,47	-18	-10	-18	39	-28	0,39	9
11 D	1,2	4,8	7,8	81,4	4,8	0,81	0,31	-4	-12	-20	39	-31	0,39	10
12 D	4,8	4,2	9,0	76,0	6,0	0,76	0,42	-12	-3	-21	36	-24	0,36	11
13 D	5,4	33,5	4,2	52,7	4,2	0,53	0,36	0	-18	-11	35	-32	0,35	12
14 B	4,2	59,9	25,7	4,8	5,4	0,60	0,47	-5	31	-6	-22	-30	0,31	13
15 D	13,2	34,1	8,4	29,9	14,4	0,30	0,27	-2	12	-15	23	-32	0,23	14
16 C	12,6	3,6	63,5	18,6	1,8	0,63	0,20	-18	-11	17	-10	-27	0,17	Elim.
17 B	1,8	63,5	25,7	6,0	3,0	0,63	0,20	-11	16	-3	1	-28	0,16	Elim.
18 C	41,9	3,0	47,9	4,8	2,4	0,48	0,16	11	-18	9	-18	-19	0,09	Elim.
19 A	83,8	3,0	5,4	6,0	1,8	0,84	0,24	29	-9	-12	-11	-26	0,29	Elim.
20 B	4,8	56,3	29,9	6,0	3,0	0,56	0,53	-6	34	-17	-13	-25	0,34	Elim.
21 A	92,2	1,2	1,8	3,0	1,8	0,92	0,22	38	-9	-16	-17	-29	0,38	Elim.
22 C	14,4	3,6	80,2	0,6	1,2	0,80	0,29	-18	-16	35	-7	-32	0,35	Elim.
23 C	9,0	4,2	76,6	4,8	5,4	0,77	0,51	-22	-17	44	-13	-26	0,44	15
24 B	1,8	58,1	23,4	4,2	12,6	0,58	0,44	-8	31	-7	-7	-30	0,31	Mod.18
25 C	7,8	11,4	25,3	49,1	5,4	0,26	0,20	-16	-18	14	20	-27	0,14	Mod.16
26 D	25,1	28,7	11,4	26,3	8,4	0,26	0,11	-12	23	-8	11	-25	0,11	Mod.17
27 A	44,3	31,1	2,4	16,2	6,0	0,44	0,51	36	-10	-13	-9	-34	0,36	19
28 B	9,6	44,3	24,6	12,6	9,0	0,44	0,40	4	34	-13	-7	-34	0,34	20
29 D	9,6	4,2	6,6	75,4	4,2	0,75	0,58	-32	-14	-19	58	-38	0,58	21
30 D	2,4	21,6	4,2	64,7	7,2	0,65	0,58	-10	-18	-12	47	-43	0,47	22
31 A	35,3	32,3	10,8	13,8	7,8	0,35	0,44	37	-1	-15	-5	-38	0,37	23
32 C	1,8	5,4	77,8	4,2	10,8	0,78	0,47	-14	-15	50	-11	-43	0,50	24
33 B	26,3	40,1	16,2	8,4	9,0	0,40	0,18	7	17	-2	4	-40	0,17	Elim.
34 C	7,2	6,0	67,7	8,4	10,8	0,68	0,51	-1	-19	45	-19	-36	0,45	26
35 D	10,8	9,6	7,8	62,3	9,6	0,62	0,51	-16	-9	-15	48	-39	0,48	25
36 A	62,9	10,2	4,2	11,4	11,4	0,63	0,69	56	-26	-7	-17	-39	0,56	27
37 C	7,2	9,6	60,5	12,0	10,8	0,60	0,51	-17	-8	48	-12	-41	0,48	28
38 B	6,0	58,1	12,0	12,6	11,4	0,58	0,42	-7	34	-10	6	-42	0,34	29
39 A	79,0	4,2	6,0	1,8	9,0	0,79	0,49	54	-13	-20	-21	-40	0,54	Mod.30

\*Accanto ad ogni numero identificativo dell'item è indicata l'alternativa corretta

#Nuova numerazione degli item sul fascicolo dello studio principale; segnalazione degli item eliminati (Elim.) o modificati (Mod.)

Nella parte sinistra della tabella è riportata l'analisi degli item secondo il modello classico, con le percentuali di scelta per ogni alternativa (colonne A-D) e per le omissioni (Omis.). Notiamo di passaggio che non si è resa necessaria una colonna per riportare eventuali errori di risposta<sup>63</sup>, poiché la chiarezza delle istruzioni, la tecnica adottata per raccogliere le risposte (modello ottico) e un'accurata fase di pulizia dei dati hanno escluso qualsiasi errore di questo tipo.

Nella parte destra è riportata l'analisi effettuata con il modello IRT. Le prime colonne (A-D e Omis.) riportano le correlazioni esistenti fra ogni alternativa e il tratto latente della prova ( $\theta$ ), che dovrebbe essere positiva e il più alta possibile per l'alternativa corretta, negativa per i distrattori e le omissioni. L'ultima colonna (PBt) riporta la misura equivalente del punto biseriale per il metodo IRT, cioè la correlazione (Pearson) fra le risposte esatte/non esatte e il tratto latente della prova (e non il punteggio totale come per il punto biseriale). Sono state evidenziate in grigio i valori critici.

L'analisi svolta secondo il modello classico evidenzia una serie di item che mostrano problemi, ed in particolare:

gli item 2, 6, 18, 26, 33, con un indice di discriminatività inferiore a 0,20, sono senza dubbio i più critici, in quanto non distinguono con chiarezza gli studenti più abili da quelli che hanno fornito meno risposte alle prove;

gli item 6 (ancora), 9, 11, 19, 21, 22, con un indice di facilità superiore a 0,80 si rivelano troppo facili per la popolazione cui è indirizzata la prova;

al contrario l'item 2 (già segnalato) con un indice di facilità 0,14 risulta essere troppo difficile;

negli item 2, 4, 5, 15, 25, 26, la risposta corretta presenta una percentuale di scelta inferiore a una o più alternative di risposta errate;

Parallelamente l'item analisi IRT segnala problemi molto simili a quelle già osservate nel modello classico:

gli item 2, 16, 17, 18, 25, 26 e 33 presentano un punto biseriale (PBt) molto basso;

gli item 2, 18, 25 e 26 sono segnalati a rischio errore di chiave di risposta;

<sup>63</sup> A differenza delle omissioni, gli errori di risposta comportano un'indicazione non univoca da parte del rispondente (per esempio vengono scelte due alternative)



gli item 2, 5, 7, 15, 17, 18, 25, 26, 28, 33, 38 presentano un distrattore con correlazione positiva.

Da sottolineare comunque che, secondo i parametri previsti da XCA-LIBRE, nessun item è stato eliminato dall'analisi, segno che anche in presenza di qualche valore critico, tutti gli item possono essere considerati per la ponderazione dei risultati ottenuti dai rispondenti.

Nel complesso l'intero pacchetto presenta una facilità media di 0,58 articolata nella tabella seguente.

Tabella 4.4 – Facilità media dei diversi testi di comprensione della lettura

Testo	N. item	Posiz. item	Facilità media
Legge sul prestito d'onore	4	1-4	0,34
Assunzioni previste nelle imprese	8	5-12	0,73
La rappresentazione nella professione	3	13-15	0,48
Dallo Statuto dei lavoratori	7	16-22	0,69
La Francia e le trentacinque ore	5	23-27	0,46
Introduzione al saggio di John Dewey	12	28-39	0,61

Una notazione particolare riguarda le percentuali di risposta omessa: crescono negli item finali, a dimostrazione che il tempo effettivamente assegnato (70 minuti in luogo dei 60 previsti dalle istruzioni) è risultato critico per circa il 10% dei rispondenti, anche se gli studenti che non hanno risposto a tutte le ultime cinque domande sono solo il 7%, e che le omissioni dell'ultima domanda si riducono rispetto alle cinque precedenti, a dimostrazione che diversi studenti hanno saltato domande ritenute difficili, preferendo rispondere all'ultima, oggettivamente più facile. Questo andamento conferma quindi i problemi evidenziati nel Paragrafo 3.5 circa il trattamento delle omissioni in relazione alle possibili strategie di risposta.

Da notare al riguardo anche che due item intermedi (15 e 24) sono stati saltati da oltre il 10% degli studenti, e se il primo dei due risulta essere leggermente difficile e con un distrattore fin troppo efficiente, per l'altro non sembrano esserci motivi particolari che spieghino la crescita delle omissioni.

Il controllo complessivo della facilità delle domande in una prova fa riferimento a uno degli aspetti più critici dei test lineari, la capacità di

stimare soggetti con livelli di difficoltà molto diversi tra loro, condizione caratteristica di una prova indirizzata a un gruppo ampio di rispondenti. Se quindi la facilità ideale di un singolo item si dovrebbe collocare intorno a 0,50, è opportuno che la prova presenti quesiti a diversi livelli di facilità, in modo tale che ogni quesito identifichi la facilità media per gruppi di studenti di diverso livello. Solo con un'ampia gamma di indici di facilità la prova può riuscire a distinguere rispondenti che presentano differenti livelli di competenza.

Bisogna considerare anche che nel corso della compilazione della prova la presenza di domande troppo difficili, per alcuni soggetti, può comportare un calo della motivazione con relativa interferenza sulla performance, per cui il controllo complessivo della facilità della prova, in termini di equilibrio delle difficoltà relative dei singoli item, può contribuire a migliorarne l'affidabilità. Si tratta di un tema ben presente per chi deve costruire prove di comprensione (Boncori 1993) soprattutto se devono essere utilizzate su un gruppo di rispondenti particolarmente ampio (OECD 2002).

L'analisi IRT non risolve questo specifico problema, perché l'assunto che guida il giudizio su ogni item è che la sua scelta è tanto più opportuna quanto più si può collegare all'abilità del rispondente, in tal modo quell'item risulta maggiormente informativo: in altre parole se uno studente risolve un quesito otteniamo un'informazione su una sua capacità, se non lo risolve otteniamo una non-informazione, cioè non possiamo stimare quanto non sia capace. Il problema si è posto nell'indagine OCSE-PISA 2000 sulla comprensione della lettura, in cui sono stati previsti, sulla base dei risultati delle prove analizzati con l'IRT, 5 livelli di competenza dei quindicenni<sup>64</sup>.

A posteriori però ci si è resi conto che una minoranza apprezzabile di loro (6%, con forti differenze da Paese a Paese) non raggiungeva nemmeno la soglia del primo livello della scala. Evidentemente per questi studenti la compilazione della prova deve essere risultata un'esperienza molto frustrante, e anche per il 12% che ha raggiunto solo il primo livello non deve essere stato piacevole confrontarsi con domande prevalentemente incomprensibili.

<sup>64</sup> I cinque livelli, frutto di un calcolo specifico di valori soglia basato sulla distribuzione dei punteggi (OECD 2001), ha permesso di classificare gli studenti in termini di profili di lettura (cosa sa fare il lettore di ogni livello).

Per concludere la verifica sulla prova TCL sottoposta a collaudo, il calcolo del coefficiente Kuder Richardson (KR 20<sup>65</sup>) ha portato ad un risultato di 0,81, per cui l'intero pacchetto risulta essere sostanzialmente affidabile.

In questa fase non sono stati considerati i punteggi degli studenti secondo i due metodi, poiché lo scopo principale dell'analisi degli item era la verifica di efficienza degli strumenti.

#### 4.5. Allestimento degli strumenti per lo studio principale

Alla luce dell'analisi svolta sui dati della prova pilota, considerando anche le informazioni raccolte circa i tempi di compilazione delle prove, è stato definito il pacchetto definitivo di strumenti da sottoporre ai nuovi iscritti dei corsi di laurea di Scienze dell'Educazione delle università di Roma "La Sapienza" e di Bologna. Il lavoro di selezione è stato svolto in stretta collaborazione con la collega Truffelli e sotto la guida dei docenti tutor. Il pacchetto degli strumenti per la somministrazione definitiva è riassunto nella Tabella 4.5, nell'ordine di somministrazione e con i tempi previsti.

Tabella 4.5 – Strumenti per lo studio principale

Strumento	Sessione	Tempo
Prova di comprensione della lettura con quesiti a scelta multipla TCL	I	60 min.
Questionario "Punti di vista sulla scelta universitaria"	I	45 min.
Pausa	Circa 50 min.	
Prova di comprensione della lettura Cloze	II	60 min.
Questionario per i nuovi iscritti.	II	15 min.

Tra le prove di comprensione e i questionari sono previsti due brevi pause (5 minuti).

<sup>65</sup> Il coefficiente  $\alpha$  (di Cronbach) può essere calcolato, per le prove composte da item di tipo binario, anche attraverso il coefficiente KR20, utilizzato da XCalibre (Lord, Novick 1968)

Di seguito vengono presentati i quattro strumenti utilizzati, con particolare riguardo alle prove di comprensione.

Una breve notazione circa i tempi di somministrazione: l'esperienza della prova pilota ha suggerito di prevedere un impegno continuativo per i rispondenti non superiore alle due ore, per cui la rilevazione è stata suddivisa in due sessioni, ognuna aperta con uno strumento prova di comprensione e completata con uno strumento questionario. La soglia delle due ore, interrotte solo da un breve intervallo, è stata verificata dalle recenti indagini PISA (OECD 2001 e 2004), in cui gli studenti, quindi, devono affrontare la compilazione dei fascicoli di prova nelle prime due ore, e la compilazione dei questionari dopo una congrua pausa che permette di recuperare lo sforzo.

#### 4.5.1. *Prova di comprensione della lettura con quesiti a scelta multipla TCL*

Per la prova di comprensione con quesiti a scelta multipla (TCL) sono stati selezionati trenta item fra quelli collaudati nella prova pilota. Considerando i risultati dell'analisi degli item sono state effettuati i seguenti interventi:

- eliminazione della prova "Dallo statuto dei lavoratori"; si tratta di un testo normativo abbastanza simile al primo ("Legge sul prestito d'onore"), ma con indici complessivamente peggiori (facilità media più alta, discriminatività quasi sempre inferiore a 0,30, scarso peso delle risposte corrette secondo l'analisi IRT);

- eliminazione della domanda 6, dal testo "Assunzioni previste dalle imprese", in considerazione degli indici critici rilevati nell'analisi, della difficoltà di modifica dell'item mantenendo inalterato l'oggetto del quesito e considerando che comunque le domande del brano restano 7, un numero adeguato per verificare la comprensione di un testo in formato tabella;

- eliminazione della domanda 33 dal testo "Introduzione al saggio di John Dewey", che presentava una bassa discriminatività e valori contraddittori per diversi distrattori, considerando anche che per il brano erano previste 12 domande;

- modifica dei quesiti 2, 8, 25 e 26, in considerazione del livello critico degli indici di analisi degli item, con interventi sui distrattori che hanno creato maggiori problemi agli studenti;

- modifica del quesito 24, pur con indici positivi, considerando le osservazioni effettuate durante la somministrazione dello studio pilota, apportando di conseguenza piccoli interventi linguistici su alcuni distrattori per migliorarne la comprensione;
- eliminazione dell'ultimo capoverso del brano "La Francia e le trentacinque ore", considerando che le domande non riguardano direttamente tale parte e che la conseguente non lettura non compromette la comprensione della prima parte del brano (si tratta di un taglio consistente, circa un quarto dell'intera lunghezza del brano).

Gli interventi sono stati già evidenziati nell'ultima colonna a destra della tabella riassuntiva di item analisi (Tabella 4.3), utile anche per verificare i piccoli spostamenti nell'ordine degli item dovuti alla nuova impaginazione del fascicolo TCL definitivo.

Per effetto di queste modifiche le domande della prova scendono a trenta, e conseguentemente si riduce l'impegno per gli studenti. Anche dal punto di vista grafico le pagine del fascicolo contenenti le prove si riducono da 10 a 6 (più la copertina con le istruzioni), per cui è stato fissato un tempo di lavoro, non derogabile, di 60 minuti.

Nella Tabella 4.6 viene riepilogata la struttura definitiva del TCL considerando i brani e la loro tipologia<sup>66</sup>, il numero identificativo di ogni item e la relativa tipologia<sup>67</sup>. Si tratta di una classificazione indicativa, in quanto la relativa brevità della prova non permette certo di considerare l'esistenza di sottoscale specifiche relative a una tipologia specifica. Si è semplicemente voluto garantire una accettabile varietà di testi e quesiti per una singola scala sintetica di competenza di lettura.

<sup>66</sup> È stata utilizzata la classificazione prevista in Benvenuto, Lastrucci, Salerno (1995), con la distinzione in testi informativi e pragmatici, distinguendo per entrambi i tipi la presenza di elementi chiaramente scientifici.

<sup>67</sup> È stata considerata l'articolazione nelle quattro categorie principali identificate nell'Indagine IEA SRL (SAL per l'Italia), che prevede la distinzione fra quesiti finalizzati al riconoscimento delle parole (lessicali), alla comprensione globale del testo, (individua l'idea principale), alla comprensione particolare del testo (localizza le informazioni) e alla comprensione di inferenze. Non sono stati considerati item finalizzati a parafrasare e seguire istruzioni (Lucisano 1994).

Tabella 4.6 – Struttura della prova di comprensione TCL

Testo	Tipologia testo	N. item	Tipologia item
Legge sul prestito d'onore	Informativo	1	Padronanza lessicale
		2	Localizzazione informazioni
		3	Individuazione idea principale
		4	Localizzazione informazioni
Assunzioni previste nelle imprese	Pragmatico	5	Localizzazione informazioni
		6	Inferenza
		7	Localizzazione informazioni
		8	Inferenza
		9	Localizzazione informazioni
		10	Localizzazione informazioni
La rappresentazione nella professione	Pragmatico Scientifico	11	Inferenza
		12	Inferenza
		13	Individuazione idea principale
La Francia e le trentacinque ore	Informativo	14	Inferenza
		15	Localizzazione informazioni
		16	Localizzazione informazioni
		17	Localizzazione informazioni
		18	Localizzazione informazioni
Introduzione al saggio di John Dewey	Informativo Scientifico	19	Localizzazione informazioni
		20	Localizzazione informazioni
		21	Localizzazione informazioni
		22	Individuazione idea principale
		23	Inferenza
		24	Inferenza
		25	Inferenza
		26	Localizzazione informazioni
		27	Localizzazione informazioni
		28	Padronanza lessicale
		29	Padronanza lessicale
		30	Localizzazione informazioni

Poiché la funzione esplicita della prova riguarda la selezione in ingresso, l'articolazione dei testi e degli item può essere considerata accettabile. D'altra parte per rappresentare in modo equilibrato i testi che gli studenti possono incontrare nella loro esperienza universitaria sono stati privilegiati (20 domande su 30) i testi di tipo informativo, provenienti

però da fonti linguistiche molto diverse: legislativa (“Prestito d’onore”), giornalistica, (“La Francia e le trentacinque ore”) e saggistica (“Introduzione al saggio di John Dewey”).

Le 10 domande rimanenti sono legate a una tipologia di testi meno frequenti nel corso delle letture di studio, ma non meno importanti, poiché permettono di ricavare informazioni dalla decodifica di una tabella (“Assunzioni previste nelle imprese”) o di grafici (“La rappresentazione nella professione”); si tratta di testi pragmatici o funzionali, anche definiti “non continui” (OECD 2004) rispetto a quelli composti esclusivamente da elementi linguistici verbali.

Anche l’articolazione degli item, basta sul modello della citata indagine IEA-SAL, ma che trova riscontro anche alle tipologie caratteristiche degli item di comprensione della lettura identificate anche in altre indagini internazionali (Campbell J. *et alii* 2001, OECD 2003), può essere considerata adeguata, considerando che la localizzazione di informazioni e le inferenze testuali costituiscono le competenze specifiche più frequentemente utilizzate nella lettura di testi complessi. Inoltre la relativa mancanza di domande di tipo lessicale è compensata dalla presenza della prova Cloze, in cui la componente lessicale è predominante.

Nella Tabella 4.7 sono riepilogate le diverse tipologie di quesiti per ogni brano che compone la prova TCL

Tabella 4.7 – Composizione del pacchetto TCL per tipologia dei quesiti

Tipologia Testo	Padronanza lessicale	Localizzazione informazioni	Individuazione idea principale	Inferenza	Tot.
Prestito d’onore	1	2		1	4
Assunzioni		4		3	7
Professione			1	2	3
35 ore		5			5
John Dewey	2	5	1	3	11
Totale	3	16	2	9	30

#### 4.5.2. Questionario Punti di vista sulla scelta universitaria

Per la prova relativa ai Punti di vista sulla scelta universitaria si rimanda al rapporto della collega Truffelli, ricordando qui che il tempo di somministrazione fissato, alla luce del collaudo svolto nella prova pilota, è di 45 minuti.

#### 4.5.3. *Prova di comprensione della lettura Cloze*

Per la prova di comprensione Cloze si è preferito in accordo con la collega Truffelli, operare una nuova selezione di testi, in considerazione del fatto che i testi utilizzati nello studio pilota erano tratti da manuali della scuola secondaria superiore, e non sempre rappresentavano il livello di difficoltà dei manuali universitari con cui si devono misurare gli studenti oggetto della rilevazione. Bisogna anche ricordare che la proposta di utilizzare le prove di Cloze per la rilevazione è emersa nel corso dell'incontro di dottorato di aprile 2003, quindi a ridosso del collaudo previsto per il TCL, per cui le prove utilizzate nella fase di collaudo sono state quelle considerate nel quadro della collaborazione esistente fra il Corso di laurea in Scienze dell'Educazione e della Formazione di Roma "La Sapienza" e un istituto superiore romano.

La scelta dei testi per la somministrazione principale si è svolta nel periodo giugno-luglio 2003, considerando le aree disciplinari affrontate dagli studenti nei primi anni di corso di Scienze dell'Educazione e della Formazione. Pertanto è stata avviata la ricerca di testi di Pedagogia, Antropologia, Sociologia, Filosofia, Psicologia e Storia. Considerando però l'andamento della prova pilota svolta a Roma, e in particolare il tempo necessario per completare un testo con circa 30 termini mancanti, si è convenuto sull'opportunità di limitare la prova a soli quattro testi, per un impegno complessivo di un'ora. Va anche considerato che la prova di comprensione Cloze è prevista dopo la compilazione del TCL e del questionario "Punti di vista". Dall'analisi preventiva dei testi si è ritenuto opportuno escludere quelli di Sociologia e Storia.

Il pacchetto della prova Cloze è quindi composto da quattro testi, per un totale di 125 "buchi" da riempire.

Nella Tabella 4.8 sono riepilogati i 4 testi che compongono la versione definitiva della prova di comprensione Cloze.



Tabella 4.8 – Composizione della prova Cloze

Area disciplinare	Argomento	Fonte	N° buchi
Antropologia	Metodo lavoro	M. Mead – L'adolescenza in Samoa	31
Psicologia	Madre-Bambino	AA.VV - Introduzione alla psicologia sociale	32
Pedagogia	Freire	M. Gadotti – Leggendo Paulo Freire	31
Filosofia	Locke	Abbagnano e Foriero – Fare filosofia vol.II	31

Come si può notare è stato recuperato il brano di Filosofia (tratto da un manuale di Storia della filosofia utilizzabile anche per la preparazione di esami di livello universitario) già utilizzato nella prova pilota di Roma.

Ognuno dei testi selezionati è lungo circa 220-250 parole, è stato lasciato inalterato il primo capoverso e dal secondo si è proceduto, secondo la classica tecnica Cloze collaudata nella prova pilota, ad eliminare una parola ogni cinque, per un totale di circa trenta parole sostituite da spazi. Il brano viene completato con un ulteriore capoverso per dare senso compiuto e rendere possibile la ricostruzione degli ultimi termini mancanti.

Durante il lavoro di predisposizione dei testi sono stati provati diversi schemi di bucatatura del testo, basati sempre sul passo una parola su cinque, ma differenziati secondo la parola da cui iniziare la bucatatura. In tre testi su quattro lo schema di buchi più equilibrato si è rivelato quello basato sulla quinta parola bucata della frase da cui ha inizio la compilazione, mentre per il testo di Pedagogia lo schema più convincente è risultato quello basato sull'eliminazione della prima parola della frase di partenza.

Per l'equilibrio fra i diversi tipi di termini da ricostruire è stata soprattutto considerata la loro varietà rispetto alla classificazione che prevede distinzioni fra i diversi tipi di parole (Kintsch, Van Dijk 1983).

Si parte da una prima distinzione fra parole piene (il cui significato prescinde dal contesto, in genere sostantivi e verbi) e parole vuote (che acquistano significato specifico nel contesto d'uso, dette anche parole grammaticali). Per le parole piene interviene poi una distinzione fra termini di base, ricostruibili anche attraverso indizi grammaticali evidenti quali le concordanze di genere o numero con altri termini, e termi-

ni appartenenti al registro medio-elevato, in cui la capacità di ricostruzione è legata soprattutto alla comprensione complessiva del testo. Una serie di termini però sfugge a questa semplice bipartizione, perché pur classificabili come termini pieni/vuoti, vengono ricostruiti attraverso un'inferenza congiuntiva, cioè un collegamento con altre parti del testo: il collegamento può riguardare i nessi sintattici fra le diverse parti del testo o puntuali ancoraggi a termini presenti nel brano e non bucati (come accade nelle ripetizioni di frasi simili). Questi termini si collocano quindi in posizione intermedia fra quelli strettamente grammaticali e quelli lessicali.

Può risultare utile rappresentare materialmente le cinque categorie di termini appena classificate, per cui nella Tabella 4.9 sono riportati gli esempi tratti dal primo testo della prova (Antropologia).

Tabella 4.9 – Esempi relativi ai diversi termini della prova Cloze

Abilità	Esempio (fra parentesi il termine mancante)
Conoscenza grammaticale di base	Il solo studio (della) famiglia francese
Collegamenti sintattici fra frasi	...l'antropologo sceglie popoli molto semplici, popoli primitivi, la (cui) società non abbia mai raggiunto la complessità della nostra
Collegamenti ai significati espressi nel testo	Nelle civiltà complicate come quelle europee, o come (quelle) più elevate dell'oriente
Lessicale di base	capire da quali forze (sono) animate
Lessicale avanzata	uno studioso esperto può rendersi conto in pochi mesi della (struttura) fondamentale di una società primitiva

La tabella che segue riepiloga, per i diversi testi usati per la prova Cloze, l'articolazione dei diversi tipi di termini bucati. L'insieme dei 125 termini risulta ben suddiviso fra le diverse categorie, con un sostanziale equilibrio fra le tre abilità principali e una soddisfacente articolazione interna delle abilità di collegamento e lessicale.

Tabella 4.10 – Composizione della prova Cloze per tipologia dei termini mancanti

Abilità	Grammaticale	Di collegamento		Lessicale		Tot.
	47	36		42		125
Testo	Conoscenza grammaticale di base	Collegamenti sintattici tra frasi	Collegamento ai significati espressi nel testo	Lessicale di base	Lessicale avanzata	
Antropologia	10	2	5	11	3	31
Psicologia	13	2	7	6	4	32
Pedagogia	14	3	8	2	4	31
Filosofia	11	6	2	8	4	31
Totale	47	9	27	27	15	125

Nella fase di allestimento dei materiali sono stati considerati con attenzione gli aspetti grafici della prova. In particolare gli spazi da riempire con le parole mancanti sono di dimensioni uniformi, se il termine che precede la parola mancante è troncato è stato lasciato l'apostrofo, sono stati lasciati anche i segni di punteggiatura eventualmente presenti prima o dopo ogni parola mancante.

#### 4.5.4. Questionario nuovi iscritti

Come già accennato in precedenza questo strumento ha ripreso, con piccoli adattamenti, il questionario già in uso nel Corso di laurea in Scienze dell'Educazione e della Formazione di Roma "La Sapienza" (Benvenuto 2003). Le aree indagate da questo strumento sono quelle relative alle caratteristiche socio anagrafiche del rispondente, al rapporto con la vita universitaria, al livello di istruzione e preparazione personale, ad alcune semplici informazioni sulla famiglia di origine (in particolare il livello di cultura familiare) e infine alle motivazioni che hanno condotto alla scelta del corso di studi.

Si è provveduto in particolare ad una nuova stesura delle istruzioni per la compilazione, puntando ad accrescere la motivazione dei rispondenti per ottenere una maggiore precisione circa le domande squisitamente individuali, considerato che la compilazione di questo strumento si svolge al termine di una impegnativa sessione di lavoro. Inoltre sono state aggiunte due domande riprese da quelle collaudate nella prova pilota del questionario "Punti di vista sulla scelta universitaria", in quanto meno aperte rispetto alle normali domande che compongono quest'ultimo questionario.

Il prospetto che segue riassume le aree di riferimento per le 21 domande del questionario

Tabella 4.11 – Struttura del Questionario nuovi iscritti

<b>Area</b>	<b>N. domanda</b>	<b>Argomento</b>
Dati anagrafici	1	Nome e cognome
	2	Sesso
	3	Data di nascita
	4	Luogo di nascita
	5	Nazionalità
Rapporto con l'università	6	Residenza durante gli studi
	7	Distanza dall'Università
	8	Tipo di iscrizione
	9	Anno di iscrizione
	18	Impegno lavorativo attuale
	20	Previsione di frequenza studi
Istruzione e preparazione personale	10	Scuola secondaria frequentata
	11	Voto di diploma
	15	Conoscenza lingue straniere
Cultura familiare	12	Titolo di studio genitori
	13	Esperienze universitarie fratelli/sorelle
	14	Professioni in famiglia legate al corso di studi
Scelta del corso di studi	16	Tipo orientamento ricevuto
	17	Partecipazione a incontri di orientamento
	19	Motivo di scelta del corso
	21	Prospettiva lavorativa

Per la compilazione del questionario è stato stimato un tempo di 15 minuti circa, assicurando comunque tutto il tempo necessario per il completamento di tutte le domande.

#### 4.6. La scheda di rilevazione del percorso accademico

La messa a punto di uno strumento che intende definire l'abilità di lettura all'ingresso di un ciclo di studi non può rinunciare ad un riscontro degli esiti dopo un periodo significativo in cui gli studenti hanno a-

vuto modo di mettere in campo le loro competenze. Risulta quindi opportuno prevedere una rilevazione sugli esiti da effettuare al termine delle sessioni di esame relative ai due semestri che compongono il I anno di corso.

Sicuramente i successi riportati dipendono da diverse variabili oltre quelle specifiche legate alla competenza nello studio, rappresentata nella prova di ingresso dall'abilità di lettura. Diverse di queste variabili sono relative ad aspetti di vita dello studente difficilmente riconducibili ad un modello di rilevazione (situazione familiare, affettiva, di salute), ma alcuni aspetti legati al successo negli studi risultano maggiormente connessi all'esperienza di rapporto con l'istituzione universitaria, per cui è possibile considerarli per cercare di stimare il loro effetto rispetto alla competenza di studio.

Un quadro riassuntivo dei fattori rilevabili che possono incidere nel successo degli studi è il seguente:

- Disponibilità di tempo: lo studio è attività esclusiva o si affianca a un lavoro?
- Facilità di frequenza/presenza: la sede di studio è facilmente/velocemente raggiungibile da casa?
- Modalità di svolgimento degli esami: il rendimento dipende dal tipo di esame (scritto/orale)?

Alcuni di questi aspetti sono stati già considerati nel questionario di ingresso: distanza casa/università, svolgimento di un'attività lavorativa, aspettativa di frequenza. In alcuni casi si rende necessario verificare le informazioni in ingresso, per esempio l'effettiva frequenza rispetto alle previsioni.

Nella parte introduttiva della scheda sono state considerate altre due informazioni utili: la modalità di iscrizione (normale o part-time) e la data di svolgimento del primo esame. Circa l'iscrizione bisogna considerare che la media di crediti richiesti per ogni semestre si riduce sensibilmente per gli iscritti part-time, perciò il successo ottenuto deve essere rapportato all'impegno richiesto e alla previsione di completamento del ciclo di studi: nell'ipotesi di completamento del corso in 4 anni e non in 3 il numero di crediti previsti per ogni anno scende da 60 a 45.

La segnalazione della data del primo esame permette di capire se ci sono stati problemi di avvio, possibili e frequenti per i nuovi iscritti, in particolare se l'iscrizione è avvenuta in ritardo, con la possibile perdita

della parte didattica, e delle relative sessioni di esame, per il primo semestre.

Naturalmente il principale oggetto della raccolta di informazioni da effettuare è costituito dal numero degli esami sostenuti, che potrebbe essere rilevato anche in modo automatico tramite il controllo dei piani di studio. Ma le informazioni contenute nel piano di studio da una parte non risultano ancora tempestivamente aggiornate, considerando i problemi strutturali che riguardano la gestione amministrativa del corso di laurea, dall'altra non risultano comunque sufficienti per la stima di alcune delle variabili indicate sopra, per cui si è ritenuto necessario costruire un apposito strumento di rilevazione, sulla scorta anche di quanto è stato fatto negli anni precedenti per monitorare l'andamento del corso di laurea (cfr. Benvenuto 2003).

Considerati gli scopi del presente lavoro di ricerca è stata allestita una scheda informativa che intende rilevare gli esami sostenuti con le seguenti informazioni accessorie: voto riportato, numero dei crediti, modalità di svolgimento dell'esame, frequenza del corso relativo. Le prime due informazioni possono essere riscontrate con la rilevazione dei piani di studio, gli ultimi due dati sono necessariamente forniti dagli studenti, anche se la procedura di informatizzazione delle verbalizzazioni di esame, in corso di applicazione, potrebbe fornire in futuro anche dati affidabili circa questi aspetti.

Circa i dettagli delle categorie di risposta la compilazione degli studenti è agevolata dalle le istruzioni relative, notando che per la frequenza rispetto alla scelta secca Sì/No è stata considerata una possibilità di risposta intermedia che definisce precisamente la frequenza parziale, laddove siano state seguite tra il 40 e il 605 delle ore previste per il corso. In una tabella specifica sono state considerate le idoneità, che non comportano un voto e possono essere superate con la presentazione di specifiche attestazioni di competenza nella materia oggetto di esame. Da notare che si propone l'uso di una tabella di codifica degli esami, per facilitare la compilazione da parte degli studenti, ma soprattutto l'inserimento dei dati.

Per completare il quadro informativo sui successi nel primo anno di corso si è ritenuto opportuno raccogliere informazioni anche sui tentativi che non hanno avuto successo, con una tabella che rileva gli esami non completati, con il relativo credito e se l'insuccesso si è verificato nella prova scritta o orale. Anche un tentativo non riuscito può risultare utile in termini di esperienza: l'esame fallito può essere sostenuto di nuovo con maggiori possibilità di superamento (e sarà possibile verificare se

nelle sessioni disponibili ciò è stato fatto), ed è possibile disporre di un indicatore accessorio utile che affianca il numero dei crediti ottenuti, cioè quello dei crediti affrontati.

La Tabella 4.12 riepiloga il contenuto della scheda, articolato per aree di contenuto.

Tabella 4.12 – Struttura della scheda di rilevazione del percorso accademico

Area	Argomento	
Dati anagrafici	Nome e cognome	
Tipo di iscrizione	Normale/part time	
Primo esame	Data di svolgimento	
Esami sostenuti	Esame	Intestazione esame
	Voto	In trentesimi (31 = 30 e lode)
	Crediti	Numero
	Modalità	Scritto, Scritto/orale, Orale
	Frequenza	Regolare, parziale, non frequentante
Idoneità sostenute	Idoneità	Intestazione idoneità
	Crediti	Numero
	Modalità	Attestazione, Prova
Esami non completati	Esame	Intestazione esame
	Crediti	Numero
	Modalità	Scritto, Orale

Le informazioni raccolte con questo strumento risulteranno determinanti per la verifica dell'ipotesi di ricerca.





## Capitolo quinto

### Risultati della prova di ingresso

#### 5.1. La somministrazione della prova di ingresso

Tutte le procedure da attivare per la realizzazione della somministrazione principale sono state definite in stretta collaborazione fra le due Università i cui studenti nuovi iscritti sono stati oggetto di indagine, allo scopo di garantire condizioni di somministrazione equivalenti nelle due realtà, premessa necessaria per il riscontro di validità degli strumenti utilizzati.

Di seguito verrà illustrata la situazione relativa all'Università di Roma "La Sapienza", rimandando al rapporto di ricerca della collega Truffelli per la somministrazione effettuata a Bologna.

In seguito alla revisione dei materiali utilizzati nella somministrazione pilota è stata in primo luogo operata una omogeneizzazione grafica degli strumenti, curando formati, caratteri e impaginazione che garantissero una buona leggibilità per gli studenti, ma anche facilità di riproduzione fotostatica rispetto agli originali. Sono state anche riviste le istruzioni delle singole prove, in particolare per renderle più omogenee sugli scopi dell'impegno richiesto agli studenti e più chiare circa le modalità di risposta e di eventuale correzione.

Si è convenuto di seguire un ordine di somministrazione che agevolasse lo sforzo degli studenti, partendo quindi con la prova considerata più difficile (comprensione lettura a scelta multipla TCL) seguita dal questionario a domande aperte ("Punti di vista sulla scelta universitaria"), sicuramente impegnativo, ma considerato meno stressante dal punto di vista dell'ansia di dover fornire per forza una risposta giusta. Subito dopo si è deciso di lasciare un tempo di intervallo di circa 40-45 minuti, permettendo un adeguato riposo prima di affrontare la seconda parte della somministrazione, con la prova di comprensione Cloze e la compilazione del questionario.

Nel complesso quindi gli studenti sono stati impegnati, presentazione iniziale e pause comprese, per circa quattro ore. La possibilità di racco-

gliere le iscrizioni degli studenti alla prova ha permesso di pre-compilare i moduli ottici da utilizzare nella compilazione della prima prova (TCL), per cui, almeno a Roma, la somministrazione è stata programmata fra le 9 e le 13 del giorno 29 settembre 2003, con una sessione pomeridiana fra le 14 e le 18 per gli studenti impossibilitati a partecipare al mattino.

La riproduzione dei materiali è stata realizzata sotto l'attenta supervisione dei responsabili della somministrazione, senza coinvolgere altre persone, in modo da garantire l'assoluta riservatezza dei materiali.

Grande importanza ha assunto la pubblicizzazione della prova fra i nuovi iscritti. In tutti gli incontri realizzati a cavallo dell'estate 2003 per promuovere l'iscrizione al corso di laurea e informare i potenziali nuovi iscritti, è stata anche presentata la prova, chiarendo il carattere esclusivamente orientativo della stessa<sup>68</sup>, ma richiedendo un impegno adeguato per una efficace raccolta di informazioni sugli studenti che vede impegnato il corso di laurea fin dalla sua costituzione (Benvenuto 2003), allo scopo di garantire una migliore offerta didattica.

Una volta fissata la data e il luogo di somministrazione è stata predisposta un'informazione capillare, con avvisi pubblici, anche sui siti internet del Dipartimento e della Facoltà cui afferisce il corso di laurea. Infine sono state attivamente coinvolte le diverse segreterie interessate (di facoltà, del servizio di orientamento e del servizio civile di supporto al Dipartimento).

Per l'organizzazione logistica delle somministrazioni sono state scelte quattro aule di capienza adeguata, per fare in modo che ogni studente fosse sufficientemente lontano dagli altri, prevenendo in tal modo possibili tentativi di copiatura: rispetto al consueto svolgimento delle lezioni ogni studente era separato dai vicini, sia laterali sia longitudinali, da un posto vuoto, anche per favorire una compilazione comoda degli strumenti. Considerando il contesto simulato della prova di accesso è stata inoltre prevista una attenta vigilanza durante lo svolgimento della prova. Per ogni aula sono stati previsti 40-50 studenti, con 3-4 persone per la vigilanza.

Sottolineiamo però ancora una volta il rinforzo della motivazione svolto prima della prova, illustrando con chiarezza, fin dalla comunica-

<sup>68</sup> In particolare insistendo sulla distinzione fra ingresso ed accesso (Andreani Dentici, Amoretti 2000).

zione della data di svolgimento negli incontri di orientamento svolti presso l'università, gli scopi orientativi della rilevazione e l'assoluta non utilizzabilità dei risultati per la valutazione accademica, sottolineando la necessità di raccogliere informazioni precise circa l'effettiva abilità di lettura prima dell'inizio degli studi universitari, quindi chiedendo ai partecipanti di impegnarsi al meglio delle loro capacità e in maniera leale.

La somministrazione principale si è regolarmente svolta il 29 settembre, con 170 studenti partecipanti (18 nella sessione pomeridiana) a fronte dei 240 nuovi iscritti comunicati dalla segreteria il giorno precedente la prova. Considerando che le iscrizioni al corso di laurea erano però possibili fino al successivo 5 novembre, si è ritenuto opportuno riproporre alcune sessioni di recupero, per far partecipare anche i 70 studenti assenti nella prima somministrazione e gli iscritti successivi.

Sono state realizzate tre sessioni di recupero a Roma (25 ottobre, 8 novembre, 15 novembre), aggiornando periodicamente gli elenchi dei nuovi iscritti secondo le indicazioni fornite dalla segreteria amministrativa di ateneo. La somministrazione del 15 novembre si è resa necessaria perché l'Università ha concesso una deroga sui tempi di iscrizione fino a dicembre, ma si è scelto comunque di non superare la metà del mese di novembre, considerando che gli iscritti successivi a tale data di fatto avevano perso buona parte delle lezioni del primo semestre, per cui il confronto con i risultati del primo anno di corso poteva risultare molto parziale.<sup>69</sup>

Prima dello svolgimento delle tre sessioni di recupero è stata rinforzata la pubblicizzazione della prova, sia per quanto riguarda le motivazioni (buona parte degli iscritti successivi non aveva partecipato ad alcun incontro di orientamento), sia per illustrarne gli aspetti pratici di svolgimento. Considerando il parallelo inizio dell'attività didattica, la partecipazione è stata sollecitata anche nei primi incontri dei corsi previsti per il primo semestre di studio.

Nel complesso sono stati raccolti i dati relativi a 419 studenti (vedi Tabella 5.1 per il dettaglio delle diverse sessioni di somministrazione).

<sup>69</sup> Alla data di scadenza prevista per le iscrizioni dell'A.A. 2003-04 risultavano 448 nuovi iscritti, quindi è stato raggiunto circa il 94% della popolazione di riferimento. In virtù della successiva proroga delle iscrizioni al mese di dicembre 2003 il numero definitivo di nuovi iscritti per l'A.A. 2003-04 è stato di 516 studenti.

Tabella 5.1 – Dati riassuntivi della somministrazione (Roma “La Sapienza”)

Sessioni	Partecipanti
29 settembre 2003	170
25 ottobre 2003	79
8 novembre 2003	95
15 novembre 2003	75
Totale	419

Pur con qualche difficoltà organizzativa, legata alla minore disponibilità di aule libere nel periodo di sovrapposizione con lo svolgimento dei corsi, è stato possibile assicurare anche alle tre sessioni di recupero le stesse condizioni logistiche della prima, ricalcando fedelmente anche la scansione oraria delle diverse prove.

Da sottolineare che il comportamento degli studenti nel corso di tutte le sessioni di somministrazione, è stato esemplare, permettendo di svolgere la prova nelle condizioni ideali. Gli spazi predisposti si sono rivelati perfettamente adeguati, anche per lo svolgimento della pausa prevista a metà della sessione, per cui non si sono verificati abbandoni nella seconda parte di tutte le quattro sessioni. Durante le diverse fasi di compilazione dei materiali non si sono verificati problemi particolari, anche perché le istruzioni predisposte, sia per la presentazione complessiva della prova che per i singoli strumenti, si sono rivelate chiare ed esaustive. Anche l'utilizzazione del modello ottico di risposta non ha comportato problemi<sup>70</sup>.

L'inserimento dei dati è stato anticipato per quanto riguarda le prove di comprensione della lettura TCL e il questionario informativo. Successivamente, in stretta collaborazione con la collega Truffelli, è stato avviato l'inserimento della prova Cloze.<sup>71</sup>

<sup>70</sup> La vigilanza durante le sessioni è stata assicurata dai docenti del corso di laurea e dal personale del Servizio civile in forza al Dipartimento di ricerche storico-filosofiche e pedagogiche. A loro va il dovuto ringraziamento per aver contribuito allo svolgimento di questa delicata fase dell'indagine.

<sup>71</sup> Ricordiamo che il questionario aperto “Punti di vista sulla scelta universitaria” è stato elaborato da Truffelli per il suo lavoro di dottorato.

## 5.2. Verifica del campione partecipante alla prova di ingresso

Il questionario informativo per gli studenti che hanno partecipato alla prova, già presentato in 4.5.4, è stato elaborato in relazione ai dati delle matricole dei precedenti anni accademici, in collaborazione con il gruppo di studenti borsisti del corso di laurea, coordinato da Guido Benvenuto<sup>72</sup>. L'analisi delle risposte fornite dai 419 partecipanti alla prova permette in primo luogo di capire se si tratta di un campione effettivamente rappresentativo della popolazione dei nuovi iscritti per l'A.A. 2003-04.

La prima distinzione importante da fare è quella fra le matricole e i nuovi iscritti provenienti da precedenti esperienze universitarie.

Tabella 5.2 – Provenienza dei nuovi iscritti (419 soggetti)

	Frequenza	%	% 2001-02
Matricole	332	79,2	56,7
Già iscritti	87	20,8	43,3
Totale	419	100,0	100,0

Nel confronto bisogna considerare che il 2001-02 è stato il primo anno di attivazione del corso di laurea, per cui è giustificabile un primo impatto importante di trasferimenti da altre università o corsi di laurea.

Tra gli 87 studenti già iscritti, 59 provengono da altri corsi di laurea (di cui 9 anche da altre università), 10 risultano già laureati e 9 riattivano gli studi dopo un'esperienza di abbandono (9 studenti non specificano l'esperienza precedente).

### 5.2.1. Caratteristiche anagrafiche

Considerando che la prova di accesso deve essere sostenuta da tutti i nuovi iscritti, trasferimenti compresi, risulta opportuno analizzare l'intero gruppo dei partecipanti, distinguendo, laddove necessario, le diversità esistenti fra matricole e re-isritti. Questo già si verifica per il genere.

<sup>72</sup> Cfr. Benvenuto (2003), da cui sono stati ricavati i dati per i confronti relativi al campione raggiunto.

Tabella 5.3 – Genere dei nuovi iscritti (419 soggetti)

		Provenienza		Totale	
		Matricole	Già iscritti	Frequenza	%
Sesso	Maschi	17	20	37	8,8
	Femmine	315	67	382	91,2
Totale		332	87	419	100,0

La preponderanza femminile è evidente, ma notiamo che le non matricole maschi superano addirittura il numero delle matricole, risultando una importante minoranza dei già iscritti. Il dato complessivo conferma l'andamento rispetto agli anni precedenti, con un sostanziale 90% di presenza femminile.

Un altro elemento interessante riguarda l'età dei nuovi iscritti. Anche in questo caso risulta opportuno distinguere le matricole dai già iscritti. Si riporta anche l'età media distinta per i due gruppi

Tabella 5.4 – Età dei nuovi iscritti (419 soggetti)

		Provenienza		Totale
		Matricole	Già iscritti	
Età	18-20	65,7%	11,4%	54,4%
	21-23	16,9%	31,9%	19,8%
	24-30	7,2%	23,9%	10,7%
	>30	10,2%	32,8%	15,1%
Totale		100,0%	100,0%	100,0%
Età media		22,1	28,8	23,6

### 5.2.2. Esperienze di studio

Per quanto riguarda le matricole il confronto con i dati del monitoraggio del corso di laurea indicano una crescita della fascia di età proveniente direttamente dall'Esame di Stato (18-20), in precedenza al 45% circa, con una parallela discesa degli ultratrentenni. Anche questa differenza può essere spiegata con la progressiva normalizzazione dopo la novità dei primi anni del corso di laurea. Nel complesso si può notare che circa un quarto dei nuovi iscritti affronta l'esperienza di studio a una certa distanza di tempo (almeno 5 anni) dagli studi secondari superiori.

A questo proposito consideriamo, nella Tabella 5.5, il tipo di diploma conseguito. Sotto i valori percentuali sono indicati i valori assoluti.

Tabella 5.5 – Tipo di diploma dei nuovi iscritti<sup>73</sup> (419 soggetti)

		Provenienza				Totale	
		Matricole		Già iscritti			
Licei	Classico		12,3%		19,5%		13,8%
	Scientifico		13,6%		25,3%		16,0%
	Educativo	62,6%	28,9%	71,2%	19,5%	64,5%	27,0%
	Linguistico	(208)	4,2%	(62)	6,9%	(270)	4,8%
	Artistico		2,4%				1,9%
	Generico		1,2%				1,0%
Tecnici	Generico	28,0%	18,4%	19,6%	17,2%	26,2%	18,1%
	Educativo	(93)	9,6%	(17)	2,3%	(110)	8,1%
Professionali	Generico	8,7%	7,2%	5,7%	3,4%	8,1%	6,4%
	Educativo	(29)	1,5%	(5)	2,3%	(34)	1,7%
Omissioni		(2)	,6%	(3)	3,4%	(5)	1,2%
Totali di gruppo		(332)	100,0%	(87)	100,0%	(419)	100,0%

Fra parentesi i valori assoluti.

Rispetto al passato risultano in calo le matricole provenienti da Licei e Istituti Professionali, con un consistente aumento interno di diplomati degli Istituti Tecnici (erano circa il 16%). Ma se consideriamo trasversalmente i diplomi specifici del settore educativo, sempre per le matricole, arriviamo al 40%, confermando il dato del monitoraggio precedente. Il calo dei Licei è quindi dovuto a una discesa degli indirizzi non di tipo specificamente educativo.

Un elemento molto importante per capire le caratteristiche del gruppo di nuovi iscritti è il voto riportato nell'Esame di Stato. Come potevamo già intuire dalla classificazione per classi di età, non sono pochi i

<sup>73</sup> L'indicazione "Generico" per i Licei si riferisce alla non specificazione dello studente del tipo di liceo. Invece per Tecnici e Professionali "Generico" si riferisce a tutti gli indirizzi non tipo educativo.

L'indicazione "Educativo" per i Licei si riferisce agli indirizzi magistrale e socio-psico-pedagogico. Invece per Tecnici e Professionali a tutti gli indirizzi di tipo educativo (Servizi sociali, Dirigente di comunità, Comunità infantili etc.)

partecipanti<sup>74</sup> che hanno sostenuto l'Esame di Stato prima della riforma del 1999, che ha disegnato una nuova scala di punteggi, passata dalla gamma 36-60 a quella 60-100. Pertanto prima di analizzare i dati si è resa necessaria una trasformazione matematica per uniformare tutti i punteggi dell'Esame di Stato sulla scala 60-100.

In questo caso è utile articolare la Tabella secondo il tipo di diploma conseguito

Tabella 5.6 – Voto dell'Esame di Stato per tipo di diploma (419 soggetti)

		Tipo di diploma			Totale
		Licei	Tecnici	Professionali	
Fasce voto Esame di Stato nor- malizzato	60	10,4%	7,3%	2,9%	8,8%
	61-69	18,9%	29,1%	5,8%	20,5%
	70-79	28,2%	32,6%	50,0%	31,0%
	80-89	20,3%	16,2%	20,5%	19,1%
	90-99	8,9%	7,2%	8,8%	8,3%
	100	7,0%	2,7%	2,9%	5,5%
Omissioni		3,0%	1,8%		2,6%
Totale		100,0%	100,0%	100,0%	100,0%
Statistiche riassuntive					
Media voto		76,8 (0,8)	74,4 (1,3)	78,6 (1,7)	76,4 (0,6)
Dev std. voto		12,3	10,9	9,7	11,8
C.d.V. Voto		16,0	14,7	12,3	15,4
Curtosi		-0,88 (0,30)	-0,39 (0,46)	-0,53 (0,78)	-0,79 (0,24)
Asimmetria		0,41 (0,15)	0,74 (0,23)	0,25 (0,40)	0,47 (0,12)

Fra parentesi l'errore standard relativo.

Come si può notare esistono differenze nella distribuzione dei voti secondo il tipo di diploma di provenienza. Il dato di asimmetria risulta sempre positivo, quindi con una coda destra più pronunciata, soprattutto per licei e tecnici, per i quali si può parlare di marcata asimmetria positiva<sup>75</sup>. Il dato di curtosi risulta sempre negativo, quindi abbiamo a che fare con distribuzioni meno raggruppate e code tagliate; in particolare il dato dei Licei (curtosi superiore al doppio dell'errore standard) indica

<sup>74</sup> Nel complesso sono 75 (17,9%) i partecipanti alla prova che hanno indicato il voto in sessantesimi.

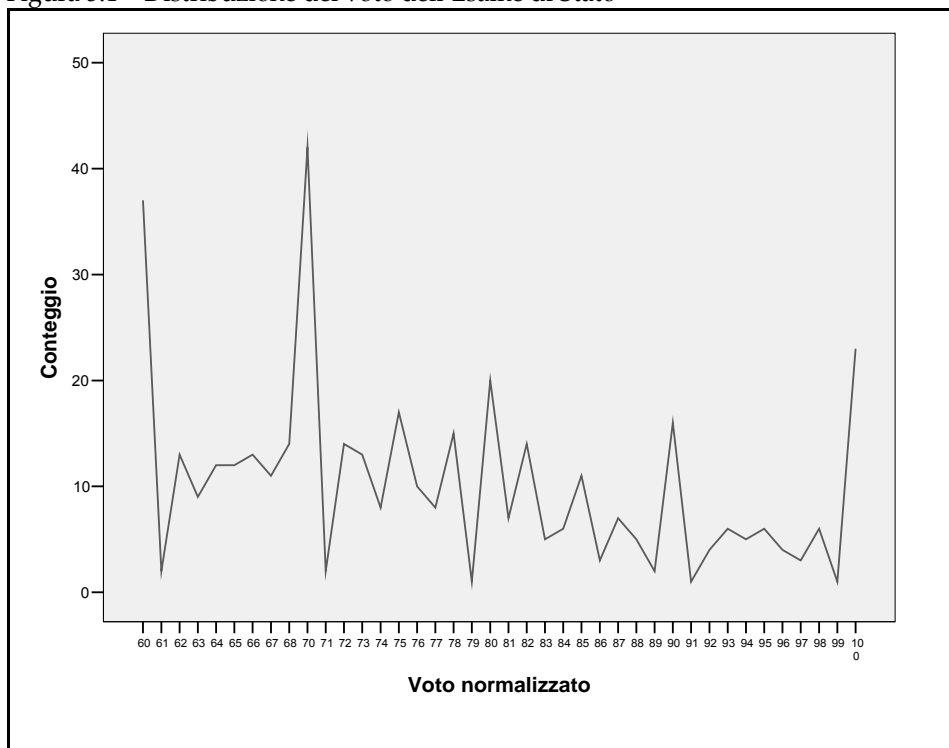
<sup>75</sup> Si può parlare di asimmetria non normale quando il rapporto fra questa e l'errore standard relativo è superiore a +2 o inferiore a -2.



una scarsa normalità della distribuzione, influenzando conseguentemente il dato relativo all'intera distribuzione.

Bisogna però considerare che la scala dei punteggi dell'Esame di Stato è utilizzata dalle commissioni in modo anomalo, con una maggiore frequenza dei punteggi considerati come una soglia psicologica (in particolare i multipli di 10). Questo è evidente nel grafico 5.1, dove si nota una serie di picchi e di abissi complementari.

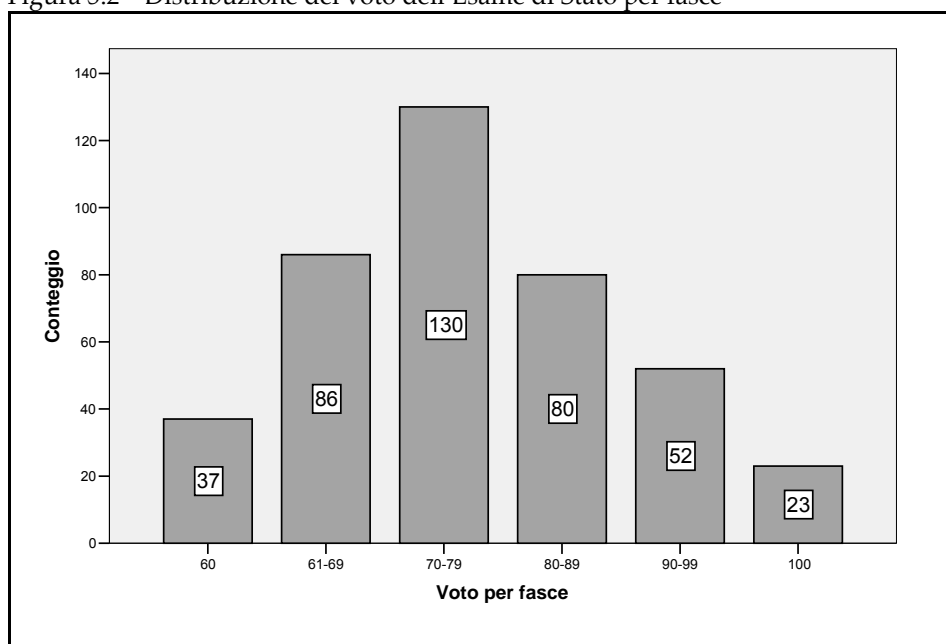
Figura 5.1 – Distribuzione del voto dell'Esame di Stato



Se però raggruppiamo i voti in fasce di punteggio (vedi Tabella 5.6) considerate su un *continuum*, otteniamo una distribuzione con minori problemi di normalità. Per esempio considerando le sei fasce dell'intera popolazione il valore di curtosi scende a  $-0,54$  ( $0,24$ ), appena sopra la soglia di normalità dell'appiattimento della curva. Anche il valore di asimmetria scende a  $0,26$  ( $0,12$ ), avvicinandosi alla simmetria normale di distribuzione dei punteggi.

La relativa normalizzazione della distribuzione appare evidente nel successivo grafico in cui sono raggruppate le diverse fasce del voto dell'Esame di Stato.

Figura 5.2 – Distribuzione del voto dell'Esame di Stato per fasce



La distribuzione risulta sostanzialmente imperniata sulla fascia 70-79, con discese progressive verso i punteggi minimo e massimo della scala dell'Esame di Stato.

### 5.2.3. *Contesto familiare*

Passiamo a un altro gruppo di domande del questionario nuovi iscritti, relativo al contesto culturale familiare, anche in rapporto alla scelta di intraprendere lo specifico corso universitario.

Un elemento considerato importante per il successo negli studi è il livello di cultura familiare, rilevato nel questionario attraverso le domande sul titolo di studio dei genitori, riepilogati nella Tabella 5.7.

Tabella 5.7 – Titolo di studio dei genitori (risposte valide su 419 soggetti)

		Titolo di studio padre		Titolo di studio madre	
		Frequenza	%	Frequenza	%
Titolo di studio	Nessun titolo	3	,8	1	,2
	Lic. elementare	52	13,1	53	13,2
	Lic. media	134	33,7	136	33,8
	Diploma secondario	168	42,2	176	43,8
	Diploma universitario	3	,8	8	2,0
	Laurea	26	6,5	25	6,2
	Titolo post laurea	12	3,0	3	,7
Totali di gruppo		398	100,0	402	100,0

Sono state escluse le omissioni di risposta e le risposte non utilizzabili.

Rispetto alla popolazione precedente degli iscritti al corso di laurea cambia il titolo di studio più rappresentato, con il diploma secondario che sorpassa, per entrambi i genitori, la licenza media, ma parallelamente diminuiscono anche i genitori con titolo post secondario, e se si distinguono le matricole rispetto ai già iscritti si può notare l'accentuazione di questo fenomeno.

Tabella 5.8 – Titolo di studio dei genitori per tipo di iscrizione (risposte valide su 419 soggetti)

		Titolo di studio padre (398)		Titolo di studio madre (402)	
		Matricole	Già iscritti	Matricole	Già iscritti
Titolo di studio	Nessun titolo	0,6%	1,3%		1,2%
	Lic. elementare	12,9%	13,9%	12,8%	14,6%
	Lic. media	37,3%	19,0%	37,2%	20,7%
	Diploma secondario	41,7%	44,3%	42,2%	50,0%
	Diploma universitario	0,6%	1,3%	1,6%	3,7%
	Laurea	4,4%	15,2%	5,6%	8,5%
	Titolo post laurea	2,5%	5,1%	0,6%	1,2%
Totali di gruppo		100,0%	100,0%	100,0%	100,0%

Sono state escluse le omissioni di risposta e le risposte non utilizzabili. Fra parentesi il numero dei soggetti validi.

Le differenze più marcate riguardano i genitori con licenza elementare, praticamente il doppio per le matricole, e quelli con titolo post secondario, nettamente più presenti fra i già iscritti. Ricordiamo che le percentuali di genitori delle matricole con titolo superiore al diploma secondario erano in precedenza del 13,7 (padre) e 11,6 (madre), per cui l'aumento dei genitori diplomati (secondari) è avvenuto proprio per la diminuzione dei titoli superiori, non per innalzamento dei più bassi, che risultano sostanzialmente stabili. Da notare che i titoli di studio più elevati per i genitori dei già iscritti, la cui età media è plausibilmente più alta, risulta in controtendenza con l'accertato elevamento del titolo di studio fra le generazioni più giovani, ma risulta un indubbio fattore di motivazione per i nuovi iscritti che riprendono l'esperienza di studio dopo un'esperienza poco positiva o addirittura di abbandono.

Per operare un raffronto con la situazione nazionale si possono considerare i dati raccolti dall'OCSE<sup>76</sup>. Nel 2003 la fascia di età 45-54, presumibilmente quella in cui si colloca la maggior parte dei genitori dei nuovi iscritti, risultava solo per il 39% in possesso di almeno un titolo di scuola secondaria superiore, comprendendo quindi anche un 10% di laureati (o meglio secondo la classificazione OCSE, dotati di titolo terziario).

Come si vede il profilo del gruppo di nuovi iscritti al Coro di laurea è comunque superiore, per titolo di studio, da qualsiasi punto di vista al dato nazionale. Se però consideriamo la fascia di adulti più giovani (35-44 anni), in cui potrebbe trovarsi una piccola parte dei genitori dei nuovi iscritti, il vantaggio si riduce, poiché i dati OCSE ci dicono che in questa fascia di età il diploma secondario è stato ottenuto dal 50% degli italiani, che corrisponde alle percentuali rilevate per le matricole, ma considerando che in Italia è in atto un progressivo innalzamento del titolo di studio posseduto dalla popolazione, è ragionevole ipotizzare che il dato positivo di questa fascia di età sia imputabile soprattutto agli under 40, che difficilmente sono genitori di matricole universitarie. In definitiva si può sostenere che la cultura media familiare dei nuovi iscritti, rilevata attraverso il titolo di studio raggiunto dai genitori, sia superiore a quella nazionale.

<sup>76</sup> Ci riferiamo in particolare all'ultima edizione di *Education at a Glance*, che presenta proprio gli indicatori dell'istruzione dell'anno 2003. In particolare l'indicatore A1 per il livello di completamento degli studi della popolazione adulta (OECD 2005).

Un altro fattore familiare che può influenzare la scelta degli studi universitari è il modello rappresentato da fratelli o sorelle che hanno intrapreso o completato lo stesso tipo di studi. Come si può osservare nella Tabella 5.9 per circa metà degli studenti si tratta della prima esperienza universitaria di famiglia, e meno del 10% ha fratelli o sorelle che hanno già completato gli studi universitari.

Tabella 5.9 – Esperienza universitaria di fratelli/sorelle (risposte valide su 419 soggetti)

		Matricole e già iscritti	
		Frequenza	%
Fratelli e/o sorelle	Laureati	33	8,4
	Frequentanti	124	31,6
	Non frequentanti	194	49,4
	Figlio unico	42	10,7
Totale		393	100,0

Sono state escluse le omissioni di risposta e le risposte non utilizzabili.

Resta da aggiungere che la presenza di modelli familiari o parentali che possono aver influenzato la scelta dell'indirizzo universitario è stata dichiarata da 92 nuovi iscritti (22% del totale). Nel complesso quindi gli stimoli derivati dall'ambiente familiare circa la scelta di intraprendere studi universitari, scegliendo lo specifico indirizzo educativo, risultano minoritari.

#### 5.2.4. Rapporto con l'università

Alcune domande del questionario indagavano su alcuni elementi anagrafici da considerare in rapporto con le ipotesi formulate dagli studenti circa il rapporto con la struttura universitaria.

Un elemento anagrafico che può incidere sulla scelta e sullo svolgimento degli studi universitari è la facilità con cui è possibile raggiungere la sede. In primo luogo è possibile distinguere gli studenti originari di Roma rispetto a coloro che provengono da altre località.

Tabella 5.10 – Luogo di nascita (419 soggetti)

		Matricole e già iscritti	
		Frequenza	%
Luogo di nascita	Roma	222	53,0
	Provincia Roma	36	8,6
	Lazio	51	12,2
	Italia	101	24,1
	Eestero	9	2,1
Totale		419	100,0

Anche se la maggioranza assoluta studia nella sua città di origine, non sono pochi gli studenti provenienti da fuori regione, circa un quarto. Per quanto riguarda i 9 studenti nati all'estero bisogna aggiungere che solo 3 di loro sono di nazionalità non italiana, dato che risulta inferiore anche alla già bassa percentuale nazionale di studenti stranieri iscritti nel nostro sistema universitario<sup>77</sup>.

Ma il dato più indicativo circa la facilità di rapporto con l'università è quello relativo all'effettiva residenza durante l'anno accademico, con un terzo esatto dei nuovi iscritti che dichiara di abitare fuori Roma e conseguentemente si prepara ad affrontare un anno accademico da non frequentante o da pendolare, come è possibile evincere anche dall'analisi del tempo necessario per raggiungere l'università; la netta maggioranza dei non residenti impiega un tempo compreso fra un'ora e due ore per arrivare in facoltà, e circa un quinto di loro deve affrontare una vera e propria trasferta. Di contro circa il 60% dei residenti a Roma raggiunge l'università in meno di un'ora.

<sup>77</sup> Per il 2003 la percentuale di studenti stranieri presenti nelle università italiane era del 1,9% (Cfr. OECD 2005).

Tabella 5.11 – Tempo necessario per raggiungere l'università (risposte valide su 419 soggetti)

		Residenti a Roma		Residenti fuori Roma	
		Frequenza	%	Frequenza	%
Tempo in minuti	<29	76	28,1	-	-
	30-59	115	42,6	15	10,8
	60-89	61	22,6	48	34,5
	90-119	14	5,2	47	33,8
	>120	4	1,5	29	20,9
Totali di gruppo		270	100,0	139	100,0

Sono state escluse le omissioni di risposta e le risposte non utilizzabili.

È interessante incrociare l'informazione sul domicilio con la previsione, richiesta nel questionario, circa la frequenza delle lezioni. Il risultato è interessante perché la percentuale di non residenti intenzionata a frequentare in modo regolare le lezioni non è molto lontana da quella dei residenti, i quali peraltro si dividono quasi equamente tra l'intenzione di frequenza regolare e quella saltuaria.

Tabella 5.12 – Previsione di frequenza delle lezioni (risposte valide su 419 soggetti)

		Residenti a Roma		Residenti fuori Roma	
		Frequenza	%	Frequenza	%
Frequenza delle lezioni	Regolare	128	46,9	54	38,6
	Parziale	130	47,6	72	51,4
	Non frequentante	15	5,5	14	10,0
Totali di gruppo		273	100,0	140	100,0

Sono state escluse le omissioni di risposta e le risposte non utilizzabili.

In effetti la distanza dall'università sembra essere un fattore determinante per la frequenza delle lezioni, e considerando l'ampiezza del territorio romano la distanza temporale finisce per prescindere dalla residenza.

Tabella 5.13 – Frequenza secondo la distanza dall'università (risposte valide su 419 soggetti)

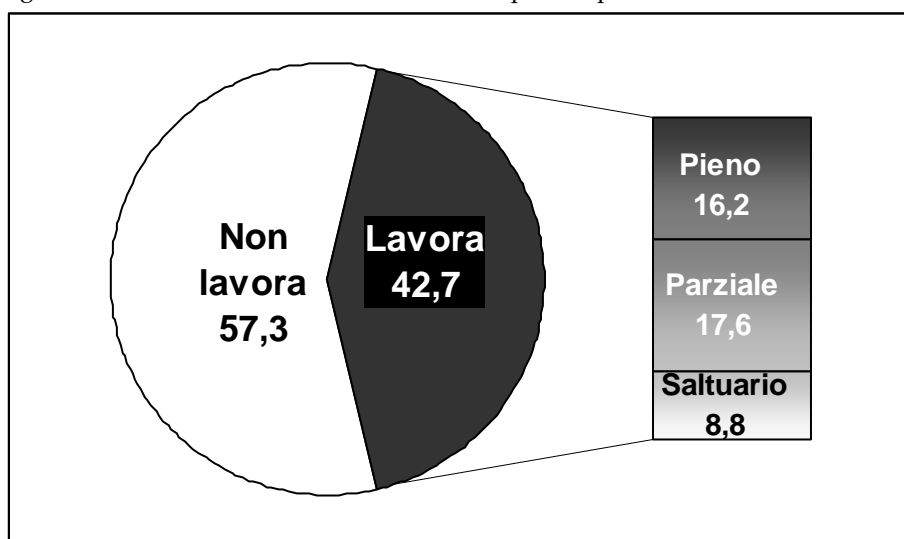
		Frequenza regolare		Frequenza saltuaria		Non frequentante		Totale	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%
Tempo in minuti	<29	40	52,6	32	42,1	4	5,3	76	100,0
	30-59	62	47,3	63	48,1	6	4,6	131	100,0
	60-89	46	41,8	59	53,6	5	4,5	110	100,0
	90-119	24	38,7	35	56,5	3	4,8	62	100,0
	>120	10	29,4	13	38,2	11	32,4	34	100,0
Totali di gruppo		182	44,1	202	48,9	29	7,0	413	100,0

Sono state escluse le omissioni di risposta e le risposte non utilizzabili. Percentuali calcolate sulle righe.

La discesa delle percentuali di frequenza segue con regolarità l'aumento dei tempi necessari per raggiungere la sede universitaria.

Bisogna però anche considerare l'esistenza di un altro fattore che può incidere sulla frequenza regolare delle lezioni, l'eventuale impegno lavorativo. Sono 179, ben il 42,7%, i nuovi iscritti che dichiarano di svolgere una qualche attività lavorativa, e di questi oltre un terzo lavora a tempo pieno.

Figura 5.3 – Distribuzione studenti lavoratori per tempo lavorativo





Risulta utile anche riepilogare, Tabella 5.14, la distribuzione specifica degli studenti lavoratori distinti per tipo di iscrizione, considerando che tra le matricole il 36,4% svolge un'attività lavorativa, mentre fra gli studenti già iscritti i lavoratori sono nel complesso ben il 66,7%, con una chiara prevalenza di quelli a tempo pieno.

Tabella 5.14 – Tempo lavorativo degli studenti lavoratori (179 soggetti)

		Matricole		Già iscritti	
		Frequenza	%	Frequenza	%
Tempo di lavoro	Tempo pieno	38	31,4	30	51,7
	Tempo parziale	50	41,3	24	41,4
	Lavoro saltuario	33	27,3	4	6,9
Totali di gruppo		121	100,0	58	100,0

Naturalmente l'impegno lavorativo rappresenta un forte vincolo per l'organizzazione degli studi universitari, e segnatamente per la frequenza delle lezioni, ma vediamo che una solida maggioranza prevede almeno una frequenza parziale dei corsi, dimostrando quindi l'intenzione di contemperare studio e lavoro.

Tabella 5.15 – Frequenza secondo l'impegno lavorativo (179 soggetti)

		Frequenza regolare		Frequenza saltuaria		Non frequentante	
		Frequenza	%	Frequenza	%	Frequenza	%
Tempo di lavoro	Pieno	6	8,8	42	61,8	20	29,4
	Parziale	21	28,4	43	58,1	10	13,5
	Saltuario	16	43,2	21	56,8	-	-
Totali di gruppo		43	24,0	106	59,2	30	16,8

Percentuali calcolate sulle righe.

### 5.2.5. Scelta del corso di studi

Al di là dei modelli familiari o del tempo disponibile, cosa spinge a iscriversi proprio al Corso di laurea in Scienze dell'Educazione e della Formazione? Di fronte a questa richiesta i nuovi iscritti dell'A.A. 2003-04 oscillano tra la possibilità di sbocco professionale offerto dal corso e la pura passione. In questo caso si è mantenuta la distinzione fra il gruppo

delle matricole e quello dei già iscritti, e come si può osservare le motivazioni strettamente culturali di questi ultimi risultano prevalenti, laddove la maggioranza delle matricole privilegia la possibilità di sbocco professionale offerta dal titolo di studio.

Tabella 5.16 – Motivi per la scelta del corso (419 soggetti)

		Matricole		Già iscritti	
		Frequenza	%	Frequenza	%
Motivi della scelta	Sbocchi professionali	167	50,3	35	40,2
	Bagaglio culturale	80	24,1	41	47,1
	Per esclusione	10	3,0	2	2,3
	Passione	126	38,0	43	49,4
	Altro	10	3,0	4	4,6

Totali percentuali superiori a 100 (possibili più risposte).

Le motivazioni che guidano la scelta sono classificate da Catalano e Figà Talamanca (2002), con la distinzione fra motivi vocazionali, funzionali, familiari e casuali, con i primi due considerati quelli più decisivi nella modalità di scelta e una progressiva crescita negli ultimi anni delle motivazioni di tipo funzionale<sup>78</sup>. Questa crescita è confermata anche per i nuovi iscritti di questo corso di laurea, infatti rispetto alla rilevazione 2001-02, limitando il confronto alle motivazioni espresse dalle matricole, lo stacco esistente fra gli sbocchi professionali e la semplice passione cresce, con una parallela diminuzione della scelta motivata dall'arricchimento culturale.

#### 5.2.6. Autovalutazione in lingua straniera

Un ultimo elemento descrittivo per l'anno 2003-04, rilevato dal questionario dei nuovi iscritti, riguarda l'autovalutazione nella conoscenza delle lingue straniere. La Tabella 5.17 riepiloga non solo la dichiarazione di conoscenza, ma anche la distribuzione degli studenti sulla scala di conoscenza.

<sup>78</sup> Si tratta delle rilevazioni del terzo ciclo di EURO STUDENT analizzate da Catalano e Figà Talamanca (2002), in cui però le motivazioni legate al completamento personale dell'istruzione e all'interesse specifico per gli studi risultano ancora prevalenti.

Tabella 5.17 – Autovalutazione conoscenza lingue straniere (419 soggetti)

		Inglese		Francese		Spagnolo		Tedesco	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%
Auto- valu- tazione lingue	Ottimo	10	2,4	6	1,4	3	,7	2	,5
	Buono	117	27,9	65	15,5	8	1,9	4	1,0
	Suffic.	187	44,6	72	17,2	18	4,3	6	1,4
	Scarso	79	18,9	33	7,9	16	3,8	5	1,2
	No	26	6,2	243	58,0	374	89,3	402	95,9
Totali di gruppo		419	100,0	419	100,0	419	100,0	419	100,0

Naturalmente l'inglese risulta ampiamente la lingua straniera più conosciuta, ma risulta in crescita, rispetto a precedenti leve di studenti del corso di laurea, la lingua francese, con il 42% che dichiara di conoscerla, anche ad un livello migliore dell'inglese, se consideriamo la distribuzione interna dei dati.

In leggero aumento anche la conoscenza dello spagnolo, mentre il tedesco resta limitato a pochissimi studenti. Bisogna considerare che nel questionario era possibile indicare più lingue straniere conosciute, e risultano nettamente in aumento rispetto al 2001-02 (quando erano circa il 30%) i nuovi iscritti che dichiarano di conoscere più lingue straniere.

Tabella 5.18 – Lingue straniere conosciute (419 soggetti)

		Matricole e nuovi iscritti	
		Frequenza	%
Lingue straniere	4	1	0,2
	3	51	12,2
	2	130	31,0
	1	230	54,9
	Nessuna	7	1,7
Totale		419	100,0

Naturalmente si tratta sempre di autodichiarazioni degli studenti, che andrebbero verificate attraverso un'analisi puntuale dei risultati relativi alle prove di idoneità linguistica obbligatorie previste nel piano di studi.

Nel complesso quindi, considerando da una parte il confronto con le rilevazioni relative alle precedenti leve dello stesso corso di laurea, dall'altra alcuni riferimenti con la situazione nazionale dei nuovi iscritti,

il gruppo dei 419 studenti che hanno partecipato alla prova di ingresso appare sostanzialmente rappresentativo e caratterizzato da una accettabile varianza per quanto riguarda le variabili di sfondo indagate dal questionario.

### 5.3. Item analisi del TCL

Anche se lo strumento è stato collaudato nella prova pilota, risulta opportuno, prima di passare all'analisi della graduatoria prodotta, verificare nuovamente il funzionamento degli item della prova e la sua attendibilità complessiva.

L'immissione delle risposte degli studenti è stata svolta, per gli studenti di tutte le sedi e di tutte le sessioni, sempre con l'ausilio di un lettore ottico, con successiva pulizia dei dati segnalati dallo stesso programma di riconoscimento delle risposte. In questa fase non si sono riscontrati problemi particolari.

In accordo con la procedura messa a punto nel corso della prova pilota l'item analisi è stata svolta secondo i due modelli, classica e IRT, e in questa fase l'analisi IRT è stata realizzata secondo il modello standard, con la distinzione fra omissioni volontarie e quesiti non raggiunti, poiché lo scopo prioritario è la verifica dell'efficienza dello strumento.

I risultati delle due item analisi sono riepilogati nella Tabella 5.19, costruita secondo il modello della Tabella 4.3, con l'item analisi classica a sinistra e quella IRT nella parte destra. In questo caso sono state fatte due integrazioni: si è reso necessario aggiungere una colonna relativa agli errori di risposta, riscontrati in cinque item (ma per un totale complessivo di soli sei errori su un totale di 17.010 risposte; inoltre il calcolo della discriminatività per il modello classico, è stato effettuato sia attraverso il confronto degli estremi inferiore e superiore della distribuzione, come è stato fatto per l'item analisi della prova pilota (cfr. 4.4) sia con la correlazione punto biseriale per ogni item, sempre calcolata per mezzo del foglio di calcolo Excel.

La tabella si riferisce all'intero gruppo che ha compilato la prova, quindi ai 567 studenti di Roma (419) e Bologna (148). Per considerare la qualità degli item è opportuno verificare l'andamento complessivo delle risposte, mentre per l'analisi specifica delle graduatorie relative agli studenti dell'Università "La Sapienza" verranno considerati i punteggi derivati dalle item analisi specifiche dei 419 rispondenti del gruppo di Roma.

Tabella 5.19 - Item analisi prova TCL (567 rispondenti)

Item	Ch.	Modello classico									Modello IRT (2 parametri)					
		A	B	C	D	Err.	Omis.	Fac.	Dis.	Pbis	A	B	C	D	Altro	PBt
1	C	24,7	10,1	39,0	21,5	0	4,8	0,39	0,16	0,13	1	4	12	-12	-11	0,12
2	A	15,2	24,7	24,7	34,2	0	1,2	0,15	0,17	0,18	22	9	-24	1	-11	0,22
3	D	4,4	7,9	19,4	63,7	0	4,6	0,64	0,34	0,28	-9	-21	-8	30	-19	0,30
4	B	39,5	27,0	29,5	1,8	0	2,3	0,27	0,33	0,32	-20	39	-11	-5	-10	0,39
5	B	1,8	52,6	0,5	44,1	0	1,1	0,53	0,31	0,29	-9	36	-9	-30	-11	0,36
6	A	85,0	1,8	3,2	7,6	0	2,5	0,85	0,15	0,15	13	-3	-9	-2	-13	0,13
7	A	76,5	4,2	7,4	9,2	0	2,6	0,77	0,25	0,26	27	-11	-14	-13	-11	0,27
8	D	12,5	1,9	4,6	76,2	0	4,8	0,76	0,35	0,33	-22	-13	-14	38	-21	0,38
9	D	20,6	3,0	2,3	70,2	0	3,9	0,70	0,35	0,31	-20	-9	-16	35	-22	0,35
10	D	3,2	6,5	10,4	75,0	0	4,9	0,75	0,40	0,37	-14	-20	-19	44	-27	0,44
11	D	2,8	6,7	4,2	78,5	0	7,8	0,78	0,36	0,33	-9	-11	-12	34	-28	0,34
12	D	7,6	31,9	9,3	46,4	0,4	4,4	0,46	0,44	0,37	-11	-22	-13	39	-10	0,39
13	B	10,4	67,4	11,3	6,3	0	4,5	0,67	0,37	0,30	-16	28	-5	-11	-19	0,28
14	D	12,0	27,5	8,8	33	0	18,7	0,33	0,35	0,30	4	-4	-23	30	-18	0,30
15	C	5,5	3,0	82,7	2,5	0,2	6,2	0,83	0,27	0,26	-19	-6	26	-8	-13	0,26
16	C	4,9	9,9	22,4	59,1	0	3,7	0,22	0,19	0,18	0	-14	19	-4	-10	0,19
17	D	18,7	44,4	5,8	23,3	0	7,8	0,23	0,18	0,21	-18	5	-6	21	-11	0,21
18	B	7,1	54,9	20,1	6,0	0	12	0,55	0,33	0,30	-1	31	-17	-6	-21	0,31
19	A	37,6	38,1	4,8	12,9	0	6,7	0,38	0,33	0,29	34	-11	-14	-13	-18	0,34
20	B	13,2	32,6	21,5	19,8	0	12,9	0,33	0,42	0,34	-9	32	-3	-15	-16	0,32
21	D	6,0	5,3	4,4	78,5	0,2	5,6	0,78	0,43	0,42	-15	-20	-23	38	-12	0,38
22	D	0,9	16,0	6,7	61,7	0,2	14,5	0,62	0,57	0,47	-4	-16	-26	40	-21	0,40
23	A	36,9	37,6	4,6	10,1	0	10,9	0,37	0,50	0,43	41	-20	-13	-13	-18	0,41
24	C	6,7	5,5	67,5	0,7	0	19,6	0,68	0,53	0,45	-11	-13	30	-17	-19	0,30
25	D	4,1	18,2	2,3	57,7	0,2	17,6	0,58	0,61	0,52	-12	-29	-15	42	-16	0,42
26	C	10,8	5,8	49,4	4,6	0	29,5	0,49	0,75	0,58	-21	-18	47	-6	-29	0,47
27	A	57,7	5,8	2,8	2,3	0	31,4	0,58	0,73	0,59	45	-12	-25	-15	-29	0,45
28	C	3,4	6,9	41,3	19,0	0	29,5	0,41	0,71	0,56	-18	-15	46	-22	-22	0,46
29	B	7,1	45,9	5,5	9,5	0	32,1	0,46	0,63	0,49	-13	30	-11	-4	-25	0,30
30	A	61,9	1,8	1,9	5,9	0	28,6	0,62	0,67	0,54	28	-17	-12	-18	-2	0,28

Nella parte sinistra della tabella è riportata l'analisi degli item secondo il modello classico, con la discriminatività calcolata attraverso il confronto degli estremi (Dis.) e il calcolo del punto biseriale (Pbis). Nella parte destra è riportata l'analisi effettuata con il metodo IRT a due parametri (software XCalibre). Le prime colonne (A-D e Altro, l'insieme di errori di risposta e omissioni) riportano le correlazioni esistenti fra ogni alternativa e il tratto latente della prova ( $\theta$ ), che dovrebbe essere positiva e il più alta possibile per

l'alternativa corretta, negativa per i distrattori e Altro. L'ultima colonna (PBt) riporta la misura equivalente del punto biseriale per il metodo IRT, cioè la correlazione (Pearson) fra le risposte esatte/non esatte e il tratto latente della prova (e non il punteggio totale come per il punto biseriale). Sono state evidenziate in grigio le misure critiche.

Ci sono alcune differenze rispetto all'item analisi della prova pilota, e non sempre riguardano i quesiti leggermente modificati in seguito al collaudo preventivo. In particolare per l'item analisi classica notiamo che:

- l'item 2 risulta abbastanza difficile, confermando quanto già verificato nella prova pilota (0,14);
- gli item 6 e 15 risultano abbastanza facili, con piccoli aumenti degli indici relativi, peraltro già vicini alla soglia di 0,80 nella prova pilota (rispettivamente 0,76 e 0,77), mentre l'indice degli item 8 e 10 è migliorato rispetto alla prova pilota;
- gli item 1, 2, 6, e 16 presentano indici di discriminatività critici ( $<0,20$ ), rilevati nella prova pilota solo per l'item 2; l'item 17 oscilla intorno alla soglia critica considerando le due modalità di calcolo utilizzate, anche se risulta migliorato rispetto alla prova pilota;
- negli item 2, 4, 16 e 17 la risposta corretta presenta una percentuale di scelta inferiore a una o più alternative di risposta errate, come già verificato nella prova pilota, mentre l'item 14 non presenta più questo problema.

Considerando tutti gli aspetti indicati, nel complesso secondo l'item analisi classica risultano critici gli item 2, 6 e 16.

L'item analisi IRT riscontra problemi simili per quasi tutti gli item già evidenziati in precedenza:

- gli item 1, 6 e 16 presentano un punto biseriale (PBt) basso, ma solo nel caso del 16 si tratta di una conferma rispetto alla prova pilota, mentre per gli item 2 e 17 il dato relativo della prova pilota risulta migliorato;
- gli item 1, 2, 14 e 17 hanno 1 o 2 (item 1) distrattori con correlazione positiva, dato già riscontrato per gli item 2, 14 e 17 nella prova pilota, d'altra parte per gli item 5, 6, 20 e 29 il dato critico della prova pilota è migliorato nella somministrazione principale.

In complesso però l'unico item segnalato come potenzialmente problematico da XCalibre per l'analisi IRT è il 6, poiché uno dei parametri di stima (quello relativo alla facilità) risulta appena oltre il limite relati-

vo, ma considerando che l'altro parametro (basato sulla discriminatività) risulta normale l'item può essere conservato.

Di passaggio notiamo che gli item 2 e 16, segnalati come critici dall'item analisi classica, riportano parametri di stima pienamente accettabili, cioè il modello IRT riesce ad assorbire le criticità senza incidere negativamente sull'analisi degli item e conseguentemente sui punteggi dei rispondenti.

Nel complesso la prova TCL presenta una facilità media di 0,54, quindi leggermente più difficile rispetto alla versione collaudata nella prova pilota<sup>79</sup>. La facilità relativa ai singoli testi è riepilogata nella tabella seguente, e ricalca quella verificata (Tabella 3.4) nella prova pilota.

Tabella 5.20 – Facilità media dei diversi testi della prova TCL

Testo	N. item	Posiz. item	Facilità media
Legge sul prestito d'onore	4	1-4	0,36
Assunzioni previste nelle imprese	7	5-11	0,73
La rappresentazione nella professione	3	12-14	0,49
La Francia e le trentacinque ore	5	15-19	0,44
Introduzione al saggio di John Dewey	11	20-30	0,54

La prova presenta una buona consistenza interna ( $KR_{20} = 0,74$ ), in discesa rispetto al valore misurato nella prova pilota, ma ancora su livelli che denotano una buona affidabilità del pacchetto di quesiti.

Per quanto riguarda i punteggi degli studenti colpisce l'incidenza delle non risposte nella parte finale della prova, soprattutto se si considera la distinzione operata dal modello di Rasch fra omissioni e domande non raggiunte (omissioni nella parte finale della prova); come si può vedere nella tabella di item analisi per le ultime 5 domande la percentuale delle non risposte è intorno al 30%.

Nella Tabella 5.21 vengono riepilogate le percentuali di rispondenti che non hanno raggiunto l'ultima domanda della prova. Nel complesso si tratta del 28,2 %, una percentuale nettamente più alta rispetto a quella rilevata nella prova pilota, in cui la maggior parte degli studenti aveva

<sup>79</sup> Ricordiamo però che la precedente versione era composta da 39 quesiti e con un testo in più, la cui facilità media era 0,69.

completato la prova<sup>80</sup>. Ciò può essere dovuto a una effettiva maggiore velocità di lettura degli studenti che hanno svolto la prova pilota, che ricordiamo erano prossimi allo svolgimento dell'Esame di Stato, quindi in una fase in cui si avvicina la massima capacità di performance. Un effetto negativo sui nuovi iscritti che hanno partecipato alla prova è rappresentato dal carattere non selettivo della prova stessa (esplicitamente dichiarato nelle istruzioni), che probabilmente ha indotto una certa rilassatezza nell'affrontare i quesiti, dedicando molta attenzione alla lettura dei primi testi per poi ritrovarsi con poco tempo per il testo finale, che comportava ben 11 quesiti. Non è un caso se due picchi di percentuale di domande non raggiunte si hanno in corrispondenza di alcune soglie psicologiche, come gli ultimi 11 quesiti (2,3%, vuol dire che è stato completamente saltato l'ultimo testo) o degli ultimi 5 quesiti (4,8%, vuol dire che è stata saltata la seconda colonna di domande prevista per l'ultimo testo).

Tabella 5.21 – Riepilogo risposte prova TCL – Rispondenti che non completano la prova (NR=Non Raggiunge)

	Stud.	%NR ultima domanda	%NR1-3	%NR4-6	%NR7-9	%NR>9
Totale	567	28,2	5,6	10,1	7,9	4,6
Bologna	148	23,6	2,0	6,1	9,5	6,1
Roma	419	29,8	6,9	11,5	7,4	4,1
Rm-09	170	26,5	8,8	8,2	7,1	2,4
Rm-10	79	24,1	5,1	8,9	5,1	5,1
Rm-11.1	95	33,7	2,1	14,7	10,5	6,3
Rm-11.2	75	38,7	10,7	17,3	6,7	4,0

Da notare che il tasso di coloro che non hanno raggiunto l'ultima domanda è sensibilmente più alto nel gruppo romano (29,8%) rispetto a quello di Bologna (23,6), ma con marcate differenze rispetto alle sessioni che si sono svolte a Roma: mentre per le prime due (settembre-ottobre 2003) i tassi di domande non raggiunte sono vicini al dato di Bologna (rispettivamente 26,5% e 24,1%) le omissioni finali salgono di molte per

<sup>80</sup> Ricordiamo (cfr 4.3) che nella prova pilota erano stati necessari a diversi studenti circa 10 minuti in più del tempo assegnato per completare la prova, che però era costituita da sei testi e da 39 quesiti, rispetto ai 5 testi e 30 quesiti della versione definitiva.



entrambe le sessioni di novembre, con percentuali del 33,7% (8 novembre 2003) e del 38,7% (15 novembre 2003).

Ricordiamo che in particolare quest'ultimo gruppo era costituito prevalentemente da studenti iscritti negli ultimi giorni utili, presumibilmente meno motivati o arrivati all'iscrizione nel Corso di laurea in Scienze dell'Educazione e della Formazione per esclusione rispetto ad altri indirizzi per cui, già in quell'anno accademico, erano previste prove selettive di accesso.

Naturalmente, come già accennato nel Paragrafo 2.4, la possibile diversa considerazione delle omissioni rispetto alle domande non raggiunte può influire sulla graduatoria determinata dall'analisi IRT (come vedremo tra poco), ma ricordando quanto sia complesso distinguere fattori di tipo personali (quali la motivazione o l'interesse) rispetto alla competenza specifica di comprensione, della quale peraltro fa parte anche la velocità di lettura.

Considerando complessivamente l'incidenza delle omissioni sul totale delle risposte fornite, le differenze fra i due gruppi principali e i sottogruppi di Roma restano evidenti.

Tabella 5.22 – Riepilogo risposte prova TCL – Incidenza delle omissioni

	Stud.	%Esatte	%Errate	%Omissioni
Totale	567	53,9	34,9	11,2
Bologna	148	59,3	31,2	9,5
Roma	419	52,0	36,2	11,8
Rm-09	170	54,5	34,7	10,8
Rm-10	79	50,9	36,9	12,2
Rm-11.1	95	49,1	38,9	12,0
Rm-11.2	75	51,2	35,2	13,6

Gli oltre sette punti percentuali di differenza fra studenti romani e bolognesi sono spiegati soprattutto dal maggior tasso di errore (+5%), ma anche dall'aumento netto delle omissioni (+2,3%), e all'interno del gruppo romano, il sottogruppo della prima somministrazione (Rm-09) raggiunge percentuali migliori del gruppo bolognese, sia per le risposte esatte che per le errate e le omissioni, confermando le discrete differenze fra i sottogruppi romani.

La relativa maggiore incidenza di omissioni rilevata nel gruppo di studenti romani si riflette sull'andamento delle statistiche relative ai due gruppi, con una netta prevalenza degli studenti di Bologna, che rispondono correttamente in media a due quesiti in più.

Entrambi i gruppi presentano un coefficiente di variazione (CdV) ben superiore a 20, denotando quindi una discreta disomogeneità nei punteggi, più marcata nel gruppo romano, dove la gamma cresce di 2 punti (26) rispetto ai colleghi di Bologna (24), espandendosi in entrambe le direzioni, visto che i valori estremi di minimo e massimo sono entrambi stati raggiunti a Roma.

Bisogna ricordare che a Bologna la partecipazione è stata caratterizzata da una assoluta volontarietà di adesione, mentre a Roma, pur non presentandosi come prova obbligatoria, la presenza alla prova è stata fortemente incoraggiata e ha interessato la stragrande maggioranza dei nuovi iscritti al corso di laurea (94%), coinvolgendo presumibilmente anche studenti di fasce di livello inferiore.

Tabella 5.23 – Confronto delle statistiche descrittive per i due gruppi: prova TCL

	Roma	Bologna
Soggetti	419	148
Media	15,6	17,8
Dev std.	4,8	4,9
CdV	29,9	27,5
Esatte (%)	52,0	59,3
Min	4	5
Max	29	28
Asimmetria	0,25	-0,43
Curtosi	-0,37	-0,56

Da segnalare anche la diversità esistente fra i due gruppi nella distribuzione dei punteggi, in particolare per il dato riguardante l'asimmetria, che risulta leggermente positiva a Roma, denotando quindi una maggiore abbondanza di punteggi inferiori alla media, mentre a Bologna il dato è più marcatamente negativo, con uno spostamento consistente dei punteggi verso i valori superiori alla media.

L'item analisi, sempre secondo i due modelli oggetto di indagine, è stata replicata considerando le risposte dei soli studenti di Roma, in quanto le graduatorie simulate per l'accesso devono necessariamente riferirsi a una realtà universitaria omogenea.

La Tabella 5.24 riepiloga tutti i dati relativi alle risposte del gruppo di Roma.

Tabella 5.24 - Item analisi prova TCL (419 rispondenti - Roma)

Item	Ch.	Modello classico									Modello IRT (2 parametri)					
		A	B	C	D	Err.	Omis.	Fac.	Dis.	Pbis	A	B	C	D	Altro	PBt
1	C	23,6	9,8	36,8	24,6	0	5,3	0,37	0,09	0,13	0	1	11	-8	-8	0,11
2	A	14,8	22,7	27,7	33,7	0	1,2	0,15	0,11	0,18	21	7	-20	-1	-8	0,21
3	D	4,8	10,0	20,0	60,1	0	5,0	0,60	0,24	0,26	-6	-18	-5	26	-18	0,26
4	B	39,1	27,2	28,4	2,1	0	3,1	0,27	0,29	0,35	-19	44	-17	-7	-10	0,44
5	B	2,4	51,3	0,7	44,6	0	1,0	0,51	0,34	0,33	-8	40	-10	-34	-8	0,40
6	A	85,0	2,4	3,1	7,4	0	2,1	0,85	0,14	0,13	12	-2	-11	-4	-7	0,12
7	A	75,4	4,1	7,9	10,3	0	2,4	0,75	0,20	0,23	25	-8	-16	-11	-11	0,25
8	D	13,1	1,7	5,5	74,2	0	5,5	0,74	0,31	0,31	-20	-13	-13	36	-18	0,36
9	D	21,2	3,1	2,6	68,5	0	4,5	0,68	0,34	0,32	-18	-10	-18	35	-21	0,35
10	D	3,6	7,4	11,7	71,4	0	6,0	0,71	0,31	0,34	-14	-19	-14	40	-25	0,40
11	D	3,8	8,1	4,1	74,5	0	9,5	0,74	0,31	0,33	-8	-10	-13	33	-27	0,33
12	D	8,4	34,4	10,0	42,5	0	4,8	0,42	0,36	0,32	-9	-17	-11	32	-11	0,32
13	B	12,2	66,3	7,6	7,9	0	6,0	0,66	0,35	0,32	-14	29	-6	-10	-19	0,29
14	D	12,4	25,5	10,5	29,8	0	21,7	0,30	0,24	0,29	4	-2	-24	29	-16	0,29
15	C	7,2	3,1	80,0	2,6	0	7,2	0,80	0,20	0,21	-18	-1	20	-8	-8	0,20
16	C	4,8	9,1	21,0	60,6	0	4,5	0,21	0,20	0,20	-2	-13	20	-5	-9	0,20
17	D	20,8	42,0	5,5	22,9	0	8,8	0,23	0,19	0,25	-20	4	-5	24	-11	0,24
18	B	6,4	52,5	21,0	6,2	0	13,8	0,53	0,28	0,31	-5	33	-17	-6	-20	0,33
19	A	35,3	40,3	4,5	12,6	0	7,2	0,35	0,25	0,26	30	-7	-13	-13	-18	0,30
20	B	13,1	32,0	19,8	21,7	0	13,4	0,32	0,34	0,30	-7	30	-7	-13	-14	0,30
21	D	6,7	6,4	5,7	75,7	0	5,5	0,76	0,34	0,39	-16	-18	-24	38	-13	0,38
22	D	0,5	16,7	8,4	59,7	0	14,8	0,60	0,53	0,47	-1	-20	-26	43	-19	0,43
23	A	33,4	40,3	4,8	11,7	0	9,8	0,33	0,44	0,42	43	-23	-11	-14	-15	0,43
24	C	8,1	6,2	65,9	1,0	0	18,9	0,66	0,46	0,43	-9	-13	31	-19	-19	0,31
25	D	4,1	20,3	2,6	55,8	0	17,2	0,56	0,54	0,47	-8	-29	-12	39	-12	0,39
26	C	11,5	6,2	48,2	4,3	0	29,8	0,48	0,69	0,55	-20	-15	43	-7	-26	0,43
27	A	56,6	5,7	3,1	2,6	0	32,0	0,57	0,69	0,58	46	-14	-28	-15	-26	0,46
28	C	3,1	7,4	37,9	21,5	0	30,1	0,38	0,64	0,54	-14	-16	45	-23	-19	0,45
29	B	7,4	44,9	5,5	8,6	0	33,7	0,45	0,56	0,50	-11	32	-14	-11	-20	0,32
30	A	60,4	1,9	2,6	5,0	0	30,1	0,60	0,61	0,51	27	-20	-12	-15	-1	0,27

Nella parte sinistra della tabella è riportata l'analisi degli item secondo il modello classico, con la discriminatività calcolata attraverso il confronto degli estremi (Dis.) e il calcolo del punto biseriale (Pbis). Nella parte destra è riportata l'analisi effettuata con il metodo IRT a due parametri (software XCalibre). Le prime colonne (A-D e Altro, l'insieme di errori di risposta e omissioni) riportano le correlazioni esistenti fra ogni alternativa e il tratto latente della prova ( $\theta$ ), che dovrebbe essere positiva e il più alta possibile per

l'alternativa corretta, negativa per i distrattori e Altro. L'ultima colonna (Pbt) riporta la misura equivalente del punto biseriale per il metodo IRT, cioè la correlazione (Pearson) fra le risposte esatte/non esatte e il tratto latente della prova (e non il punteggio totale come per il punto biseriale). Sono state evidenziate in grigio le misure critiche.

Rispetto alla Tabella 5.19, che presenta l'item analisi relativa all'intero gruppo dei rispondenti, le differenze risultano minime.

Per l'item analisi classica vengono confermati i due item con eccessiva facilità (6 e 15) e l'unico item molto difficile (2).

Problemi di discriminatività sembrano confermati per gli item 1, 2, 6 e 16, che presentano valori critici sia per il confronto degli estremi, sia per il calcolo del punto biseriale, mentre risulta leggermente peggiorato il dato di discriminatività riscontrato per gli item 3, 4 e 14, per i quali però il punto biseriale resta soddisfacente.

Nel complesso quindi, secondo l'item analisi classica, la prova si mostra sostanzialmente equivalente tra l'intero gruppo di rispondenti e il solo gruppo di Roma, con due soli quesiti dubbi (2 e 6).

Per l'item analisi IRT si può notare che solo due item (1 e 6) presentano valori critici di discriminatività, mentre il 16, che risultava critico nell'item analisi IRT relativa all'intero gruppo dei rispondenti, raggiunge un valore accettabile. Da considerare comunque che in questo caso XCalibre non ha segnalato pericoli di eliminazione per alcun item, quindi il pacchetto di quesiti utilizzati per la prova risulta complessivamente efficiente, anche se riferito in maniera esclusiva al solo gruppo di rispondenti romani.

Allo stesso modo viene confermato il dato relativo alla consistenza interna della prova, con un indice KR20 di 0,73, di pochissimo inferiore a quello rilevato per l'intero gruppo dei 567 studenti partecipanti.

#### 5.4. Item analisi del Cloze

La seconda prova di comprensione della lettura, basata sul Cloze, è stata svolta dall'intero gruppo di nuovi iscritti di Roma "La Sapienza" (419), mentre 6 studenti di Bologna, pur avendo compilato la prova TCL, non hanno completato la sessione di lavoro (142 rispondenti validi). Per cui il totale dei rispondenti, sulla cui base sono state svolte le procedure di item analisi, è di 561.

Il lavoro di correzione della prova Cloze si è naturalmente rivelato più complesso rispetto a quello svolto per la prova TCL. In quanto si è deciso, in accordo con la collega Truffelli, di considerare corretti non solo i termini esatti rispetto al testo originale, ma anche i termini equivalenti da tutti i punti di vista, sia grammaticale (genere e numero per sostantivi e aggettivi, persona e numero per le forme verbali<sup>81</sup>) sia semantico. Questa decisione è stata presa pur considerando che è stata dimostrata la sostanziale equivalenza delle graduatorie ottenute con la semplice considerazione del termine esatto e con l'accettazione dei termini equivalenti<sup>82</sup>, in quanto l'accettazione di più termini permette di ampliare la scala dei punteggi, e di conseguenza può favorire la definizione di una graduatoria utile ai fini selettivi<sup>83</sup>.

Il lavoro di scelta dei possibili termini equivalenti è stato svolto in collaborazione con la collega Truffelli, partendo dal riscontro dei termini alternativi effettivamente utilizzati dagli studenti in luogo dei termini esatti. Oltre ai criteri sopra specificati è stata considerata accettabile anche una variante grafica ("ke" per "che") ormai entrata nell'uso abituale del linguaggio giovanile, anche se il termine non è attestato nei dizionari in modo specifico<sup>84</sup>. Il quadro complessivo delle alternative considerate equivalenti è stato infine verificato con alcuni docenti del collegio di dottorato.

Prima di passare all'item analisi della prova Cloze è però necessario premettere che l'andamento dei punteggi relativi ai singoli testi che

<sup>81</sup> In alcuni casi particolari sono stati considerati equivalenti dal punto di vista grammaticale anche modi verbali diversi, quali indicativo e congiuntivo, laddove l'uso alternativo di tali modi producesse sfumature di significato ugualmente accettabili rispetto al testo (cfr. Rohlf 1969).

<sup>82</sup> Questo vuol dire che la considerazione anche dei termini equivalenti non falsifica i risultati della prova; cfr. Lucisano (1992), pur pronunciandosi per l'accettazione della sola risposta esatta, cita studi precedenti in cui è stata accertata l'esistenza di una specifica capacità di riempire i buchi del Cloze con termini accettabili, soprattutto dal punto di vista sintattico (Martinez 1978).

<sup>83</sup> Nello specifico, a titolo esemplificativo, la gamma dei punteggi rilevata per il primo testo della prova (testo antropologico) è passata da 27 (solo termini esatti) a 29 (anche termini equivalenti), premiando soprattutto i rispondenti che già si trovavano nella parte alta della distribuzione dei punteggi.

<sup>84</sup> Bisogna però segnalare che nel dizionario De Mauro (2002), utilizzato per la verifica dei termini alternativi, nella definizione della lettera K si sostiene che "viene usata solo in forestierismi o talvolta scherzosamente o con intento denigratorio in sostituzione della c velare".

componevano la prova è risultato critico, inducendo il sospetto che il fattore stanchezza abbia influenzato il risultato in misura sempre maggiore. Se consideriamo la Tabella 5.25, vediamo che il punteggio medio tende a scendere, come pure la percentuale di risposte esatte, mentre sale nettamente anche la percentuale di omissioni, che finisce praticamente per raddoppiare tra il primo e l'ultimo testo. Inoltre diversi studenti, a partire dal secondo testo, rinunciano a compilare la prova (4 per il testo di psicologia) o effettuano una compilazione talmente parziale che non gli permette di ottenere alcun punto (13 studenti a punteggio 0 per il testo di filosofia).

Tabella 5.25 – Confronto delle statistiche descrittive per i quattro testi della prova Cloze

	Testi			
	Antropologia	Psicologia	Pedagogia	Filosofia
Item	31	32	31	31
Media	20,2	17,3	19,8	16,7
Dev std.	3,9	4,8	3,9	5,0
Esatte (%)	65,0	54,1	64,22	54,1
Omesse (%)	6,8	11,2	10,9	14,2
Min	1	0	0	0
Max	29	28	29	27
Asimmetria	-0,89	-0,90	-1,22	-1,33
Curtosi	1,75	1,39	3,55	2,34

Considerando anche altre anomalie nelle statistiche dei diversi testi, quali il valore decisamente anomalo della curtosi per il testo pedagogico o l'evidente calo di interesse mostrato verso il testo di psicologia, si è ritenuto prudente, in accordo con la collega Truffelli che si è occupata dei risultati relativi agli studenti di Bologna, limitare l'analisi al solo primo testo della prova Cloze (Antropologia), per cui anche le graduatorie relative alla prova Cloze si riferiranno esclusivamente ai punteggi conseguiti nel primo testo.

In relazione a quanto si è appena esposto, la Tabella 5.26 riepiloga le chiavi di risposta utilizzate per il testo antropologico, distinguendo il termine esatto e le alternative considerate equivalenti. Per agevolare la lettura della tabella, il termine esatto è evidenziato all'interno del testo di riferimento

Tabella 5.26 – Chiavi di risposta del testo Antropologia

Item	Testo – TERMINE ESATTO	Alternative equivalenti
1	Per questo genere di STUDI,	Ricerche - Indagini - Analisi
2	l'antropologo sceglie popoli MOLTO semplici,	Piuttosto - Alquanto - Estremamente - Decisamente - Assai - Particolarmente - Abbastanza
3	popoli primitivi la CUI	
4	società non abbia mai RAGGIUNTO la complessità della nostra.	Conosciuto - Sviluppato - Eguagliato - Avuto
5	IN questa precisa scelta	
6	di POPOLI primitivi, come gli Esquimesi,	
7	GLI australiani,	
8	gli isolani dei MARI del Sud America	
9	o GLI Indiani Pueblos,	
10	l'antropologo È guidato dalla convinzione	Viene
11	che L'analisi di una	
12	civilizzazione PIÙ semplice	Molto - Piuttosto - Alquanto - Abbastanza - Decisamente - Particolarmente - Estremamente - Strutturalmente - Sicuramente - Relativamente
13	è di più FACILE attuazione.	Immediata - Pratica
14	Nelle civiltà complicate COME quelle europee,	
15	o come QUELLE più elevate dell'Oriente	Le
16	OCCORRONO anni di lavoro	Servono - Passano - Trascorrono
17	prima CHE lo studioso	Ke
18	possa cominciare A capire	
19	da quali forze SONO animate.	Vengono - Siano - Vengano
20	Il solo studio DELLA famiglia francese	Sulla
21	comporterebbe uno STUDIO preliminare	
22	della storia francese, DELLA legislazione francese,	
23	della posizione CATTOLICA e protestante	
24	in rapporto AL sesso	
25	e anche alle RELAZIONI personali.	Interazioni - Situazioni - Condizioni
26	un popolo primitivo, SENZA linguaggio scritto,	
27	presenta un PROBLEMA assai meno complicato	Compito - Quadro - Caso - Assetto - Panorama - Sistema
28	e UNO studioso esperto	Lo - Qualsiasi
29	può rendersi CONTO in pochi mesi	
30	della STRUTTURA fondamentale	Organizzazione - Strutturazione - Composizione
31	di una società PRIMITIVA	Arcaica

Come si può notare 15 termini si sono rivelati insostituibili, mentre sono state proposte alternative per 16 termini. Di questi, quattro sono verbi, sostituibili sia per il significato (RAGGIUNTO = Conosciuto), sia per il modo (SONO = Siano); quattro sono sostantivi, di cui tre classificati dal De Mauro come termini fondamentali (STUDI, RELAZIONI, PROBLEMA) e uno come Alto Uso (STRUTTURA); tre sono aggettivi, di cui uno indefinito (MOLTO); infine troviamo anche un avverbio (PIÙ), un pronome (QUELLE), una congiunzione (CHE) una preposizione articolata (DELLA) e un articolo (UNO), che però può anche assumere valore indefinito ed essere sostituito da un aggettivo (Qualsiasi).

La Tabella 5.27, presentata nella pagina a fronte, riepiloga le due item analisi svolte sulle risposte di tutti i 561 rispondenti (Roma e Bologna) del solo testo antropologico, per valutare la qualità complessiva della prova.

Successivamente, come già fatto per la prova di comprensione TCL, verrà fornito il quadro specifico relativo alle item analisi delle risposte fornite dai soli studenti romani, utilizzate per la definizione delle ipotetiche graduatorie di accesso.



Tabella 5.27 - Analisi degli item Cloze (561 rispondenti)

Item	Chiave	Modello classico						Modello IRT (2 parametri)			
		A	B	Omis.	Fac.	Dis.	Pbis	A	B	Altro	PBt
1	A	27,5	72,5	0,0	0,27	0,14	0,12	7	-7	-	0,07
2	A	57,0	39,4	3,6	0,57	0,49	0,36	36	-30	-18	0,36
3	A	80,9	15,7	3,4	0,81	0,38	0,33	34	-29	-15	0,34
4	A	70,2	21,2	8,6	0,70	0,41	0,37	35	-22	-25	0,35
5	A	37,1	60,6	2,3	0,37	0,27	0,28	26	-21	-17	0,26
6	A	83,2	15,9	0,9	0,83	0,20	0,25	23	-21	-12	0,23
7	A	70,9	20,3	8,7	0,71	0,37	0,36	34	-23	-22	0,34
8	A	17,1	57,4	25,5	0,17	0,17	0,19	19	1	-18	0,19
9	A	73,8	16,9	9,3	0,74	0,41	0,41	40	-25	-29	0,40
10	A	88,1	10,2	1,8	0,88	0,19	0,32	29	-22	-21	0,29
11	A	74,2	22,3	3,6	0,74	0,32	0,36	35	-25	-27	0,35
12	A	52,9	33,5	13,5	0,53	0,46	0,38	37	-15	-33	0,37
13	A	62,7	24,6	12,7	0,63	0,41	0,41	39	-19	-33	0,39
14	A	94,3	5,3	0,4	0,94	0,10	0,24	21	-19	-11	0,21
15	A	96,3	2,1	1,6	0,96	0,09	0,22	19	-6	-23	0,19
16	A	51,7	41,7	6,6	0,52	0,58	0,48	53	-44	-21	0,53
17	A	91,8	4,1	4,1	0,92	0,17	0,32	30	-15	-26	0,30
18	A	93,6	5,7	0,7	0,94	0,09	0,22	21	-19	-9	0,21
19	A	92,7	2,7	4,6	0,93	0,20	0,40	39	-21	-32	0,39
20	A	83,6	14,4	2,0	0,84	0,27	0,35	32	-24	-25	0,32
21	A	83,1	10,7	6,2	0,83	0,30	0,38	36	-23	-26	0,36
22	A	88,6	7,1	4,3	0,89	0,30	0,50	47	-32	-34	0,47
23	A	34,2	44,7	21,0	0,34	0,49	0,40	43	-18	-28	0,43
24	A	83,8	12,1	4,1	0,84	0,33	0,43	40	-29	-27	0,40
25	A	16,9	73,1	10,0	0,17	0,18	0,17	15	0	-20	0,15
26	A	26,2	66,0	7,8	0,26	0,37	0,37	41	-33	-9	0,41
27	A	14,4	72,2	13,4	0,14	0,11	0,16	14	-5	-8	0,14
28	A	94,5	2,7	2,9	0,94	0,13	0,41	27	-19	-20	0,27
29	A	97,5	1,2	1,2	0,98	0,07	0,35	14	-14	-	0,14
30	A	18,7	65,8	15,5	0,19	0,28	0,31	31	-24	-8	0,31
31	A	57,6	31,0	11,4	0,58	0,36	0,30	17	-17	-	0,17

Nella parte sinistra della tabella è riportata l'analisi degli item secondo il modello classico, per convenzione la risposta corretta è classificata con la lettera A, quella errata con la lettera B; la discriminatività è calcolata sia attraverso il confronto degli estremi (Dis.), sia con il calcolo del punto biseriale (Pbis). Nella parte destra è riportata l'analisi effettuata con il metodo IRT a due parametri (software XCalibre).

Le prime colonne (A-B e Altro, corrispondente alle omissioni) riportano la correlazione esistente fra ogni alternativa e il tratto latente della prova ( $\theta$ ), che dovrebbe essere positiva e il più alta possibile per l'alternativa corretta, negativa per i distrattori e Altro. L'ultima colonna (Pbt) riporta la misura equivalente del punto biseriale per il metodo IRT, cioè la correlazione (Pearson) fra le risposte esatte/non esatte e il tratto latente della prova (e non il punteggio totale come per il punto biseriale). Sono state evidenziate in grigio le misure critiche.

La particolare struttura del Cloze, in particolare la scelta del modello di bucatura sistematica del testo, influenza l'andamento dell'item analisi, in particolare per la presenza di un discreto numero di quesiti che risultano molto facili (7 presentano una facilità  $>0,90$ ), in quanto il termine risulta agevolmente ricostruibile dal contesto.

Per quanto riguarda l'item analisi classica sono da segnalare dati critici per gli item 8, 25 e 27, che risultano troppo difficili ( $fac < 0,20$ ) e con problemi di discriminatività segnalati sia dal confronto fra gli estremi, sia dal valore del punto biseriale. Problemi di discriminatività anche per l'item 1, ma in questo caso la facilità risulta accettabile.

Se passiamo all'analisi IRT, XCalibre conferma il valore critico del punto biseriale, rispetto al tratto latente della prova, per gli item 1, 8, 25 e 27, ma non evidenzia comunque assoluta necessità di esclusione per questi item; vengono segnalati possibili problemi di discriminatività anche per gli item 15, 29 e 31, collegati, per i primi due casi, a un'eccessiva facilità, ma l'analisi dei residui, cioè il controllo specifico dei rispondenti che hanno sbagliato la risposta, non comporta valori del tutto fuori norma, per cui risulta possibile non escludere neanche tali item dall'analisi.

Nel complesso però l'item analisi della prova Cloze evidenzia maggiori criticità rispetto a quella della prova TCL (un solo item su 30 a rischio di esclusione); il miglior funzionamento della prova TCL è attestato anche dalla consistenza interna della prova Cloze, che risulta minore ( $KR20 = 0,69$ ), anche se la *reliability* degli item che compongono il testo antropologico raggiunge quasi la soglia di 0,70, comunemente considerata accettabile (Lord, Novick 1968).

Anche in questo caso procediamo al controllo delle domande non raggiunte. Nel riepilogo della Tabella 5.28 si può notare che, rispetto alla prova TCL, le domande, o meglio i termini non raggiunti, risultano notevolmente diminuiti: gli studenti che lasciano in bianco gli ultimi buchi

del testo antropologico sono poco più del 10%, e solo l'1% circa rinuncia a scrivere per più degli ultimi tre buchi. Bisogna comunque ricordare che il testo antropologico era il primo di una serie di quattro, e nonostante i problemi rilevati in precedenza sull'effetto della stanchezza, è evidente che la maggior parte dei rispondenti ha affrontato anche il secondo testo proposto, per cui è plausibile che abbia cercato di finire del tutto almeno il primo testo. Come per la prova TCL il gruppo di Bologna tende a rispondere di più, ma anche in questo caso, e con un andamento diverso, ci sono notevoli differenze fra i sottogruppi romani: per il Cloze nella terza somministrazione (Rm-11.1) è stato rilevato il minor tasso di termini non raggiunti, mentre nell'ultima somministrazione (Rm-11.2) il dato è praticamente raddoppiato.

Tabella 5.28 – Riepilogo risposte prova Cloze – Rispondenti che non completano la prova (NR=Non Raggiunge)

	Stud.	%NR ultima domanda	%NR1-3	%NR>3
Totale	561	11,4	10,3	1,1
Bologna	142	8,5	8,5	-
Roma	419	12,4	11,0	1,4
Rm-09	170	12,4	11,8	0,6
Rm-10	79	12,7	10,1	2,6
Rm-11.1	95	8,4	7,3	1,1
Rm-11.2	75	17,3	14,6	2,7

Se però consideriamo il fenomeno complessivo delle omissioni, senza distinguere i termini non raggiunti, le percentuali relative ai quattro sottogruppi romani risultano più vicine.

Tabella 5.29 – Riepilogo risposte prova Cloze – Incidenza delle omissioni

	Stud.	%Esatte	%Errate	%Omissioni
Totale	561	65,0	28,2	6,8
Bologna	142	67,2	27,2	5,6
Roma	419	64,2	28,6	7,2
Rm-09	170	64,4	28,0	7,6
Rm-10	79	64,5	28,8	6,7
Rm-11.1	95	64,9	28,5	6,6
Rm-11.2	75	63,0	29,1	8,0

Le oscillazioni sono appena superiori al punto percentuale, e allo stesso modo i tassi di risposta esatta si allineano, con l'eccezione delle risposte della quarta somministrazione (Rm-11.2), che risultano inferiori. In questo caso però il miglior risultato del gruppo di Bologna è spiegato in misura leggermente maggiore dall'aumento delle omissioni degli studenti romani, mentre per la prova TCL prevalevano le risposte errate.

Come già verificato per la prova TCL quindi il risultato degli studenti romani è inferiore a quello dei colleghi bolognesi, anche se la differenza risulta più contenuta, con uno scarto fra le due medie di poco più di un punto, e tre punti percentuali di differenza per il numero di risposte esatte. Il confronto fra i due gruppi è riepilogato nella Tabella 5.30.

Tabella 5.30 – Confronto delle statistiche descrittive per i due gruppi: prova Cloze

	Roma	Bologna
Soggetti	419	142
Media	19,9	21,1
Dev std.	3,9	3,9
CdV	19,4	18,5
Esatte (%)	60,1	64,2
Min	1	6
Max	28	29
Asimmetria	-0,95	-0,76
Curtosi	1,92	1,10

A differenza della prova TCL i coefficienti di variazioni risultano decisamente più contenuti; l'aumento dell'omogeneità dei due gruppi per la prova Cloze indica indirettamente una maggiore efficienza della prova TCL ai fini della definizione di una graduatoria. Come vedremo più avanti la distribuzione dei punteggi risulterà sostanzialmente differente fra le due prove, e risulterà influenzata dall'incidenza della deviazione standard sul punteggio medio, incidenza che risulta minore per la prova TCL, con il conseguente avvicinamento delle soglie di passaggio fra le fasce di livello, che invece risultano più larghe, con maggiori differenze fra i valori soglia, per la prova TCL.

Come abbiamo fatto nel paragrafo precedente, riportiamo nella Tabella 5,31, pagina a fronte, i dati relativi alle item analisi effettuate sulle risposte dei soli nuovi iscritti dell'Università di Roma "La Sapienza", sulla cui base sono stati definiti i punteggi individuali che compongono l'ipotetica graduatoria di accesso.

Tabella 5.31 - Analisi degli item Cloze (419 rispondenti - Roma)

Item	Chiave	Modello classico						Modello IRT (2 parametri)			
		A	B	Omis.	Fac.	Dis.	Pbis	A	B	Altro	PBt
1	A	27,2	72,8	0,0	0,27	0,16	0,15	12	-12	-	0,12
2	A	54,9	41,5	3,6	0,55	0,50	0,37	38	-34	-13	0,38
3	A	81,9	14,1	4,1	0,82	0,37	0,36	37	-33	-15	0,37
4	A	69,5	20,5	10,0	0,69	0,39	0,35	33	-16	-29	0,33
5	A	37,2	60,4	2,4	0,37	0,22	0,25	23	-17	-19	0,23
6	A	83,5	15,8	0,7	0,84	0,20	0,27	27	-23	-16	0,27
7	A	69,2	21,0	9,8	0,69	0,31	0,32	31	-20	-29	0,31
8	A	18,6	56,6	24,8	0,19	0,16	0,20	21	2	-21	0,21
9	A	71,1	18,1	10,7	0,71	0,38	0,40	39	-23	-29	0,39
10	A	86,4	11,2	2,4	0,86	0,20	0,38	35	-25	-27	0,35
11	A	72,3	23,6	4,1	0,72	0,31	0,36	35	-25	-27	0,35
12	A	53,2	31,7	15,0	0,53	0,46	0,40	39	-17	-33	0,39
13	A	61,6	24,3	14,1	0,62	0,44	0,42	41	-18	-34	0,41
14	A	94,5	5,0	0,5	0,95	0,11	0,28	24	-22	-13	0,24
15	A	95,9	2,1	1,9	0,96	0,08	0,24	20	-4	-25	0,20
16	A	49,9	43,7	6,4	0,50	0,54	0,47	52	-44	-18	0,52
17	A	91,9	4,5	3,6	0,92	0,16	0,31	27	-15	-23	0,27
18	A	93,8	5,5	0,7	0,94	0,07	0,21	19	-19	-3	0,19
19	A	92,4	3,1	4,5	0,92	0,19	0,39	37	-22	-29	0,37
20	A	83,8	14,3	1,9	0,84	0,26	0,36	33	-25	-24	0,33
21	A	83,1	10,7	6,2	0,83	0,26	0,36	34	-20	-27	0,34
22	A	87,1	8,1	4,8	0,87	0,32	0,50	47	-32	-34	0,47
23	A	32,5	45,6	22,0	0,32	0,42	0,38	40	-15	-27	0,40
24	A	82,6	13,1	4,3	0,83	0,32	0,41	38	-28	-26	0,38
25	A	16,7	72,6	10,7	0,17	0,15	0,16	15	2	-21	0,15
26	A	21,0	70,4	8,6	0,21	0,29	0,32	36	-27	-8	0,36
27	A	13,4	73,7	12,9	0,13	0,09	0,14	13	0	-14	0,13
28	A	94,3	2,6	3,1	0,94	0,13	0,42	28	-22	-17	0,28
29	A	97,6	1,0	1,4	0,98	0,06	0,36	14	-14	-	0,13
30	A	16,7	66,1	17,2	0,17	0,22	0,28	31	-24	-8	0,28
31	A	58,0	29,6	12,4	0,58	0,32	0,30	17	-17	-	0,16

Nella parte sinistra della tabella è riportata l'analisi degli item secondo il modello classico, per convenzione la risposta corretta è classificata con la lettera A, quella errata con la lettera B; la discriminatività è calcolata sia attraverso il confronto degli estremi (Dis.), sia con il calcolo del punto biseriale (Pbis). Nella parte destra è riportata l'analisi effettuata con il metodo IRT a due parametri (software XCalibre).

Le prime colonne (A-B e Altro, corrispondente alle omissioni) riportano la correlazione esistente fra ogni alternativa e il tratto latente della prova ( $\theta$ ), che dovrebbe essere positiva e il più alta possibile per l'alternativa corretta, negativa per i distrattori e Altro. L'ultima colonna (Pbt) riporta la misura equivalente del punto biseriale per il metodo IRT, cioè la correlazione (Pearson) fra le risposte esatte/non esatte e il tratto latente della prova (e non il punteggio totale come per il punto biseriale). Sono state evidenziate in grigio le misure critiche.

L'item analisi IRT segnala gli stessi problemi di discriminatività rilevati con l'intero gruppo dei rispondenti per gli item 1, 25, 27, 29 e 31; piccoli movimenti intorno alla soglia dello 0,20 del Pbt per alcuni item, con l'item 18 che scende sotto la soglia della piena adeguatezza, mentre gli item 8 e 15, segnalati come critici nella precedente tabella, in questo caso risultano adeguati.

Nel complesso XCalibre segnala gli stessi problemi di coerenza con l'intera prova già rilevati per l'intero gruppo dei rispondenti (item 14, 15, 17, 18, 28 e 29), ma anche in questo caso l'analisi dei residui permette di non escludere tali item dal calcolo dei punteggi individuali degli studenti.

### 5.5. Analisi dei risultati delle prove degli studenti di Roma

Dopo aver verificato la funzionalità degli strumenti utilizzati per la rilevazione si può passare all'analisi dei risultati ottenuti dagli studenti romani che hanno partecipato alla prova di ingresso. Verranno considerati parallelamente gli andamenti dei punteggi secondo i tre modelli di item analisi adottati, classica, IRT standard e IRT con penalizzazione di tutte le omissioni<sup>85</sup>. In una prima tabella vengono fornite le statistiche descrittive standard per i tre diversi modelli; subito dopo, per favorire una immediata comparazione, viene fornita la stessa tabella con i dati normalizzati. Per quanto riguarda i punteggi ottenuti dagli studenti secondo l'item analisi classica, si è proceduto alla loro normalizzazione in punti T<sup>86</sup>. La normalizzazione si renderà comunque indispensabile nel

<sup>85</sup> Cfr. 3.4. Per brevità questo terzo modello nel seguito viene definito IRT Om.

<sup>86</sup> Ricordiamo che la normalizzazione in punti T dei punteggi grezzi si basa sulla seguente formula:  $\text{punti T} = 50 + 10z$ . Per il calcolo dei punti z, necessario per la trasformazione in punti T la formula applicata è:  $\text{punti } z = (\text{punteggio grezzo} - \text{punteggio medio}) / \text{deviazione standard}$ . Di fatto l'unità di misura adottata risulta essere la deviazione standard dell'insieme dei punteggi grezzi, mentre la trasformatio-

momento in cui verranno considerati insieme i risultati del TCL e del Cloze, per cui appare opportuno utilizzare subito punteggi normalizzati. Anche se i punteggi ottenuti secondo l'analisi IRT risultano già normalizzati<sup>87</sup>, si è provveduto a un'ulteriore normalizzazione per renderli immediatamente confrontabili con i punteggi di item analisi classica, naturalmente utilizzando sempre la trasformazione in punti T<sup>88</sup>.

### 5.5.1. Le graduatorie della prova di comprensione TCL

Nella Tabella 5.32 sono riepilogate le statistiche descrittive della prova di comprensione TCL secondo le tre modalità di analisi delle risposte.

Tabella 5.32 – Statistiche descrittive prova TCL

	Classica	IRT	IRT Om
Mediana	15	-0,18	-0,19
Media	15,60 (0,23)	-0,04 (0,05)	-0,07 (0,06)
Dev std.	4,80	1,13	1,22
Min	4	-3,05	-3,36
Max	29	5,06	5,27
Asimmetria	0,25 (0,12)	0,78 (0,12)	0,66 (0,12)
Curtosi	-0,37 (0,24)	1,89 (0,24)	1,45 (0,24)

Fra parentesi l'errore standard relativo.

Per tutti e tre i modelli si può osservare che la mediana si trova a destra della media, cioè presenta un valore più piccolo, quindi la maggioranza dei punteggi si trova sotto la media. Da notare anche che il dato relativo alla media dei due modelli IRT è leggermente diverso da 0: i punteggi forniti da XCalibre sono arrotondati al 4° decimale, per cui il

ne in punti T trasla l'origine 0 (cioè la media) a 50, permettendo di operare con punteggi tutti di segno positivo.

<sup>87</sup> L'unità di misura dei punteggi, già normalizzati, ottenuti con l'analisi IRT è costituita dal logit, che corrisponde al logaritmo naturale del rapporto tra risposte corrette ed errate rispetto ad item che si trovano all'origine della scala, cioè che hanno valore 0 (cfr. paragrafo 2.1).

<sup>88</sup> Nelle indagini internazionali in genere i punteggi derivati dal trattamento IRT dei punteggi vengono normalizzati su una scala basata su media = 500 e deviazione standard = 100, in pratica si tratta della scala punti T moltiplicata per 10 (OECD 2005; Campbell J., Kelly D., Mullis I., Martin M., Sainsbury M. 2001).

loro trattamento successivo con SPSS<sup>89</sup> per l'elaborazione delle statistiche descrittive sconta la somma dei piccoli arrotondamenti precedenti.

Si notano anche i valori massimi dei punteggi IRT, che risultano molto più distanti dalla media rispetto ai valori minimi. Nessuno studente ha raggiunto il punteggio massimo teorico (30 risposte esatte) e solo due hanno sbagliato una sola risposta (ma non la stessa, ottenendo quindi diversi punteggi per le analisi IRT).

Come già accennato in precedenza si è proceduto alla trasformazione in punti T di tutti i punteggi, riepilogati nella Tabella 5.33

Tabella 5.33 – Statistiche descrittive prova TCL (punti T)

	Classica	IRT	IRT Om
Mediana	48,75	48,71	49,06
Media	50,00 (0,49)	50,00 (0,49)	50,00 (0,49)
Dev std.	10,00	10,00	10,00
Min	25,78	23,49	22,97
Max	77,98	94,89	93,86
Asimmetria	0,25 (0,12)	0,78 (0,12)	0,66 (0,12)
Curtosi	-0,37 (0,24)	1,89 (0,24)	1,45 (0,24)

Fra parentesi l'errore standard relativo.

Confrontando le statistiche descrittive secondo le tre diverse modalità di analisi degli item si può notare una sostanziale diversità fra le gamme dei punteggi, che risultano più ampie per i due modelli IRT (in entrambi i casi >71) rispetto al modello classico (52). Bisogna considerare che la gamma non normalizzata del modello classico era di 26 punti con una deviazione standard di 4,80<sup>90</sup> (un rapporto quindi di 5,4/1) mentre la gamma IRT era di 8,11 logit con deviazione standard 1,13 (rapporto 7,2/1) e quella IRT Om era di 8,63 logit con deviazione standard 1,22 (rapporto 7,1/1). Per considerare il diverso andamento dei punteggi è opportuno confrontare i grafici relativi alle tre distribuzioni, presentati nelle Figure 5.4-5-6 della pagina a fronte.

<sup>89</sup> Per l'analisi statistica delle graduatorie è stata utilizzata la versione 13.0 del programma prodotto da SPSS, in dotazione al Dipartimento di Ricerche Storico-filosofiche e Pedagogiche dell'Università di Roma "La Sapienza".

<sup>90</sup> La deviazione standard costituisce l'unità di misura base per la normalizzazione dei punteggi, per cui il rapporto con la gamma permette di stimare l'estensione della scala.



Figura 5.4 – Distribuzione punteggi TCL (Classica)

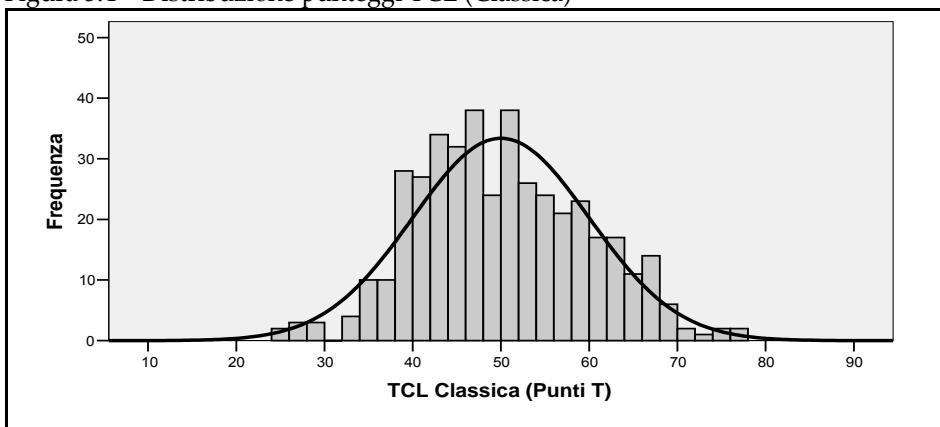


Figura 5.5 – Distribuzione punteggi TCL (IRT)

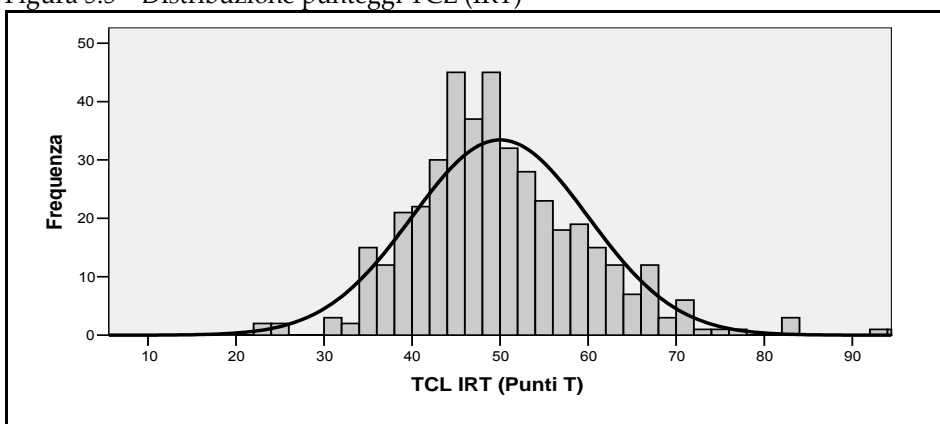
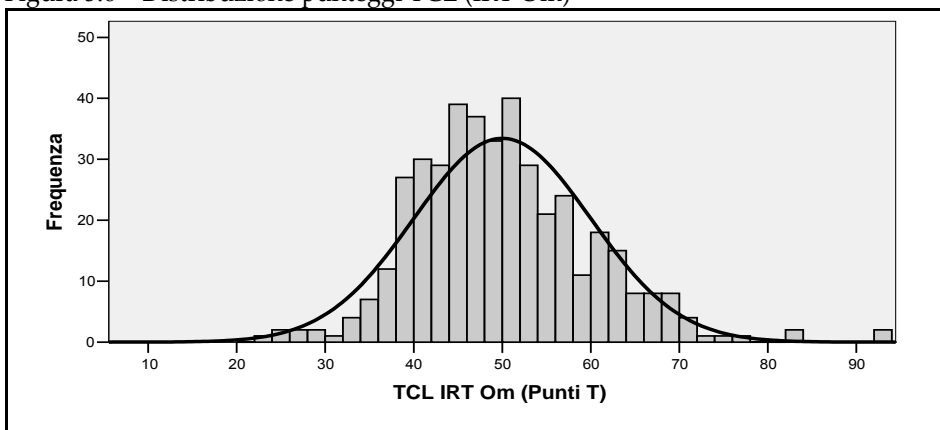


Figura 5.6 – Distribuzione punteggi TCL (IRT Om)



Per agevolare il confronto i tre grafici mantengono per entrambi gli assi le stesse scale e le stesse aggregazioni di punteggio (ogni barra rappresenta due punti).

Tutte e tre le curve presentano una chiara asimmetria positiva, con i punteggi inferiori alla media che scavalcano spesso, e in modo accentuato, la linea relativa alla distribuzione normale, confermando la precedente osservazione sulla posizione reciproca di media e mediana. L'asimmetria positiva dei punteggi IRT risulta però più spiccata, per cui l'estensione della scala dei punteggi riguarda soprattutto la parte destra della distribuzione, confermando anche in questo caso la prima osservazione riguardante i valori massimi, decisamente più marcati rispetto ai valori minimi.

La maggiore articolazione delle scale IRT è confermata anche dai valori di curtosi, che risultano saldamente positivi, a fronte di un valore negativo rilevato per i punteggi ottenuti dall'item analisi classica<sup>91</sup>. La prima curva (curtosi negativa) risulta leggermente platicurtica, cioè appiattita nella parte centrale, mentre le due distribuzioni IRT (curtosi positive) risultano fortemente leptocurtiche, cioè più alte nella parte centrale e, nel caso specifico con frequenze accertate nella parte destra della distribuzione.

Da una parte quindi i modelli IRT rivelano qualche problema di distribuzione dei punteggi in relazione al marcato sviluppo verticale della curva, ma dall'altra si ottengono punteggi che sfruttano meglio la scala grazie a una maggiore estensione agli estremi, potendo contare su punteggi pesati dei singoli item e di conseguenza su un maggior numero di combinazioni possibili di punteggi<sup>92</sup>, laddove il modello classico è legato alla scala discreta costruita sulla base del valore standard assegnato a ogni item.

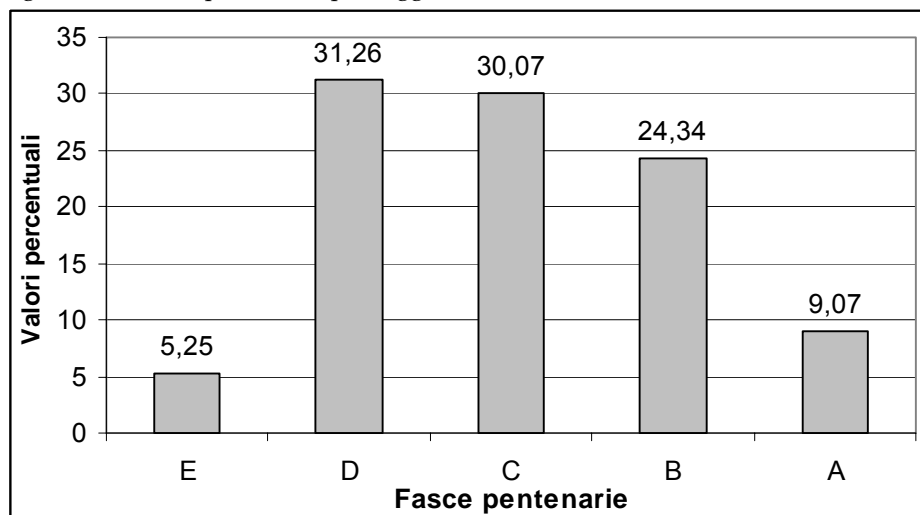
Proprio le differenze esistenti nelle statistiche di controllo della normalità di distribuzione suggeriscono di procedere a una ulteriore articolazione dei punteggi basata sulla suddivisione in fasce pentenarie, che permette di dare un primo significato ai punteggi identificando alcune

<sup>91</sup> I valori positivi della curtosi si associano a una maggiore frequenza dei punteggi nelle code della distribuzione, mentre valori positivi della statistica denotano un aumento dei punteggi intorno alla media rispetto alla distribuzione normale.

<sup>92</sup> La graduatoria ottenuta con i punteggi IRT per la prova TCL prevede 415 valori diversi su 419 rispondenti, per cui sono solo 4 i punteggi che si ripetono per più di un soggetto. Per la scala IRT Om i valori diversi sono 412, con solo 7 punteggi che si ripetono.

soglie significative<sup>93</sup>. In questo caso risulta chiara una rappresentazione grafica attraverso una serie di grafici a barre, su cui è opportuno specificare le percentuali relative a ogni fascia. Il primo grafico (Figura 5.7) riguarda l'articolazione in fasce pentenarie dei punteggi secondo il modello di item analisi classica.

Figura 5.7 – Fasce pentenarie punteggi TCL (Classica)



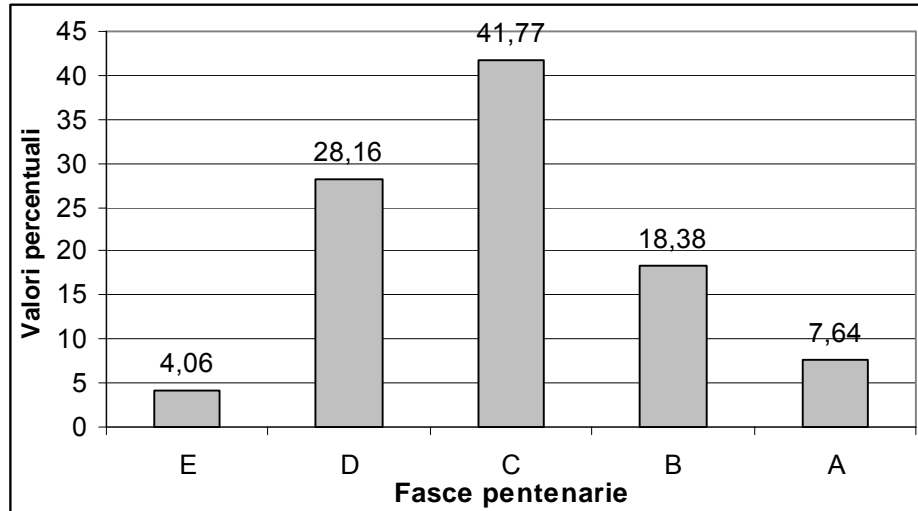
La distribuzione per fasce pentenarie risulta anomala, in quanto i punteggi della fascia D risultano più rappresentati di quelli della fascia centrale, nettamente più bassa rispetto alla normale distribuzione percentuale per fasce. L'asimmetria positiva quindi si distribuisce in modo irregolare fra le due fasce inferiori alla media, la cui somma comunque risulta più alta della somma delle fasce A e B. Ciò è il frutto della particolare distribuzione dei punteggi derivata da una scala discreta, per cui

<sup>93</sup> E' stato utilizzato il modello tradizionale di suddivisione in fasce pentenarie, che prevede l'incardinamento della fascia centrale (C) sul valore di media e l'ampiezza di ogni fascia corrispondente a una deviazione standard. E' possibile considerare diverse ampiezze della fascia (da 1,33 DevStd a 0,58 DevStd), ma il modello tradizionale risulta essere quello meno infedele rispetto alla distribuzione dei punteggi (Gattullo e Giovannini 1989).

le soglie previste per il passaggio da una fascia all'altra possono creare questo tipo di distorsione<sup>94</sup>.

Vediamo ora (Figura 5.8) come si distribuiscono nelle cinque fasce i punteggi derivati dall'analisi IRT tradizionale.

Figura 5.8 – Fasce pentenarie punteggi TCL (IRT)



Considerando la maggiore articolazione dei punteggi IRT la distribuzione nelle cinque fasce risulta molto più vicina a una distribuzione di tipo normale, anche se permane uno squilibrio a favore delle fasce inferiori (soprattutto la fascia D)<sup>95</sup>.

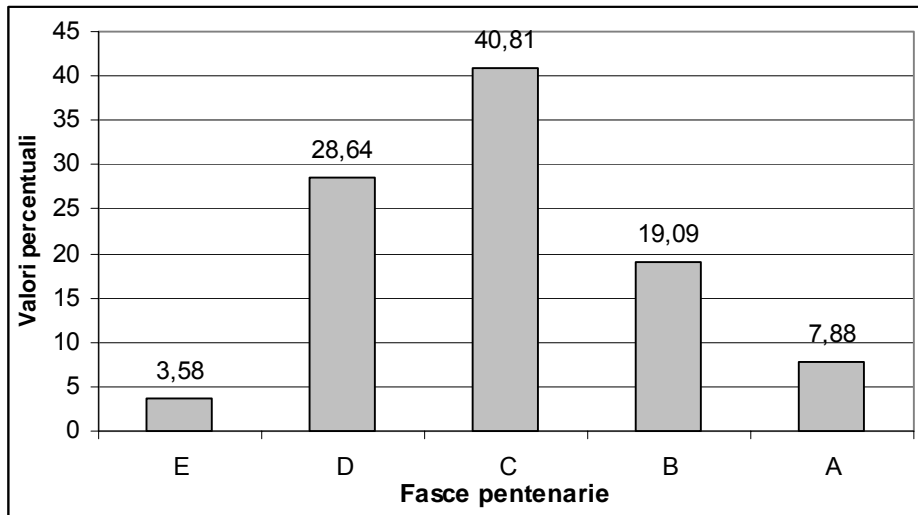
Un andamento molto simile si può osservare nel grafico a barre (Figura 5.9) relativo alle fasce centenarie calcolate secondo il modello IRT

<sup>94</sup> Nel caso specifico i punteggi minimo e massimo della fascia D (36,2 e 44,6 sulla scala normalizzata) risultano molto più vicini ai valori soglia (35 e 45) rispetto agli stessi due valori della fascia C (46,7 e 52,9, con valori soglia 45 e 55). Allo stesso modo minimo e massimo della fascia B (rispettivamente 55,01 e 63,4) sono molto vicini ai valori soglia specifici (55 e 65), in particolare per il valore minimo.

<sup>95</sup> Le percentuali di riferimento per una distribuzione normale articolata nelle fasce pentenarie, considerando il modello utilizzato, sono le seguenti: 38% per la fascia centrale, 24% per le fasce intermedie (B e D), 7% per le fasce estreme (A e E) (Gattullo e Giovannini 1989).

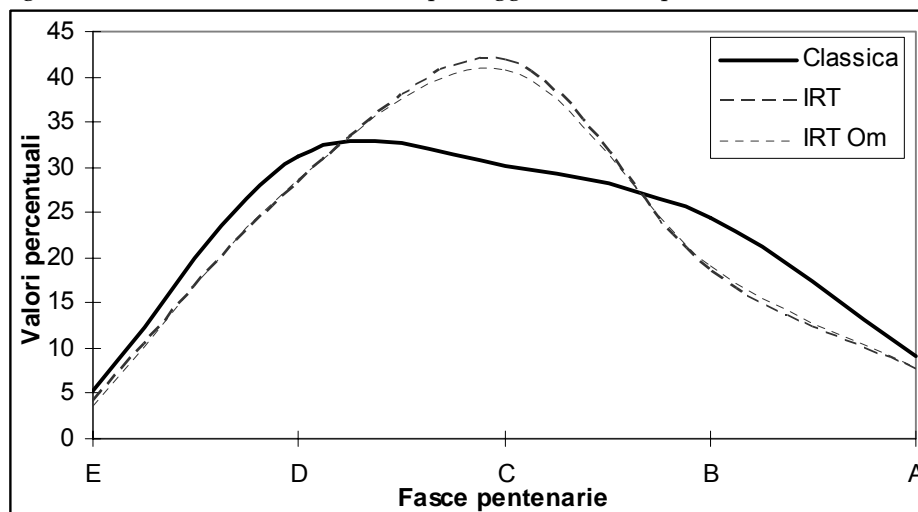
Om, con un contenuto riequilibrio delle fasce A e B, che recuperano un punto percentuale dalla fascia centrale della distribuzione.

Figura 5.9 – Fasce pentenarie punteggi TCL (IRT Om)



Per concludere il confronto fra le tre diverse distribuzioni in fasce pentenarie riuniamo i valori dei tre grafici precedenti in un grafico riassuntivo a linee (Figura 5.10) che presenta il diverso andamento delle fasce dei punteggi per i tre diversi modelli di analisi.

Figura 5.10 – Confronto distribuzione punteggi TCL (fasce pentenarie)



Si può notare che le due curve disegnate sulle distribuzioni dei punteggi di tipo IRT, ma articolate in fasce pentenarie, risultano molto simili e più vicine alla normalità, con una marcata introflessione in corrispondenza della fascia B. Di contro la curva dei punteggi di tipo classico risulta molto schiacciata al centro e più alta in corrispondenze delle due fasce intermedie (D e B).

Questo comporta una tendenziale maggiore pendenza delle curve IRT (a conferma della maggiore articolazione della scala) che sicuramente favorisce il riscontro di differenze più significative nella graduatoria dei punteggi.

### 5.5.2. Le graduatorie della prova di comprensione Cloze

Come abbiamo già fatto per la prova di comprensione TCL, ecco un quadro riassuntivo delle principali statistiche descrittive relative alla prova di comprensione Cloze, sempre considerando i punteggi definiti dai tre tipi di analisi. Ricordiamo che, come specificato nel Paragrafo 4.4, l'analisi si riferisce al solo testo antropologico, composto complessivamente da 31 item.

Tabella 5.34 – Statistiche descrittive prova Cloze

	Classica	IRT	IRT Om
Mediana	20	0,00	0,01
Media	19,92 (0,19)	-0,02 (0,05)	-0,03 (0,05)
Dev std.	3,87	1,10	1,11
Min*	1	-4,20	-5,74
Max	28	3,24	3,19
Asimmetria	-0,95 (0,12)	-0,18 (0,12)	-0,44 (0,12)
Curtosi	1,92 (0,24)	0,37 (0,24)	1,59 (0,24)

Fra parentesi l'errore standard relativo.

\* Risente di un caso anomalo (studente con una sola risposta corretta). Escludendo tale studente i valori minimi salgono rispettivamente a 5 (Classica), -3,23 (IRT) e -3,65 (IRT Om).

A differenza della prova TCL il valore della mediana, per tutte e tre i tipi di analisi, risulta maggiore rispetto alla media, quindi l'asimmetria di distribuzione dei punteggi diventa sempre negativa, in misura molto accentuata per la distribuzione dei punteggi ottenuti con l'analisi classica, ma risulta anomala anche l'asimmetria della distribuzione IRT Om (il valore è superiore al doppio dell'errore standard relativo). Considerando l'esclusione dell'unico punteggio anomalo segnalato nella nota della Tabella 5.34, la gamma dei punteggi più alta, rapportata con la deviazione standard, è quella dei punteggi da item analisi classica (6,20/1), ma in questo caso non si riscontrano le stesse differenze viste per la prova TCL, infatti la gamma dei punteggi IRT Om (6,16/1) e IRT standard (5,88/1) risultano molto vicine a quella di tipo classico. La relativa difficoltà della prova è attestata dal fatto che non solo nessun rispondente ha raggiunto il punteggio massimo, ma l'unico migliore si è fermato a sole 28 risposte corrette su 31.

Anche per la prova Cloze la tabella delle statistiche descrittive è stata rielaborata considerando i punteggi normalizzati per tutte e tre le scale di misura.

Tabella 5.35 – Statistiche descrittive prova Cloze (punti T)

	Classica	IRT	IRT Om
Mediana	50,22	50,16	50,17
Media	50,00 (0,49)	50,00 (0,49)	50,00 (0,49)
Dev std.	10,00	10,00	10,00
Min*	1,09	11,90	-1,52
Max	70,90	79,71	79,05
Asimmetria	-0,94 (0,12)	-0,18 (0,12)	-0,44 (0,12)
Curtosi	1,92 (0,24)	0,37 (0,24)	1,59 (0,24)

Fra parentesi l'errore standard relativo.

\* Risente di un caso anomalo (studente con una sola risposta corretta). Escludendo tale studente i valori minimi salgono rispettivamente a 11,43 (Classica), 20,78 (IRT) e 17,37 (IRT Om).

Da notare i diversi valori di curtosi per le tre distribuzioni: risulta marcatamente positiva per i punteggi da item analisi classica, ma molto vicina al valore rilevato per la distribuzione IRT Om. Una curtosi di questo tipo denota una curva che cresce nella parte centrale (leptocurtica), cioè l'esatto contrario di quanto rilevato per i punteggi da item analisi classica della prova di comprensione TCL. La distribuzione del Cloze presenta quindi maggiori problemi di normalità, in particolare per i punteggi della classica.

Un confronto grafico delle tre distribuzioni illustra chiaramente i diversi andamenti.



Figura 5.11 – Distribuzione punteggi Cloze (Classica)

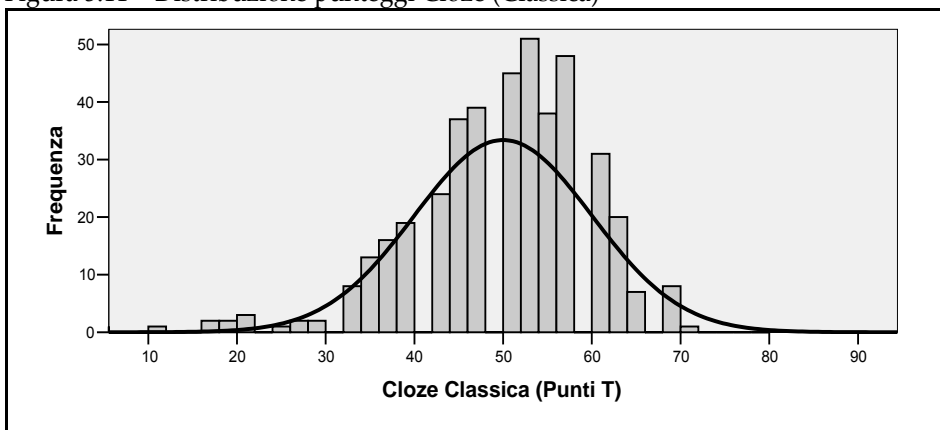


Figura 5.12 – Distribuzione punteggi Cloze (IRT)

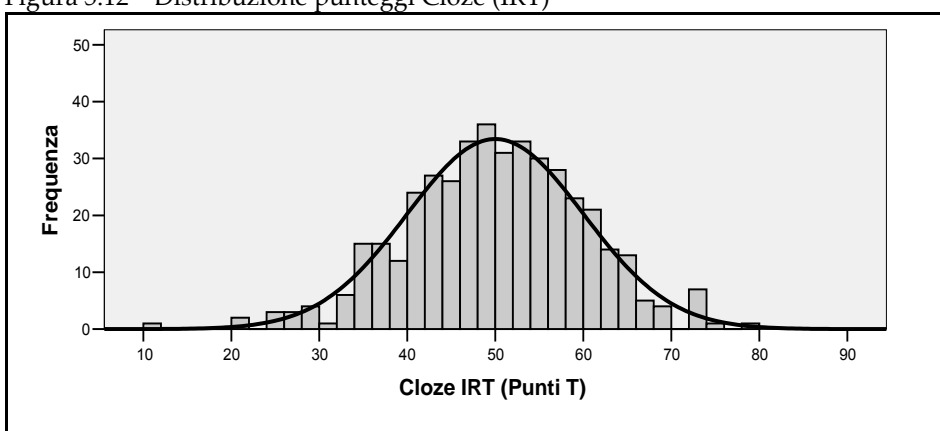
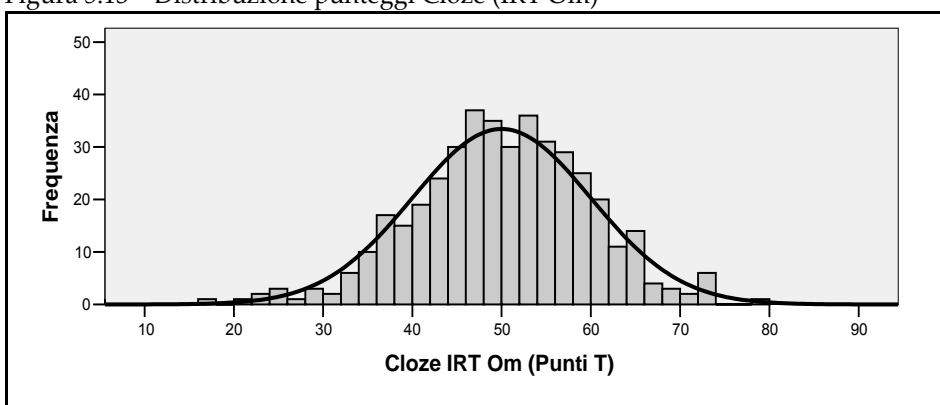


Figura 5.13 – Distribuzione punteggi Cloze (IRT Om)

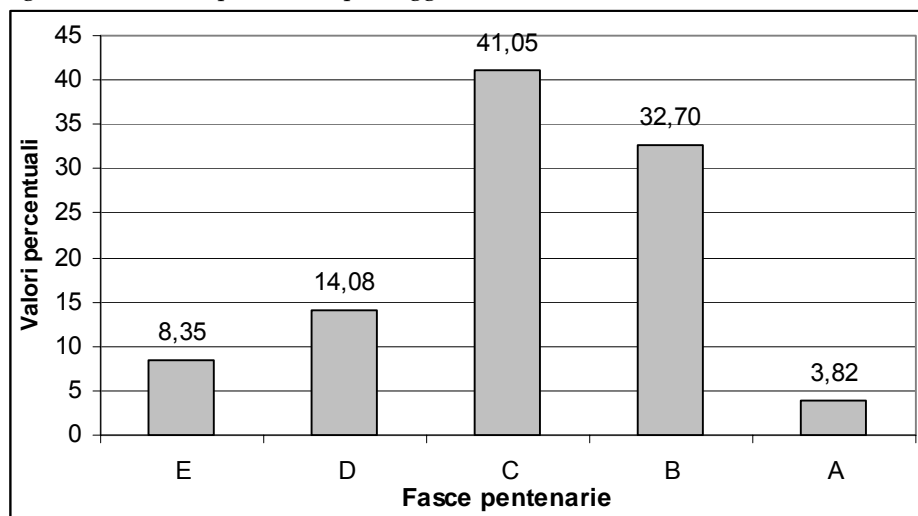


Anche in questo caso sono state mantenute per tutti e tre i grafici le scale relative ai due assi, che risultano anche uguali a quelle dei tre grafici simili dedicate ai punteggi della prova TCL (Figure 5.4-5-6). Sono molto evidenti le anomalie di distribuzione dei punteggi da item analisi classica, in parte spiegate dai buchi di distribuzione derivati dalla presenza di punteggi scalati unicamente secondo il numero dei termini da ricostruire. Le due distribuzioni IRT risultano invece molto più fluide e aderenti alla curva normale.

Pur considerando l'esclusione dell'unico punteggio anomalo segnalato nella nota della Tabella 5.35 è evidente che la graduatoria dei punteggi da item analisi classica si allunga maggiormente verso la coda negativa, mentre le due distribuzioni IRT si articolano maggiormente verso i punteggi più alti. Ciò è dovuto alla presenza di alcuni item difficili nella prova Cloze, che pesano in misura maggiore sulla determinazione del punteggio e quindi premiano gli studenti che li hanno risolti positivamente, allungando la distribuzione verso destra.

Come già fatto per la prova di comprensione TCL si procede alla determinazione delle fasce pentenarie per i risultati della prova Cloze. Il primo grafico a barre, Figura 5.14, è relativo ai punteggi ottenuti dall'item analisi classica.

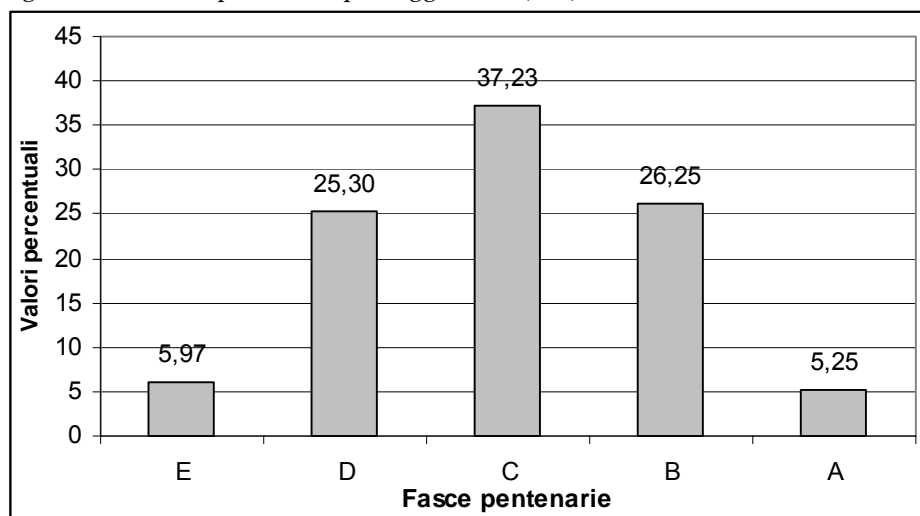
Figura 5.14 – Fasce pentenarie punteggi Cloze (Classica)



La relativa mancanza di normalità della curva di distribuzione dei punteggi, in particolare per l'asimmetria negativa, si riflette evidentemente anche sull'articolazione in fasce, con un marcato squilibrio fra le fasce intermedie, accentuato dal fenomeno già segnalato nella distribuzione relativa alla prova TCL<sup>96</sup>. Da notare anche che la prevalenza di punteggi superiori alla media comporta però la scarsa consistenza dei punteggi che si trovano nella fascia di eccellenza, praticamente dimezzati rispetto a una distribuzione normale.

Vediamo ora la stessa distribuzione in fasce, ma relativa ai punteggi ottenuti dall'item analisi IRT standard.

Figura 5.15 – Fasce pentenarie punteggi Cloze (IRT)



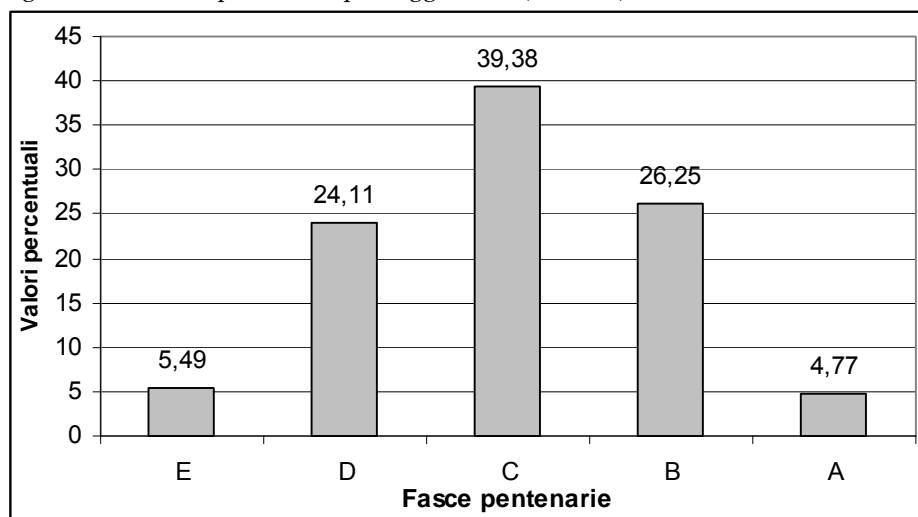
In questo caso l'equilibrio fra le fasce è molto vicino al modello di riferimento, con una leggera contrazione delle percentuali delle fasce estreme (A e E) a favore delle fasce intermedie (B e D), mentre la fascia centrale sfiora il valore normale (38%). Anche in questo caso la particola-

<sup>96</sup> In questo caso sono le fasce B e C a presentare una gamma maggiore di punteggi considerando i valori minimi e massimi reali interni alle fasce rispetto ai valori soglia: la gamma di queste due fasce è di 7,8 punti normalizzati (su 10 di ampiezza di fascia), mentre la fascia D presenta una gamma ristretta (5,2), inoltre i valori minimi delle fasce B e C si trovano molto vicini (55,4 e 45,01) ai valori di soglia inferiori.

re conformazione dei punteggi ottenuti grazie all'analisi IRT favorisce una distribuzione più omogenea all'interno delle fasce<sup>97</sup>.

Proprio per questa articolazione dei punteggi anche la distribuzione pentenaria relativa all'analisi IRT Om risulta molto vicina alle percentuali di normalità (Figura 5.16), con una leggera crescita della fascia centrale C e una maggiore contrazione delle due fasce inferiori alla media (D e E), confermando anche un andamento leggermente leptocurtico della curva, rivelata dal valore di curtosi.

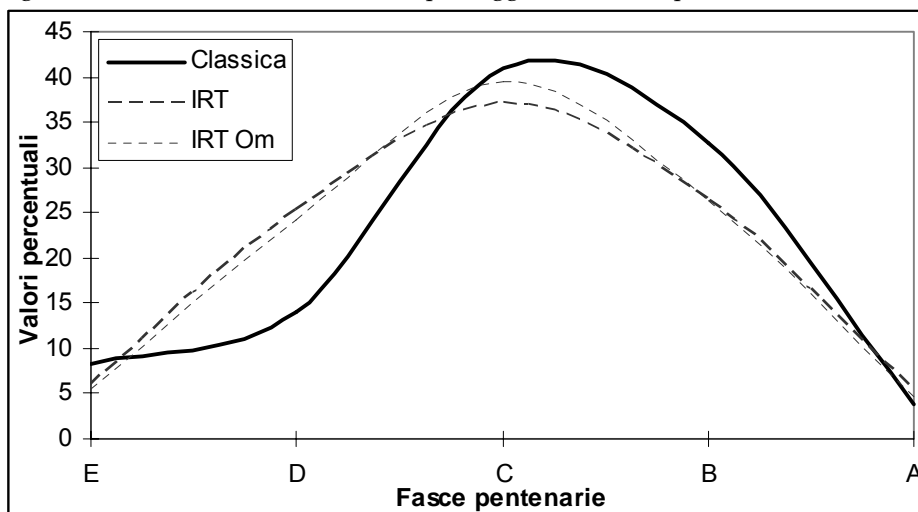
Figura 5.16 – Fasce pentenarie punteggi Cloze (IRT Om)



Anche per quanto riguarda la prova di comprensione Cloze è utile riunire le tre curve di distribuzione per fasce in un unico grafico a linee (Figura 5.17) per un confronto immediato delle tre distribuzioni.

<sup>97</sup> Sia per la distribuzione di punteggi IRT, sia per quella IRT Om, i punteggi effettivi rilevati sono 411, con solo cinque punteggi ottenuti da due studenti e un punteggio da tre studenti.

Figura 5.17 – Confronto distribuzione punteggi Cloze (fasce pentenarie)



La distribuzione dei punteggi da item analisi classica presenta una marcata introflessione nella parte sinistra e una doppia pendenza nella parte destra; inoltre si presenta nettamente meno normale rispetto alle due curve relative alle distribuzioni di punteggi IRT. Questi andamenti comportano un maggiore affollamento di punteggi, nella parte medio-alta della graduatoria basata sull'item analisi classica, con evidenti ripercussioni sulla definizione di soglie necessarie alla funzione selettiva della prova, mentre l'andamento dei punteggi IRT presenta una maggiore fluidità, con conseguente maggiore facilità di definizione delle soglie.

### 5.5.3. Le relazioni fra le prove di comprensione TCL e Cloze

Fino a questo punto abbiamo considerato separatamente le graduatorie provenienti dalle due prove di comprensione, verificandone il diverso andamento, soprattutto per i punteggi derivati dall'item analisi classica. Nel Paragrafo 4.2 sono stati segnalati i punti di contatto tra le due diverse forme di verifica della comprensione della lettura, per cui risulta utile verificare se anche i punteggi degli studenti che hanno partecipato a queste due prove risultano correlati, anche per verificare l'opportunità di sommare i due punteggi per arrivare a una graduatoria unica che tenga conto di entrambe le prove.

In questa fase vengono messe in relazione graduatorie omogenee, realizzate cioè con la stessa modalità di analisi. Nella Tabella 5.36 è evidenziata la correlazione esistente fra le graduatorie ottenute con i punteggi da item analisi classica, relativi a tutti gli studenti che hanno partecipato alla prova.

Tabella 5.36 – Correlazione fra prove TCL e Cloze (Classica – 561 rispondenti)

		TCL	Cloze
TCL	Correlazione di Pearson	1	,465(**)
	Sig. (2-code)		,000
	N	561	561
Cloze	Correlazione di Pearson	,465(**)	1
	Sig. (2-code)	,000	
	N	561	561

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

La correlazione fra le due prove di comprensione, pur risultando inferiore a quelle segnalate da Lucisano (1989) risultano comunque significative, confermando il sostanziale legame esistente fra le competenze indagate dalle due prove. E' interessante verificare se tale correlazione si conserva considerando il solo gruppo di rispondenti romani, per il quale saranno confrontate le graduatorie.

Tabella 5.37 – Correlazione fra prove TCL e Cloze (Classica – 419 rispondenti)

		TCL	Cloze
TCL	Correlazione di Pearson	1	,447(**)
	Sig. (2-code)		,000
	N	419	419
Cloze	Correlazione di Pearson	,447(**)	1
	Sig. (2-code)	,000	
	N	419	419

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

Non emergono sostanziali differenze se si riduce il confronto fra le due prove al solo gruppo di studenti di Roma, per cui le successive correlazioni saranno verificate considerando solo questo gruppo.

Vediamo ora il livello di correlazione raggiunto dalle graduatorie costruite con le due analisi IRT. Nella Tabella 5.38 è riportato il dato relativo alle due analisi IRT tradizionali, mentre nella Tabella 5.39 il confronto

riguarda le due analisi IRT che penalizzano le omissioni in maniera indistinta.

Tabella 5.38 – Correlazione fra prove TCL e Cloze (IRT – 419 rispondenti)

		TCL	Cloze
TCL	Correlazione di Pearson	1	,456(**)
	Sig. (2-code)		,000
	N	419	419
Cloze	Correlazione di Pearson	,456(**)	1
	Sig. (2-code)	,000	
	N	419	419

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

Tabella 5.39 – Correlazione fra prove TCL e Cloze (IRT Om – 419 rispondenti)

		TCL	Cloze
TCL	Correlazione di Pearson	1	,459(**)
	Sig. (2-code)		,000
	N	419	419
Cloze	Correlazione di Pearson	,459(**)	1
	Sig. (2-code)	,000	
	N	419	419

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

Le correlazioni risultano stabili con tutti e tre i tipi di analisi svolte, ma il valore non esaltante della misura di correlazione (sempre inferiore a 0,5), e l'andamento molto diverso delle curve di distribuzione dei punteggi relative alle due prove (cfr. Figure 5.10 e 5.17), spinge a considerare con molta prudenza la fusione delle due graduatorie, anche se la correlazione dei due punteggi sommati con i due punteggi distinti risulta molto alta. Nella Tabella 5.40 si riporta, a titolo esemplificativo, la correlazione del punteggio complessivo secondo l'analisi classica (TCL + Cloze) con il solo punteggio TCL (il dato risulta naturalmente uguale anche per il Cloze).

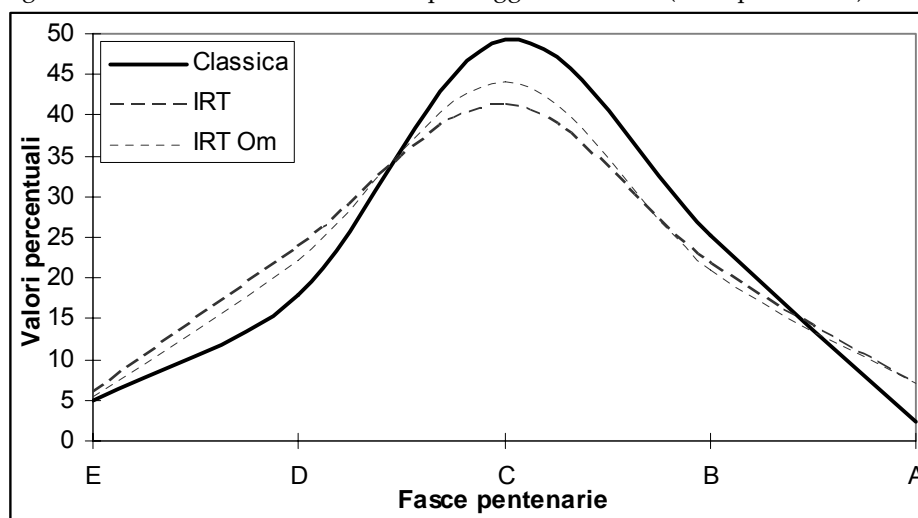
Tabella 5.40 – Correlazione fra TCL+Cloze e solo TCL (Classica – 419 rispondenti)

		TCL+Cloze	TCL
TCL+Cloze	Correlazione di Pearson	1	,851(**)
	Sig. (2-code)		,000
	N	419	419
TCL	Correlazione di Pearson	,851(**)	1
	Sig. (2-code)	,000	
	N	419	419

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

Per completare il confronto è utile riportare il grafico comparativo delle tre curve di distribuzione delle fasce pentenarie definite dalle tre modalità di analisi, osservando l'accentuato picco centrale delle tre curve (in particolare per i punteggi da item analisi classica), che hanno costretto a portare il valore massimo della scala relativa ai valori percentuali di frequenza a 50, rispetto ai precedenti grafici (cfr. Figure 5.10 e 5.17) in cui tale valore era 45.

Figura 5.18 – Confronto distribuzione punteggi TCL+Cloze (fasce pentenarie)



In definitiva proprio la buona correlazione accertata fra la graduatoria sommata (TCL + Cloze) e quelle delle singole prove permette di considerare accettabile l'approfondimento dell'analisi dando priorità alle graduatorie relative alla sola prova di comprensione TCL, che presenta



problemi molto minori, sia nell'analisi degli item, sia nella consistenza interna, rispetto alla prova Cloze che, ricordiamo, considera solo uno dei quattro testi che componevano il pacchetto originale, per i già segnalati problemi di validità relativi ai risultati degli altri tre testi.

#### 5.5.4. *Le relazioni fra le diverse graduatorie prodotte dalle analisi degli item*

Prima di passare al confronto delle diverse graduatorie relative alle prove di comprensione, verifichiamo le relazioni esistenti fra i diversi modelli di item analisi.

Ricordiamo che anche in questo caso, considerato che il confronto delle graduatorie riguarderà solo gli studenti del gruppo romano, verranno considerati solo i punteggi di questo gruppo.

Tabella 5.41 – Correlazione fra le graduatorie Classica, IRT e IRT Om. (TCL – 419 rispondenti)

		Classica	IRT	IRT Om
Classica	Correlazione di Pearson	1	,892(**)	,983(**)
	Sig. (2-code)		,000	,000
	N	419	419	419
IRT	Correlazione di Pearson	,892(**)	1	,920(**)
	Sig. (2-code)	,000		,000
	N	419	419	419
IRT Om	Correlazione di Pearson	,983(**)	,920(**)	1
	Sig. (2-code)	,000	,000	
	N	419	419	419

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

I valori relativi alle correlazioni esistenti fra le graduatorie prodotte dai tre diversi modelli di analisi risultano sempre significativi e rivelano una quasi linearità di relazione, suggerendo la possibile interscambiabilità delle tre graduatorie. In particolare molto forte risulta essere il legame esistente fra l'item analisi classica e quella IRT che penalizza tutte le omissioni. Nel complesso è proprio questa ultima modalità quella che risulta essere più correlata con le altre.

Dal quadro di correlazioni che riguarda la prova di comprensione Cloze (Tabella 5.42) emerge un legame ancora più forte tra i diversi modelli di item analisi, con valori che sfiorano quasi la correlazione perfetta.

Tabella 5.42 – Correlazione fra le graduatorie Classica, IRT e IRT Om. (Cloze – 419 rispondenti)

		Classica	IRT	IRT Om
Classica	Correlazione di Pearson	1	,980(**)	,987(**)
	Sig. (2-code)		,000	,000
	N	419	419	419
IRT	Correlazione di Pearson	,980(**)	1	,995(**)
	Sig. (2-code)	,000		,000
	N	419	419	419
IRT Om	Correlazione di Pearson	,987(**)	,995(**)	1
	Sig. (2-code)	,000	,000	
	N	419	419	419

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

In questo caso il valore più elevato si riscontra tra le due analisi IRT, ma con differenze limitate al secondo numero decimale.

E' possibile pertanto definire una prima conclusione utile per la verifica dell'ipotesi: le graduatorie prodotte dall'item analisi IRT (anche nella variante IRT Om) risultano sostanzialmente equivalenti a quelle prodotte attribuendo a ogni risposta lo stesso valore, secondo il modello dell'item analisi classica.

Consideriamo ora la correlazione dei diversi modelli di item analisi con un criterio oggettivo esterno. Nel primo capitolo, analizzando i criteri che accompagnano la formazione delle graduatorie per l'accesso ai corsi universitari, era stato evidenziato il rilievo dato al voto nell'Esame di Stato. Pertanto risulta interessante controllare le correlazioni esistenti fra questo indicatore e le diverse graduatorie prodotte.

La Tabella 5.43 riepiloga tali correlazioni; il tradizionale formato incrociato della tabella di correlazione viene in questo caso ridotto a una sola colonna, per favorire la leggibilità e sintetizzare in un solo quadro il complesso di tutte le misure di correlazione, dando priorità alla variabile relativa al voto nell'Esame di Stato.

Tabella 5.43 – Correlazione fra il voto nell'Esame di Stato e tutte le graduatorie (419 rispondenti)

		Voto Esame
Voto Esame	Correlazione di Pearson Sig. (2-code) N	1 419
TCL Classica	Correlazione di Pearson Sig. (2-code) N	,291(**) ,000 419
TCL IRT	Correlazione di Pearson Sig. (2-code) N	,297(**) ,000 419
TCL IRT Om	Correlazione di Pearson Sig. (2-code) N	,303(**) ,000 419
Cloze Classica	Correlazione di Pearson Sig. (2-code) N	,297(**) ,000 419
Cloze IRT	Correlazione di Pearson Sig. (2-code) N	,309(**) ,000 419
Cloze IRT Om	Correlazione di Pearson Sig. (2-code) N	,309(**) ,000 419

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

I valori di correlazioni delle diverse graduatorie con il voto riportato nell'Esame di Stato risultano molto simili, tutte significative e intorno al valore di 0,30. Si tratta quindi di correlazioni non molto marcate, ma rispetto alle variabili rilevate in ingresso, il voto nell'esame risulta essere quella maggiormente collegata alla performance nella prova di ingresso.

Considerando che la maggior parte delle variabili rilevate nel questionario di ingresso è di tipo nominale, è possibile costruire tavole di contingenza per verificare la distribuzione dei valori attesi. Per garantire una consistente rappresentatività delle celle risulta opportuno incrociare il dato relativo alla suddivisione in fasce pentinarie, e non la graduato-

ria completa<sup>98</sup>. L'unico aspetto che sembra essere collegabile con l'andamento dei punteggi nella prova di ingresso è quello relativo al tipo di scuola secondaria superiore frequentata. Nella Tabella 5.44 sono riportati gli incroci relativi a questa variabile con la fascia di appartenenza della distribuzione pentenaria della graduatoria TCL Classica.

Tabella 5.44 – Tavola di contingenza fra scuola secondaria frequentata e fasce TCL Classica (risposte valide su 419 rispondenti)

Scuola secondaria frequentata		Fasce TCL Classica					Totale
		A	B	C	D	E	
Licei	Conteggio	32	70	88	69	11	270
	Atteso	24,8	65,2	82,2	83,5	14,3	270,0
Tecnici	Conteggio	3	24	27	49	7	110
	Atteso	10,1	26,6	33,5	34,0	5,8	110,0
Professionali	Conteggio	3	6	11	10	4	34
	Atteso	3,1	8,2	10,3	10,5	1,8	34,0
Totale	Conteggio	38	100	126	128	22	414
	Atteso	38,0	100,0	126,0	128,0	22,0	414,0

Sono state escluse le omissioni di risposta e le risposte non utilizzabili.

	Valore	df	Sig. asint. (2 vie)
Chi-quadrato di Pearson	22,813(a)	8	0,004
N. di casi validi	414		

(a) 2 celle (13,3%) hanno un conteggio atteso inferiore a 5. Il conteggio atteso minimo è 1,81.

La verifica del Chi-quadrato attesta la significatività di relazione fra le due variabili, mentre risulta meno marcata (solo  $<0,05$ ) la significatività rispetto alle graduatorie IRT, sempre per la prova TCL.

Naturalmente sarà importante verificare l'esistenza di correlazioni con le variabili rilevate in uscita dal primo anno di corso, ma in questa fase si può sottolineare che gli unici aspetti che sembrano in qualche modo legati ai risultati delle prove di ingresso riguardano la precedente esperienza scolastica dei nuovi iscritti.

<sup>98</sup> La misura di significatività per una tavola di contingenza maggiore della matrice 2x2 presuppone che meno del 20% delle celle che la compongono riportino un valore atteso minore di 5 casi (Bailey 1982).

## 5.6. Graduatorie di accesso secondo i modelli classico e IRT

La prova di ingresso utilizzata per la ricerca, pur essendo stata presentata agli studenti come prova di orientamento per migliorare l'offerta didattica del corso di laurea, ha permesso di scattare una fotografia del loro livello di competenza in comprensione della lettura prima di affrontare il percorso accademico. E' possibile pertanto considerare i risultati ottenuti anche come indicatori di accesso per una ipotetica prova di selezione<sup>99</sup>.

Per la definizione di una graduatoria bisogna preventivamente risolvere il problema dei punteggi *ex aequo*, in particolare per quelli derivati dall'item analisi classica, che come abbiamo visto si distribuiscono su valori grezzi interi su una gamma limitata. Nel primo capitolo abbiamo visto che questo problema viene risolto o con criteri interni alla prova (dando priorità al punteggio parziale di una parte del test) o con criteri esterni, come il voto riportato nell'Esame di Stato o l'età anagrafica.

Per la simulazione della graduatoria di accesso si è deciso di procedere in quest'ultimo modo, prevedendo, in caso di parità di punteggio nella prova, di considerare il voto riportato nell'Esame di Stato, e in caso di ulteriore parità premiare la minore anzianità anagrafica.

Un altro aspetto da considerare è la soglia fissata per l'accesso, che finisce per essere l'elemento determinante per gli effetti della prova, poiché di fatto divide il gruppo dei rispondenti fra chi può accedere al corso di studi e chi ne viene escluso. Proprio per questo, al di là della forte correlazione riscontrata nel precedente paragrafo tra le diverse graduatorie dei punteggi, è importante analizzare le differenze fra queste in relazione ai casi individuali, con lo scopo di indagare quali e quanti scarti abbiano prodotto una modifica nel gruppo che, nella simulazione, può accedere al corso di laurea.

<sup>99</sup> Naturalmente è possibile operare questa equivalenza ricordando che le motivazioni degli studenti cambiano sensibilmente nel momento in cui la prova assume effettivamente una funzione selettiva, modificando anche le tattiche di comportamento durante la prova, poiché lo scopo non è tanto dimostrare le proprie capacità, quanto precedere gli altri nella graduatoria (per la distinzione fra ingresso ed accesso Andreani Dentici, Amoretti 2000). Le differenze di atteggiamento e motivazione degli studenti sono state già segnalate nel Paragrafo 3.5, in relazione al trattamento delle omissioni, e nel successivo riscontro del loro numero relativamente alto accertato per la prova TCL (Tabelle 5.21-22).

Tabella 5.45 – Confronto graduatorie TCL - Fascia alta dei punteggi (Classica)

ID	Grezzi	Classica PT	Graduatoria classica	IRT PT	Graduatoria IRT	IRTom PT	Graduatoria IRTom
89	29	77,98	1	93,62	2	93,03	2
60	29	77,98	2	94,89	1	93,86	1
509	28	75,89	3	83,94	3	82,29	3
88	28	75,89	4	83,61	5	82,12	4
616	27	73,80	5	77,06	6	76,95	5
410	26	71,71	6	74,34	7	74,02	6
736	26	71,71	7	73,83	8	73,80	7
4	25	69,63	8	71,55	9	71,71	8
868	25	69,63	9	70,36	14	69,91	13
235	25	69,63	10	70,48	12	70,63	9
639	25	69,63	11	70,61	11	70,29	10
330	25	69,63	12	70,46	13	70,29	11
770	25	69,63	13	70,71	10	69,68	14
216	24	67,54	14	67,15	22	69,36	16
531	24	67,54	15	66,94	24	67,01	25
317	24	67,54	16	67,03	23	69,37	15
703	24	67,54	17	66,52	29	66,84	27
820	24	67,54	18	67,53	18	67,97	20
547	24	67,54	19	67,25	20	69,14	17
596	24	67,54	20	67,25	21	69,14	18
242	24	67,54	21	66,64	26	66,92	26
526	24	67,54	22	66,59	27	68,87	19
524	24	67,54	23	66,87	25	67,11	24
94	24	67,54	24	68,03	17	69,98	12
654	24	67,54	25	67,45	19	67,72	21
409	24	67,54	26	65,77	31	65,28	31
636	24	67,54	27	66,59	28	67,22	22
641	23	65,45	28	68,21	16	67,12	23
612	23	65,45	29	64,56	33	65,60	29
412	23	65,45	30	63,61	39	64,68	34
830	23	65,45	31	63,52	41	63,48	37
750	23	65,45	32	64,01	36	65,13	32
502	23	65,45	33	62,38	45	62,17	50
618	23	65,45	34	63,69	37	64,53	35
50	23	65,45	35	64,27	34	65,73	28
529	23	65,45	36	64,11	35	65,01	33
512	23	65,45	37	65,24	32	65,39	30
515	23	65,45	38	62,80	43	62,89	42

Sono state evidenziate in grigio le discrepanze più rilevanti fra le graduatorie.

E' stata assunta la soglia effettivamente usata nel successivo anno accademico per il Corso di laurea in Scienze dell'Educazione e della Formazione dell'Università di Roma "La Sapienza", che prevedeva 150 posti per la laurea triennale.

Nella Tabella 5.45 sono messe a confronto le graduatorie relative alla prova TCL<sup>100</sup> secondo le tre diverse modalità di analisi. È stata considerata la fascia alta della graduatoria, identificata in questo caso dai 38 rispondenti che si trovano nella fascia pentenaria A secondo i punteggi derivati dall'item analisi classica. Di conseguenza questa tabella è ordinata secondo la graduatoria Classica. Per ogni graduatoria è riportato anche il punteggio (in punti T) relativo. A sinistra, nelle prime due colonne, è possibile osservare il codice identificativo dei rispondenti (ID) e il punteggio grezzo (Grezzi), cioè il numero di risposte corrette. In grigio sono segnalati gli studenti che, nelle graduatorie IRT, escono dal gruppo dei primi 38.

Notiamo subito l'alto numero di *ex aequo* presenti nella graduatoria Classica, risolti grazie ai criteri esterni alla prova, come accennato sopra. Per la graduatoria IRT sono presenti solo due casi di *ex aequo* (ID 636 e 526, ID 596 e 547), come pure per la graduatoria IRT Om (ID 641 e 636, ID 596 e 547). Notiamo di passaggio che quest'ultima coppia e rappresenta l'unico caso di assoluta equivalenza nel test, con le stesse risposte esatte ed errate, e identico dato relativo a omissioni e domande non raggiunte. E' abbastanza agevole rilevare che le graduatorie risultano molto simili: le prime 10 posizioni vedono praticamente gli stessi studenti (ma si nota la mancanza dello studente in quarta posizione della graduatoria IRT) e nel complesso dei primi 38 studenti identificati dalla graduatoria Classica, 34 trovano posto anche nella graduatoria IRT e 35 nella graduatoria IRT Om. Ci sono alcuni spostamenti significativi come lo studente 641 che guadagna 12 posizioni (da 28° a 16°) nella graduatoria IRT o lo studente 502 che ne perde 17 (da 33° a 50°) nella graduatoria IRT Om, ma sostanzialmente il gruppo dei migliori è lo stesso in tutte le graduatorie.

Per verificare la posizione dei pochi studenti mancanti nelle graduatorie IRT dei migliori, riportiamo la Tabella 5.46, costruita stavolta con-

<sup>100</sup> Nell'economia del presente lavoro di ricerca l'analisi delle graduatorie è limitata alla sola prova TCL, da una parte considerando la migliore qualità complessiva della prova e il buon livello di correlazione con la prova Cloze, dall'altra il fatto che sarà la sola prova TCL ad essere effettivamente utilizzata per una graduatoria di accesso (cfr. 7.2)

siderando solo la fascia A della distribuzione IRT tradizionale. In questo caso il gruppo si riduce a 32 studenti.

Tabella 5.46 – Confronto graduatorie TCL - Fascia alta dei punteggi (IRT)

ID	Grezzi	Classica PT	Graduatoria classica	IRT PT	Graduatoria IRT	IRTom PT	Graduatoria IRTom
60	29	77,98	2	94,89	1	93,86	1
89	29	77,98	1	93,62	2	93,03	2
509	28	75,89	3	83,94	3	82,29	3
929	22	63,36	45	83,67	4	63,11	39
88	28	75,89	4	83,61	5	82,12	4
616	27	73,80	5	77,06	6	76,95	5
410	26	71,71	6	74,34	7	74,02	6
736	26	71,71	7	73,83	8	73,80	7
4	25	69,63	8	71,55	9	71,71	8
770	25	69,63	13	70,71	10	69,68	14
639	25	69,63	11	70,61	11	70,29	10
235	25	69,63	10	70,48	12	70,63	9
330	25	69,63	12	70,46	13	70,29	11
868	25	69,63	9	70,36	14	69,91	13
239	18	55,01	130	68,71	15	54,76	115
641	23	65,45	28	68,21	16	67,12	23
94	24	67,54	24	68,03	17	69,98	12
820	24	67,54	18	67,53	18	67,97	20
654	24	67,54	25	67,45	19	67,72	21
547	24	67,54	19	67,25	20	69,14	17
596	24	67,54	20	67,25	21	69,14	18
216	24	67,54	14	67,15	22	69,36	16
317	24	67,54	16	67,03	23	69,37	15
531	24	67,54	15	66,94	24	67,01	25
524	24	67,54	23	66,87	25	67,11	24
242	24	67,54	21	66,64	26	66,92	26
526	24	67,54	22	66,59	27	68,87	19
636	24	67,54	27	66,59	28	67,22	22
703	24	67,54	17	66,52	29	66,84	27
862	20	59,19	80	65,79	30	60,29	65
409	24	67,54	26	65,77	31	65,28	31
512	23	65,45	37	65,24	32	65,39	30

Sono state evidenziate in grigio le discrepanze più rilevanti fra le graduatorie.



Per favorire la lettura incrociata delle tabelle l'ordine delle graduatorie è immutato, ma naturalmente la graduatoria di riferimento è quella IRT, che si trova nella parte centrale della tabella.

Se nella tabella precedente, ordinata sulla graduatoria classica, non avevamo notato grossi sbalzi di posizione, in questa spiccano tre casi particolari (pur segnalato per l'uscita dal gruppo dei migliori 32, lo studente 512 ha perso in realtà solo poche posizioni nella graduatoria classica), che denotano un forte impatto dei punteggi IRT rispetto alle altre due modalità di analisi. Gli studenti 929 e 862 guadagnano decine di posizioni, addirittura più di un centinaio lo studente 239, che quindi secondo l'item analisi IRT si trovano nel gruppo di eccellenza identificato dalla prova di accesso. Se osserviamo il numero di risposte corrette di questi studenti (Grezzi) vediamo che risulta più basso rispetto agli altri. In questo caso è possibile rilevare gli effetti prodotti dalla distinzione operata dall'analisi IRT tra omissioni e domande non raggiunte: tutti e tre gli studenti in questione non hanno completato la prova, anzi hanno lasciato in bianco un discreto numero di domande nella parte finale.

Tabella 5.47 – Riepilogo casi particolari TCL - Fascia alta dei punteggi (IRT)

ID	Esatte	Errate	Omissioni	Non Raggiunte
929	22	2	0	6
239	18	5	0	7
862	20	5	1	4

Il caso dello studente 929 risulta esemplare: ha risposto a solo 24 domande, non raggiungendo le ultime sei, e su quelle 24 risposte l'item analisi IRT ha stimato la sua competenza, mettendola di fatto allo stesso livello degli studenti 509 e 88, che hanno completato la prova dando 28 risposte corrette. Di fatto quello che unisce questi tre studenti è il numero di risposte errate, solo 2. La modalità di trattamento delle omissioni ha quindi un forte effetto sui punteggi di alcuni studenti che, per scelta consapevole o mancanza di tempo, non rispondono alle ultime domande.

Da questo punto di vista la penalizzazione reintrodotta con il modello IRT Om, che ricordiamo penalizza allo stesso modo tutte le omissioni, sembra avere un fattore riequilibrante: i tre studenti in questione guadagnano comunque diverse posizioni (da 6 a 15) rispetto alla graduatoria Classica, per cui viene riconosciuto da una parte il buon risultato nei quesiti affrontati con certezza (per esempio lo studente 239 con 18 risposte esatte su 23 si trova, in questa graduatoria, allo stesso livello di stu-

denti che hanno dato 19 o 20 risposte esatte sull'intera prova), dall'altro riassume l'indubbia distorsione introdotta dall'analisi IRT tradizionale.

Il confronto fra le graduatorie si fa molto delicato se consideriamo la fascia corrispondente alla fascia in cui viene operato il taglio per decidere l'accesso. La prima tabella considera la graduatoria determinata dai punteggi dell'item analisi classica. L'ampiezza della fascia considerata (26 studenti) è determinata dal punteggio grezzo: la soglia attraversa il gruppo che ha dato 17 risposte corrette.

Tabella 5.48 – Confronto graduatorie TCL - Fascia di taglio dei punteggi (Classica)

ID	Grezzi	Classica PT	Graduatoria classica	IRT PT	Graduatoria IRT	IRTom PT	Graduatoria IRTom
252	17	52,92	141	49,09	200	50,50	183
520	17	52,92	142	50,04	183	52,27	149
734	17	52,92	143	49,90	186	52,82	141
350	17	52,92	144	49,22	197	53,75	127
709	17	52,92	145	50,47	173	51,96	155
510	17	52,92	146	50,06	182	52,43	148
895	17	52,92	147	49,76	189	52,27	150
236	17	52,92	148	50,75	169	52,22	151
322	17	52,92	149	50,29	176	51,47	159
846	17	52,92	150	48,58	213	50,93	172
787	17	52,92	151	50,41	175	52,75	144
904	17	52,92	152	49,91	185	51,10	166
316	17	52,92	153	50,57	171	51,39	161
206	17	52,92	154	56,15	97	53,67	131
611	17	52,92	155	49,89	187	52,55	147
852	17	52,92	156	50,54	172	51,69	157
521	17	52,92	157	49,62	193	50,35	188
327	17	52,92	158	49,66	191	51,54	158
856	17	52,92	159	49,13	198	50,49	184
78	17	52,92	160	48,78	207	51,93	156
7	17	52,92	161	50,83	167	53,06	138
21	17	52,92	162	51,00	163	53,48	135
85	17	52,92	163	56,72	87	53,28	136
938	17	52,92	164	49,83	188	51,26	164
853	17	52,92	165	49,97	184	50,71	177
541	17	52,92	166	49,55	195	50,23	190

Sono state evidenziate in grigio le discrepanze fra le graduatorie ai fini dell'ammissione.

Se la graduatoria così determinata fosse stata effettivamente usata per l'accesso, sui 26 studenti che hanno riportato il punteggio grezzo 17 solo 10 sarebbero stati ammessi al corso di laurea, e 16 si sarebbero trovati esclusi solo per il voto riportato nell'Esame di Stato (per inciso notiamo che, per la soglia di accesso non è stato necessario ricorrere al criterio anagrafico). Per gli stessi studenti l'analisi IRT non identifica nessuna coppia di punteggi uguali, e nell'analisi IRT Om si determina un solo ex aequo (ID 520 e 895, proprio in corrispondenza del taglio per l'accesso, ma con un piccolo scarto dei due studenti rispetto al 236, che risulta il primo dei non ammessi).

Il problema degli *ex aequo* passa però in secondo piano se consideriamo che 24 dei 26 studenti con punteggio grezzo 17 non sarebbero stati ammessi secondo la graduatoria IRT, anche a prescindere dai criteri esterni, e che i due rimanenti (206 e 85) non risultano invece ammessi secondo la graduatoria classica, in cui perdono molte decine di posizioni. Spicca all'opposto la situazione dello studente 846, ultimo degli ammessi nella graduatoria Classica, che precipita di oltre 60 posizione nella graduatoria IRT.

Anche in questo caso la distorsione delle graduatorie è determinata soprattutto dalla distinzione fra omissioni e domande non raggiunte operata dall'analisi IRT. Nella Tabella 5.49 appare evidente come, a parità di risposte esatte, l'analisi IRT tradizionale premia il minor numero di errori, con lo studente 85 che precede il 206 di 10 posizioni ed entrambi che staccano nettamente (oltre cento posizioni) il collega 846, che ha avuto il demerito di rispondere a tutte le domande, facendo ben 13 errori.

Tabella 5.49 – Riepilogo casi particolari TCL - Fascia di taglio dei punteggi (Classica)

ID	Esatte	Errate	Omissioni	Non Raggiunte
846	17	13	0	0
206	17	9	0	4
85	17	8	1	4

Notiamo che anche in questo caso l'analisi IRT alternativa ha decisamente attenuato l'effetto delle omissioni: da una parte lo studente 846 esce dal gruppo degli ammessi, ma perdendo solo una ventina di posizioni, dall'altra gli studenti 206 e 85 rientrano tra gli ammessi, ma con scarto molto ridotto rispetto alla graduatoria Classica.

L'estrema diversità di selezione operata dall'item analisi IRT colpisce però anche oltre la fascia di taglio.

Nella Tabella 5.50 sono riportati gli studenti sicuramente esclusi secondo la graduatoria classica, ma che risultano ammessi secondo l'analisi IRT. Si tratta di 22 studenti che si aggiungono ai 2 già evidenziati nella precedente tabella.

Tabella 5.50 – Confronto graduatorie TCL - Fascia bassa dei punteggi (Classica)

ID	Grezzi	Classica PT	Graduatoria classica	IRT PT	Graduatoria IRT	IRTom PT	Graduatoria IRTom
625	16	50,84	168	58,18	82	50,94	171
63	16	50,84	178	63,63	38	50,81	175
737	16	50,84	188	59,71	64	52,00	154
947	16	50,84	190	56,14	98	51,35	162
863	16	50,84	196	52,91	140	50,60	181
65	16	50,84	202	52,85	141	51,47	160
516	15	48,75	216	52,72	142	49,60	200
751	14	46,66	232	55,37	106	47,25	234
514	14	46,66	243	54,08	122	46,42	256
312	14	46,66	250	54,82	110	46,56	253
244	14	46,66	256	54,37	113	46,33	258
944	14	46,66	260	56,08	100	48,12	224
208	13	44,57	271	52,18	149	45,02	284
891	13	44,57	272	52,50	146	45,26	278
95	13	44,57	286	61,13	57	44,40	296
834	13	44,57	294	54,20	118	44,32	297
782	12	42,49	301	53,17	137	43,06	317
33	12	42,49	302	57,65	84	42,90	321
874	12	42,49	321	59,53	65	42,96	319
96	11	40,40	338	53,55	133	40,78	349
841	10	38,31	361	62,31	46	39,71	367
223	10	38,31	380	54,25	117	38,88	380

Sono state evidenziate in grigio le discrepanze fra le graduatorie ai fini dell'ammissione.

Tutti i 22 studenti sono accomunati da un numero relativamente alto di domande non raggiunte, che permette loro di risalire la graduatoria IRT, anche in modo vertiginoso. Analizzando nel dettaglio alcune situazioni spicca lo studente 63, che con 16 risposte esatte scala ben 140 posizioni, arrivando alla soglia della fascia di eccellenza. Addirittura clamorosa la situazione dello studente 841, che con 10 risposte esatte passa dalla 361° posizione della graduatoria Classica alla 46° secondo i pun-

teggi IRT. Impressionanti anche gli scarti degli studenti 95 e 874. Risulta opportuno verificare anche per questi studenti l'incidenza delle omissioni, e si può notare che il numero di domande non raggiunte è altissimo, difficilmente spiegabile solo con la mancanza di tempo per rispondere.

Tabella 5.51 – Riepilogo casi particolari TCL - Fascia bassa dei punteggi (Classica)

ID	Esatte	Errate	Omissioni	Non Raggiunte
63	16	5	1	8
95	13	6	0	11
874	12	4	2	12
841	10	4	0	16

La correzione introdotta dall'item analisi IRT con penalizzazione di tutte le omissioni si rivela anche in questo caso preziosa: notiamo infatti che proprio per i 4 casi segnalati nel dettaglio le differenze rispetto alla graduatoria Classica risultano minime, quindi mentre in precedenza avevamo notato una correzione in positivo dei risultati grezzi per gli studenti con un numero di omissioni complessivamente contenuto, se le risposte omesse crescono, il modello IRT Om attenua questa capacità correttiva e penalizza in modo simile errori ed omissioni: nessuno degli studenti ammessi secondo l'analisi IRT (Tabella 5.50) sarebbe stato incluso nella graduatoria IRT Om.

Per completare il confronto risulta opportuno, come già fatto per la fascia di eccellenza dei punteggi, riportare la tabella relativa alla fascia di taglio per l'accesso, ma costruita secondo la graduatoria IRT tradizionale. In questo caso, mancando un riferimento specifico per l'ampiezza della fascia (distribuzione pentenaria o stesso punteggio grezzo), sono state selezionate le posizioni a cavallo della soglia, tra il 140° e il 163° posto, compresi fra 51 e 53 punti T.

Si può notare la presenza di un solo punteggio *ex aequo* nella graduatoria IRT (ID 758 e 717) che peraltro non pregiudica la soglia dei 150 posti utili. In questo caso degli ultimi 11 studenti ammessi secondo la graduatoria IRT, 5 sarebbero esclusi considerando i soli punteggi grezzi, mentre tra i primi 12 esclusi ben 10 sarebbero riammessi in base alle sole risposte esatte.

E' interessante notare che per questo gruppo di studenti non esistono sostanziali differenze fra la graduatoria classica e quella IRT alternativa: gli scarti fra le due graduatorie non producono effetti ai fini dell'ammissione/esclusione rispetto alla soglia di taglio.

Tabella 5.52 – Confronto graduatorie TCL - Fascia di taglio dei punteggi (IRT)

ID	Grezzi	Classica PT	Graduatoria classica	IRT PT	Graduatoria IRT	IRTom PT	Graduatoria IRTom
863	16	50,84	196	52,91	140	50,60	181
65	16	50,84	202	52,85	141	51,47	160
516	15	48,75	216	52,72	142	49,60	200
233	18	55,01	132	52,68	143	53,75	128
763	18	55,01	127	52,61	144	56,25	100
802	18	55,01	131	52,52	145	53,78	126
891	13	44,57	272	52,50	146	45,26	278
416	18	55,01	128	52,41	147	53,51	133
637	18	55,01	120	52,37	148	53,70	129
208	13	44,57	271	52,18	149	45,02	284
341	18	55,01	137	52,15	150	55,34	109
823	18	55,01	122	52,13	151	52,81	142
29	18	55,01	123	51,99	152	53,49	134
545	16	50,84	204	51,85	153	52,07	153
758	18	55,01	124	51,78	154	52,80	143
717	18	55,01	134	51,78	155	53,66	132
828	18	55,01	125	51,73	156	52,84	140
724	18	55,01	140	51,55	157	54,30	122
325	9	36,22	394	51,51	158	37,49	391
677	18	55,01	119	51,48	159	52,64	146
538	18	55,01	135	51,27	160	52,65	145
34	18	55,01	126	51,19	161	53,00	139
708	18	55,01	138	51,08	162	53,90	125

Sono state evidenziate in grigio le discrepanze fra le graduatorie ai fini dell'ammissione.

Ai 10 studenti segnalati dalla Tabella 5.52 come esclusi dalla graduatoria IRT, ma ammessi per i punteggi grezzi, vanno aggiunti altri 14 studenti nella stessa situazione, ma in posizione ancora più bassa di graduatoria. Nella Tabella 5.53 sono riportati proprio gli studenti penalizzati dai punteggi IRT. In questo modo si compensano i 24 studenti al contrario si erano visti premiati rispetto alla graduatoria determinata dall'item analisi classica.

Tabella 5.53 – Confronto graduatorie TCL - Fascia bassa dei punteggi (IRT)

ID	Grezzi	Classica PT	Graduatoria classica	IRT PT	Graduatoria IRT	IRTom PT	Graduatoria IRTom
219	18	55,01	129	50,98	165	52,14	152
311	18	55,01	118	50,95	166	53,69	130
833	18	55,01	133	50,81	168	54,31	121
236	17	52,92	148	50,75	169	52,22	151
709	17	52,92	145	50,47	173	51,96	155
322	17	52,92	149	50,29	176	51,47	159
933	18	55,01	136	50,06	180	53,12	137
510	17	52,92	146	50,06	182	52,43	148
520	17	52,92	142	50,04	183	52,27	149
734	17	52,92	143	49,90	186	52,82	141
895	17	52,92	147	49,76	189	52,27	150
350	17	52,92	144	49,22	197	53,75	127
252	17	52,92	141	49,09	200	50,50	183
846	17	52,92	150	48,58	213	50,93	172

Sono state evidenziate in grigio le discrepanze fra le graduatorie ai fini dell'ammissione.

Notiamo che in questo caso dei 14 studenti ammessi secondo la graduatoria Classica, 5 sarebbero rimasti esclusi anche per la graduatoria IRT Om. In 4 casi (219, 236, 709 e 322) l'esclusione dai 150 ammessi è di misura, con scarti contenuti fra le due graduatorie. Più marcata invece la differenza per lo studente 252, per cui risulta utile approfondire il profilo di risposta di questi studenti.

Tabella 5.54 – Riepilogo casi particolari TCL - Fascia bassa dei punteggi (IRT)

ID	Esatte	Errate	Omissioni	Non Raggiunte
219	18	11	1	0
236	17	13	0	0
709	17	11	2	0
322	17	13	0	0
252	17	9	4	0

La maggiore penalizzazione determinata per lo studente 252 rispetto ai colleghi che hanno dato lo stesso numero di risposte corrette (17) è data proprio dal maggior numero di omissioni, poiché, ricordiamo, il modello IRT determina i punteggi considerando anche le risposte errate, distinguendo la penalizzazione secondo il distrattore scelto, per cui alcune

risposte errate possono risultare meno penalizzanti rispetto all'omissione. In questo caso è evidente che le 4 risposte non date dallo studente 252 risultano omesse in larga misura da studenti di livello basso, per cui il punteggio individuale risente di questa distribuzione delle omissioni. È da notare comunque che in tutti e cinque i casi segnalati i punteggi IRT Om si trovano (ancora una volta) in posizione intermedia rispetto alle altre graduatorie, confermando la capacità riequilibratrice rispetto alla presenza di omissioni di risposta.

Proprio per verificare questa capacità, concludiamo il confronto fra le diverse graduatorie con la Tabella 5.55, che riporta la fascia di taglio secondo la graduatoria IRT Om, considerando i punteggi compresi fra 51,5 e 53.

Tabella 5.55 – Confronto graduatorie TCL - Fascia di taglio dei punteggi (IRT Om)

ID	Grezzi	Classica PT	Graduatoria classica	IRT PT	Graduatoria IRT	IRTom PT	Graduatoria IRTom
34	18	55,01	126	51,19	161	53,00	139
828	18	55,01	125	51,73	156	52,84	140
734	17	52,92	143	49,90	186	52,82	141
823	18	55,01	122	52,13	151	52,81	142
758	18	55,01	134	51,78	154	52,80	143
787	17	52,92	151	50,41	175	52,75	144
538	18	55,01	135	51,27	160	52,65	145
677	18	55,01	119	51,48	159	52,64	146
611	17	52,92	155	49,89	187	52,55	147
510	17	52,92	146	50,06	182	52,43	148
520	17	52,92	142	50,04	183	52,27	149
895	17	52,92	147	49,76	189	52,27	150
236	17	52,92	148	50,75	169	52,22	151
219	18	55,01	129	50,98	165	52,14	152
545	16	50,84	204	51,85	153	52,07	153
737	16	50,84	188	59,71	64	52,00	154
709	17	52,92	145	50,47	173	51,96	155
78	17	52,92	160	48,78	207	51,93	156
852	17	52,92	156	50,54	172	51,69	157
327	17	52,92	158	49,66	191	51,54	158

Sono state evidenziate in grigio le discrepanze fra le graduatorie ai fini dell'ammissione.



Se tutti gli ultimi 12 ammessi secondo la graduatoria IRT alternativa risulterebbero esclusi per i punteggi IRT tradizionali, solo due di loro risultano esclusi dall'analisi dei punteggi grezzi, facendo comunque parte della fascia critica dei 17 punti grezzi che viene tagliata proprio dalla soglia fissata a 150 posti. Al contrario tra i primi esclusi secondo la graduatoria IRT Om figurano tre studenti (236, 219 e 709) ammessi secondo i punteggi grezzi, ma anche per due di loro, con un punteggio di 17, si tratta di una esclusione determinata da criteri esterni alla prova (voto nell'Esame di Stato. La Tabella 5.56 riporta gli unici altri casi di esclusione di questo tipo.

Tabella 5.56 – Confronto graduatorie TCL - Fascia medio-alta dei punteggi (IRT Om)

ID	Grezzi	Classica PT	Graduatoria classica	IRT PT	Graduatoria IRT	IRTom PT	Graduatoria IRTom
206	17	52,92	154	56,15	97	53,67	131
21	17	52,92	162	51,00	163	53,48	135
85	17	52,92	163	56,72	87	53,28	136
7	17	52,92	161	50,83	167	53,06	138

Sono state evidenziate in grigio le discrepanze fra le graduatorie ai fini dell'ammissione.

È facilmente constatabile che si tratta di quattro casi sempre legati alla discriminazione operata secondo il criterio esterno, poiché tutti e quattro gli studenti hanno ottenuto il punteggio grezzo di 17. È interessante notare la collocazione di questi quattro studenti nella graduatoria IRT: si formano due coppie separate da una settantina di posizioni, laddove nelle altre due graduatorie le differenze interne risultano molto contenute. Anche in questo caso la differenza è determinata dal diverso profilo dei rispondenti, con gli studenti 206 e 85 che non raggiungono le ultime quattro domande, quindi contano un numero inferiore di risposte errate.

Tabella 5.57 – Riepilogo casi particolari TCL - Fascia medio-alta dei punteggi (IRT)

ID	Esatte	Errate	Omissioni	Non Raggiunte
206	17	9	0	4
21	17	11	1	1
85	17	8	1	4
7	17	13	0	0

Ancora una volta la graduatoria IRT premia il minor numero di errori palesi, attribuendo a questi studenti un punteggio stimato che non viene penalizzato dalle omissioni nelle ultime domande della prova.

Alla luce del confronto delle diverse graduatorie prodotte secondo le diverse modalità di analisi degli item si può concludere che, pur considerando la sostanziale sovrapposizione delle stesse graduatorie rilevata dall'analisi delle correlazioni (cfr. 5.5.4), gli effetti sulle situazioni individuali degli studenti risultano rilevanti.

In particolare la determinazione di una soglia significativa all'interno delle graduatorie, utile per la selezione di accesso, rende particolarmente critica la fascia di punteggi in prossimità di tale soglia.

La graduatoria definita secondo l'item analisi classica, basata sul numero di risposte corrette, finisce per presentare notevoli problemi di discriminatività della scala proprio in questa fascia, tanto più la soglia si avvicina al punteggio medio della prova, dove, considerata la distribuzione normale dei punteggi, si addensano gli studenti dello stesso livello.

Le graduatorie definite secondo l'item analisi IRT risentono in maniera decisamente inferiore di questo fenomeno, considerando che i punteggi individuali vengono stimati secondo il profilo di risposta, per cui lo stesso numero di risposte esatte non corrisponde allo stesso punteggio, da cui la quasi totale mancanza di punteggi *ex aequo*<sup>101</sup>. Questo tipo di graduatorie risultano quindi più funzionali, secondo quanto già ipotizzato in 3.5, per la definizione di graduatorie più precise e articolate.

Se confrontiamo i diversi gruppi di studenti ammessi al corso di laurea secondo le tre graduatorie, abbiamo visto che le differenze esistenti fra graduatoria classica e graduatoria IRT tradizionale riguardano 24 studenti su 150, una percentuale contenuta (16%), ma non certo irrilevante se consideriamo la buona correlazione fra le due graduatorie. In effetti, come abbiamo visto per le diverse fasce, le differenze fra le due graduatorie si fondano soprattutto sul diverso trattamento delle omissioni, in particolare quelle relative alle domande non raggiunte.

Ricordando quanto illustrato nel Paragrafo 3.4 sui dubbi tuttora esistenti circa l'interpretazione delle omissioni rispetto alla competenza in-

<sup>101</sup> Ricordiamo che nella graduatoria IRT tradizionale si sono verificati solo quattro casi di stesso punteggio per due studenti, mentre nella graduatoria IRT Om le copie di punteggi uguali sono state sette.

dagata dalla prova, dubbi rafforzati dall'uso specifico finalizzato all'accesso universitario<sup>102</sup>, appare quantomeno problematico dare credito a una graduatoria fondata su un sistema di analisi delle risposte che, estremizzando quanto rilevato per alcuni studenti nella nostra prova di appoggio, potrebbe assegnare un punteggio brillante a uno studente che rispondesse solo a poche domande, ma tutte concentrate nella prima parte della prova.

Si tratta senza dubbio di possibilità estreme, che nella normalità dello svolgimento delle prove la maggior parte degli studenti completa tutte le domande o omette di rispondere solo in pochi casi<sup>103</sup>, ma proprio lo scopo selettivo della prova non rende accettabile risultati dettati da un andamento anomalo delle omissioni, anche se limitati a pochi studenti.

Di conseguenza s'impone all'attenzione la modalità alternativa di item analisi, IRT Om, che prevede il trattamento indistinto delle omissioni, penalizzandole normalmente in qualsiasi parte della prova esse si trovino.

In questo caso le differenze fra questa graduatoria e quella Classica comportano una diversa selezione per soli sei studenti su 150 (4%), limitandosi nella maggior parte dei casi a risolvere in maniera diversa, basandosi esclusivamente sui risultati della prova attraverso la pesatura dei diversi quesiti, i casi di ex aequo che la graduatoria Classica aveva sciolto appellandosi a criteri esterni.

Oltre a questa capacità di rendere più fluida la scala dei punteggi, l'analisi alternativa IRT corregge positivamente anche i punteggi degli studenti che non rispondono a tutte le domande, ma si tratta di una correzione parziale, laddove la graduatoria basata sui punteggi grezzi equipara semplicemente l'omissione all'errore, mentre l'analisi IRT, ignorando le domande non raggiunte, finisce per squilibrare eccessivamente i punteggi di alcuni studenti.

Considerando anche la forte correlazione esistente fra graduatoria Classica e IRT Om (ancora più forte rispetto a quella attestata per IRT

<sup>102</sup> Si tratta evidentemente di problemi non solo legati alle caratteristiche psicometriche della prova, ma anche di tipo burocratico, legati al valore legale della prova stessa.

<sup>103</sup> Anche nella prova TCL in definitiva l'incidenza delle omissioni si è attestata complessivamente intorno al 10%, ciò spiega anche la forte correlazione fra le graduatorie..

tradizionale), la graduatoria basata sull'analisi IRT alternativa risulta essere la più funzionale, se l'accesso a un corso di studi universitario è legato ai risultati di una prova selettiva basata su quesiti a risposta chiusa.

Naturalmente sarà importante controllare le relazioni esistenti fra le diverse graduatorie e gli esiti del primo anno di corso, per verificare la loro possibile capacità predittiva.

## Capitolo sesto

### Verifica dell'ipotesi

#### 6.1. La raccolta dei dati riguardanti gli esiti accademici del primo anno di corso

Come abbiamo già visto nel paragrafo 4.6, è stata predisposta una scheda di rilevazione del percorso accademico per raccogliere i dati riguardanti i nuovi iscritti dell'A.A. 2003-04 al termine del primo anno di corso<sup>104</sup>. Il periodo compreso fra febbraio e aprile 2005 è stato dedicato alla ricerca degli studenti che risultavano ancora iscritti al Corso di laurea in Scienze dell'Educazione e della Formazione dell'Università di Roma "La Sapienza" per la compilazione della scheda. Il lavoro si è rivelato abbastanza impegnativo, considerando l'ampiezza della leva 2003-04 e la sua prevedibile diminuzione al termine del primo anno di corso<sup>105</sup>.

La ricerca degli studenti è stata svolta in collaborazione con il gruppo di lavoro per il monitoraggio del corso di laurea, coordinato da Benvenuto, permettendo anche un adeguato aggiornamento delle liste degli studenti del corso di laurea. Al termine del lavoro di ricerca sono stati raccolti i dati riguardanti 242 studenti su 419 partecipanti alla prova (58%), di cui 231 avevano svolto almeno un esame nel corso del primo anno.<sup>106</sup> Dei 176 studenti mancanti è stato accertato l'abbandono o il trasferimento ad altri corsi di studio per 117 casi, mentre i restanti 59 risultano dispersi, in quanto non è stato possibile raggiungerli o avere notizie affidabili sulla loro situazione di studio.

<sup>104</sup> La stessa rilevazione è stata svolta dalla collega Truffelli per gli studenti di Bologna, ma bisogna considerare che per gli studenti di Roma è stato possibile svolgere esami del I anno fino al mese di dicembre 2004, mentre a Bologna sono considerati di primo anno anche gli esami sostenuti fino alla sessione di marzo 2005.

<sup>105</sup> Cfr. Benvenuto (2003), al termine del I anno 2001-02 è stato possibile ricostruire il percorso accademico, su 275 nuovi iscritti, di 129 studenti (47%).

<sup>106</sup> Gli 11 studenti che non hanno svolto esami sono comunque ancora regolarmente iscritti al corso di laurea.

La compilazione della scheda non ha comportato particolari problemi, tutti gli studenti raggiunti hanno specificato nelle diverse sezioni gli esami sostenuti con i voti relativi, le modalità di svolgimento dell'orale (scritto, orale, scritto/orale) e il tipo di frequenza per ogni esame sostenuto (regolare, parziale, non frequentante). Inoltre sono state indicate in un quadro apposito le idoneità conseguite (informatica e lingua inglese), poiché non comportano voto e quindi non rientrano nel calcolo della media, ma distinguendo le modalità di conseguimento (prova sostenuta o presentazione di attestazione), poiché nel primo caso possono essere equiparate alla preparazione di un esame tradizionale.<sup>107</sup>

Se la compilazione delle sezioni in positivo (esami e prove completati) non ha riservato sorprese, più problematica si è rivelata la compilazione della sezione riguardante gli esami non andati a buon fine. Le istruzioni, per evidenti problemi di *privacy* (i dati degli esami sostenuti sono pubblici e rilevabili dalle segreterie) permettevano di non compilare la sezione insuccessi, e per dichiarazioni raccolte fra gli stessi studenti, non pochi fra questi hanno ammesso esplicitamente di aver lasciato volontariamente in bianco questa sezione. Questo non permette di considerare attendibili i dati su gli esami non completati, anche se il 32,6% degli studenti ha compilato questa sezione.<sup>108</sup>

I dati raccolti sono stati inseriti nel database già allestito per la prova di ingresso, da cui in seguito è stato estratto un database specifico riguardante i 242 studenti che rappresentano il campione di indagine definitivo per verificare la possibile relazione fra risultati della prova di accesso ed esiti accademici.

## 6.2. Verifica del campione sul quale sono stati rilevati gli esiti accademici

Come abbiamo già fatto per il gruppo di studenti che ha sostenuto la prova di ingresso, risulta opportuna una descrizione accurata del grup-

<sup>107</sup> Possiamo anticipare che il 44,2% degli studenti ha conseguito idoneità, ma solo il 12,8% lo ha fatto sostenendo prove specifiche, mentre il 32,6% ha presentato attestazioni (lo 0,8% ha utilizzato entrambe le modalità).

<sup>108</sup> Di questi il 15,6% ha dichiarato l'insuccesso in un esame, il 13,2% in due esami, il 3,7% in tre o più esami. La maggior parte degli insuccessi riguarda gli esami scritti (28,1%) rispetto agli orali (5,4%, lo 0,9 ha sperimentato entrambi i tipi di insuccesso). Nel complesso i crediti mancati sono stati 638, a fronte di 7072 effettivamente conseguiti.

po di studenti in uscita dal primo anno, questa volta utilizzando come riscontro non gli iscritti di anni precedenti, ma proprio il gruppo che ha partecipato alla prova. In tal modo sarà possibile verificare se qualche fattore può aver inciso in maniera significativa nella riduzione del gruppo. Si è proceduto costruendo delle tavole di contingenza basate sulla distinzione esito/non esito accademico e, di volta in volta, su una diversa variabile rilevata dal Questionario nuovi iscritti. È stato così possibile osservare la significatività del Chi quadrato di Pearson, integrandola laddove necessario con la misura di associazione lineare-lineare<sup>109</sup>.

Per chiarezza di esposizione nelle tabelle è riportato il confronto fra il gruppo in entrata, che ha sostenuto la prova di ingresso, e il gruppo in uscita che ha compilato la scheda di rilevazione degli esiti accademici.

Nella Tabella 6.1 è riepilogata l'articolazione fra matricole e nuovi iscritti.

Tabella 6.1 – Provenienza dei nuovi iscritti in ingresso e in uscita (I anno)

	Prova di ingresso		Esiti accademici	
	Frequenza	%	Frequenza	%
Matricole	332	79,2	183	75,6
Già iscritti	87	20,8	59	24,4
Totale	419	100,0	242	100,0

Si può notare che la perdita è stata maggiore fra le matricole (45% di coloro che avevano sostenuto la prova di ingresso) rispetto a chi aveva avuto precedenti esperienze universitarie (32%), e il calcolo del Chi quadrato presenta una significatività di 0,033. Quindi il fattore esperienza risulta essere influente, almeno per il completamento del I anno di corso.<sup>110</sup>

Considerando l'articolazione del gruppo in uscita non emergono invece differenze significative riguardo al sesso, poiché le frequenze effettive risultano molto vicine a quelle attese.

<sup>109</sup> Viene calcolata attraverso la formula del Chi quadrato di Mantel-Haenszel, verifica l'esistenza di relazioni lineari fra variabili continue di riga e colonna

<sup>110</sup> Naturalmente un riscontro effettivo di questo aspetto dovrebbe considerare anche il destino di chi non ha completato il primo anno, in quanto è ben differente la situazione di chi abbandona del tutto gli studi rispetto a chi semplicemente si trasferisce in un'altra realtà universitaria.

Tabella 6.2 – Nuovi iscritti per genere in ingresso e in uscita (I anno)

	Prova di ingresso		Esiti accademici	
	Frequenza	%	Frequenza	%
Femmine	382	91,2	220	90,9
Maschi	37	8,8	22	9,1
Totale	419	100,0	242	100,0

Allo stesso modo non risultano differenze significative legate all'età dei nuovi iscritti. La composizione del gruppo di studenti in uscita ricalca sostanzialmente l'articolazione per fasce di età già presentata nella Tabella 5.4., con una perdita appena più alta per le due fasce di maggiore anzianità.

Tabella 6.3 – Nuovi iscritti per fasce di età in ingresso e in uscita (I anno)

	Prova di ingresso		Esiti accademici	
	Frequenza	%	Frequenza	%
18-20	228	54,4	138	57,0
21-23	83	19,8	48	19,8
24-30	45	10,7	22	9,1
>30	63	15,1	34	14,1
Totale	419	100,0	242	100,0

Anche il titolo di studio che ha permesso l'accesso all'università non risulta significativamente diverso fra i due gruppi (Chi-quadrato sign. 0,73) ma si può notare il peso crescente degli ex liceali, che raggiungono quasi il 70% del gruppo in uscita, a fronte di una parallela diminuzione di diplomati degli istituti tecnici e professionali.

Tabella 6.4 – Nuovi iscritti per titolo secondario superiore in ingresso e in uscita (I anno)

	Prova di ingresso		Esiti accademici	
	Frequenza	%	Frequenza	%
Liceo	270	65,2	166	69,7
Tecnico	110	26,6	56	23,6
Professionale	34	8,2	16	6,7
Totale	414	100,0	238	100,0

Sono state escluse le omissioni di risposta.



Decisamente più interessante il confronto relativo all'articolazione dei due gruppi in ingresso e uscita per il voto riportato all'Esame di Stato. In questo caso la significatività esiste (sign 0,003), ma non di tipo lineare (sign. 0,336), infatti nel gruppo di studenti in uscita le fasce di voto che crescono in percentuale rispetto al gruppo in ingresso sono la più bassa (60), la più alta (100), e una delle intermedie (80-89), mentre le altre fasce perdono di importanza.

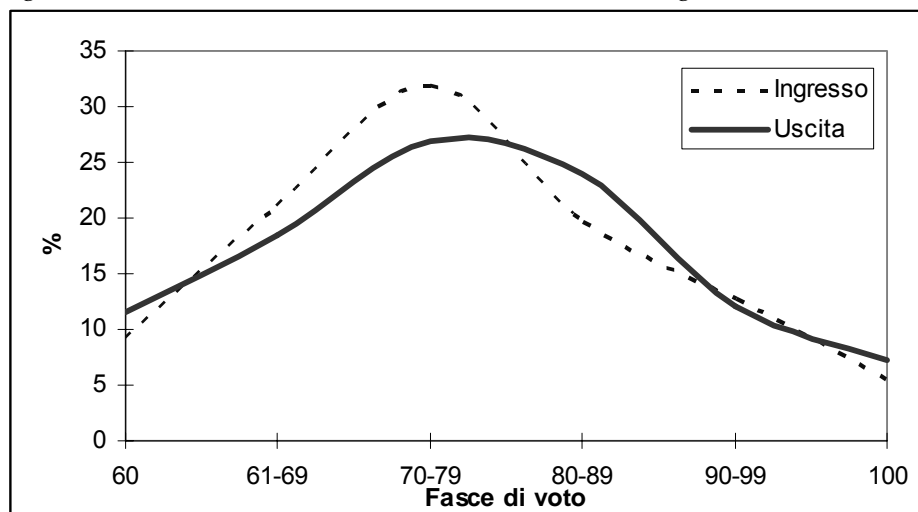
Tabella 6.5 – Nuovi iscritti per fasce di voto nell'Esame di Stato in ingresso e in uscita (I anno)

	Prova di ingresso		Esiti accademici	
	Frequenza	%	Frequenza	%
60	37	9,1	27	11,5
61-69	86	21,1	43	18,4
70-79	130	31,9	63	26,9
80-89	80	19,6	56	23,9
90-99	52	12,7	28	12,0
100	23	5,6	17	7,3
Totale	408	100,0	234	100,0

Sono state escluse le omissioni di risposta.

Questa tenuta delle fasce estreme nel gruppo in uscita determina una diversa curva di distribuzione dei voti. Nella Figura 6.1 vengono confrontate proprio le due curve risultanti dalle distribuzioni del gruppo di ingresso e del gruppo di uscita.

Figura 6.1 – Curve di distribuzione voto Esame di Stato in ingresso e in uscita



È evidente la minore pendenza della curva (anche il valore modale scende di oltre 5 punti) e la tendenza alla normalità (rispetto a una marcata asimmetria positiva), testimoniate anche dall'evoluzione dei valori di curtosi e asimmetria.

Tabella 6.6 – Curtosi e asimmetria distribuzioni in ingresso e in uscita (I anno)

	Asimmetria	Err. std	Curtosi	Err. std
Ingresso	0,26	0,12	-0,54	0,24
Uscita	0,15	0,16	-0,67	0,32

Come si può notare il valore dell'asimmetria si avvicina sensibilmente alla normalità (e risulta anche inferiore al relativo errore standard), mentre l'aumento della negatività della curtosi dimostra proprio l'appiattimento della curva. Nel complesso il voto medio del gruppo in uscita cresce di pochissimo (da 76,4 a 76,8), con una crescita di poco più marcata della deviazione standard (da 11,8 a 12,2)

Assoluta coincidenza nelle distribuzioni dei due gruppi per quanto riguarda i titoli di studio dei genitori.

Tabella 6.7 – Nuovi iscritti per titolo di studio del padre in ingresso e in uscita (I anno)

	Prova di ingresso		Esiti accademici	
	Frequenza	%	Frequenza	%
Nessun titolo	3	0,8	3	1,3
Lic. elementare	52	13,0	25	10,9
Lic. media	134	33,7	78	33,9
Diploma secondario	168	42,2	99	43,0
Diploma universitario	3	0,8	2	0,9
Laurea	26	6,5	14	6,1
Titolo post laurea	12	3,0	9	3,9
Totale	398	100,0	230	100,0

Sono state escluse le omissioni di risposta.

Tabella 6.8 – Nuovi iscritti per titolo di studio della madre in ingresso e in uscita (I anno)

	Prova di ingresso		Esiti accademici	
	Frequenza	%	Frequenza	%
Nessun titolo	1	0,2	1	0,4
Lic. elementare	53	13,2	29	12,6
Lic. media	136	33,9	80	34,5
Diploma secondario	176	43,8	99	42,9
Diploma universitario	8	2,0	5	2,2
Laurea	25	6,2	15	6,5
Titolo post laurea	3	0,7	2	0,9
Totale	402	100,0	231	100,0

Sono state escluse le omissioni di risposta.

Non si riporta il confronto relativo all'eventuale esperienza universitaria di fratelli e sorelle, in quanto i gruppi in ingresso e in uscita risultano assolutamente uguali.

Uno dei possibili fattori che potrebbe incidere sulla regolarità della frequenza, e conseguentemente sulla dispersione, è la lontananza dall'università, ma in questo caso il gruppo in uscita risulta assolutamente omogeneo al complesso degli iscritti.

Tabella 6.9 – Nuovi iscritti per lontananza dall'università in ingresso e in uscita (I anno)

	Prova di ingresso		Esiti accademici	
	Frequenza	%	Frequenza	%
<29	76	18,4	42	17,6
30-59	130	31,6	80	33,6
60-89	110	26,7	61	25,6
90-119	62	15,0	37	15,6
>120	34	8,3	18	7,6
Totale	412	100,0	238	100,0

Sono state escluse le omissioni di risposta.

Assoluta mancanza di significatività anche per le differenze di domicilio, per cui la residenza a Roma non incide assolutamente nella composizione del gruppo in uscita.

Più interessante il confronto relativo alla previsione di frequenza. In questo caso il rilevato risulta leggermente significativo (Chi quadrato sign. 0,014). In particolare cresce la percentuale degli studenti frequentanti con regolarità rispetto al valore atteso, mentre diminuiscono sia i frequentanti irregolari sia i non frequentanti.

Tabella 6.10 – Nuovi iscritti per previsione di frequenza in ingresso e in uscita (I anno)

	Prova di ingresso		Esiti accademici	
	Frequenza	%	Frequenza	%
Costante	185	44,3	118	48,8
Parziale	203	48,5	113	46,7
Non frequentante	30	7,2	11	4,5
Totale	418	100,0	242	100,0

Sono state escluse le omissioni di risposta.

Un risultato molto simile si ha considerando l'impegno lavorativo dei nuovi iscritti, per il quale si è rilevato quasi lo stesso livello di significatività (Chi quadrato sign. 0,015), con un evidente scarto negativo nel gruppo di uscita del numero di studenti lavoratori rispetto al valore atteso.

Tabella 6.11 – Nuovi iscritti per impegno lavorativo in ingresso e in uscita (I anno)

	Prova di ingresso		Esiti accademici	
	Frequenza	%	Frequenza	%
Lavoratori	179	43,1	91	38,1
Non lavoratori	236	56,9	148	61,9
Totale	415	100,0	239	100,0

Sono state escluse le omissioni di risposta.

In pratica, senza considerare le 4 omissioni di risposta alla domanda sull'impegno lavorativo, i 176 studenti non rilevati in uscita si dividono esattamente a metà fra lavoratori e non lavoratori, per cui la differenza iniziale di circa 14 punti percentuali cresce a 24 punti, mostrando un'evidente difficoltà, per alcuni studenti, di contemperare lavoro e studio.

Questa difficoltà non sembra comunque imputabile al tempo lavorativo, poiché la diminuzione dei lavoratori in uscita riguarda in misura molto simile, e comunque senza differenze significative, sia i lavoratori a tempo pieno, sia quelli a tempo parziale o saltuari.

Tabella 6.12 – Nuovi iscritti lavoratori per tempo lavorativo in ingresso e in uscita (I anno)

	Prova di ingresso		Esiti accademici	
	Frequenza	%	Frequenza	%
Tempo pieno	68	38,0	33	36,3
Tempo parziale	74	41,3	36	39,6
Lavoro saltuario	37	20,7	22	24,2
Totale	179	100,0	91	100,0

Sono state escluse le omissioni di risposta.

Il gruppo rilevato in uscita non differisce in modo sostanziale da quello in ingresso se si considerano le motivazioni che hanno portato alla scelta del corso di studi.

Tabella 6.13 – Nuovi iscritti per motivazioni di scelta del corso di studi in ingresso e in uscita (I anno)

	Prova di ingresso		Esiti accademici	
	Frequenza	%	Frequenza	%
Sbocchi professionali	202	48,2	119	49,2
Bagaglio culturale	121	28,9	71	29,3
Per esclusione	12	2,9	7	2,9
Passione	199	47,5	106	43,8
Altro	14	3,3	4	1,7

Totali percentuali superiori a 100 (possibili più risposte).

In definitiva sono poche le variabili rilevate dal questionario che si modificano in modo significativo tra il gruppo in entrata, che ha svolto la prova di ingresso, e il gruppo rilevato all'uscita del primo anno accademico: le distribuzioni relative alla precedente esperienza universitaria e al voto conseguito nell'Esame di Stato si sono modificate in modo sensibile alla fine del primo anno di studi, mentre alcuni aspetti materiali, quali gli impegni lavorativi o la possibilità di frequentare le lezioni, hanno modificato in modo appena significativo il profilo del gruppo in uscita rispetto al gruppo originale.

### 6.3. Descrizione del campione definitivo secondo le variabili della scheda di rilevazione del percorso accademico

Con la compilazione della scheda di rilevazione degli esiti accademici si è definito il campione di ricerca per la verifica dell'ipotesi. Tuttavia prima di affrontare un'analisi specifica sul rapporto esistente fra prova di ingresso e esiti accademici risulta opportuna una descrizione del campione definitivo secondo le variabili rilevate in uscita dal primo anno accademico, riguardanti il percorso di studi del primo anno di corso.

Dei 242 studenti rilevati sono 11 quelli che non hanno sostenuto alcun tipo di esami<sup>111</sup>. Per la definizione degli esiti accademici è stato utilizzato l'indicatore relativo al numero di crediti conseguiti, che diverse ricerche

<sup>111</sup> Secondo il Manifesto del Corso di laurea in Scienze dell'Educazione e della Formazione per l'anno 2003-04 gli studenti iscritti al primo anno potevano sostenere, oltre agli esami legati agli insegnamenti, anche esami di idoneità per la competenza informatica e la conoscenza della lingua straniera.

hanno attestato essere il più sensibile rispetto al numero assoluto di esami, ai voti riportati o a elaborazioni statistiche degli stessi dati. Nella Tabella 6.14 sono riassunti i crediti ottenuti nel complesso dai 231 studenti che hanno ottenuto crediti, distinti considerando gli insegnamenti tradizionali, le prove di idoneità e le idoneità conseguite tramite attestazione<sup>112</sup>.

Tabella 6.14 – Riepilogo crediti per tipologia

Statistiche riassuntive crediti I anno di corso			
Crediti	Insegnamenti	Idoneità con prova	Idoneità con attestazione
Monte crediti	7072	445	158
%	92,1	5,8	2,1
Media crediti	30,6 (1,0)	1,9 (0,2)	0,7 (0,1)
Dev std. crediti	15,5	2,8	1,8

Fra parentesi l'errore standard relativo.

Come si può notare i crediti di idoneità rappresentano una netta minoranza (7,9%) a fronte dei crediti legati a insegnamenti tradizionali, e comunque risulta utile distinguere, fra le idoneità, quelle conseguite attraverso il sostenimento di una prova, in quanto comportano di fatto un impegno simile (frequenza del corso o preparazione da non frequentante, prova di esame) a quello sostenuto per affrontare un esame tradizionale. L'attestazione di una competenza precedentemente sostenuto costituisce in vece una sorta di jolly che lo studente può giocare in qualsiasi punto del percorso accademico, e che viene spesso utilizzato per "riequilibrare" sessioni di esame con pochi risultati.

Nella Tabella 6.15 sono riportate le principali statistiche descrittive del complesso di crediti ottenuti dai 231 studenti esaminati, distinguendo nella colonna centrale i crediti ottenuti con il superamento di una prova (insegnamenti e idoneità con prova) e riportando nella colonna di destra quelli ottenuti solo con esami tradizionali.

<sup>112</sup> L'idoneità informatica si poteva conseguire, senza sostenere alcuna prova, attestando la competenza individuale specifica (es. ETCL) ; allo stesso modo era conseguibile l'idoneità in lingua straniera (es. *First certificate*).

Tabella 6.15 – Riepilogo crediti complessivi

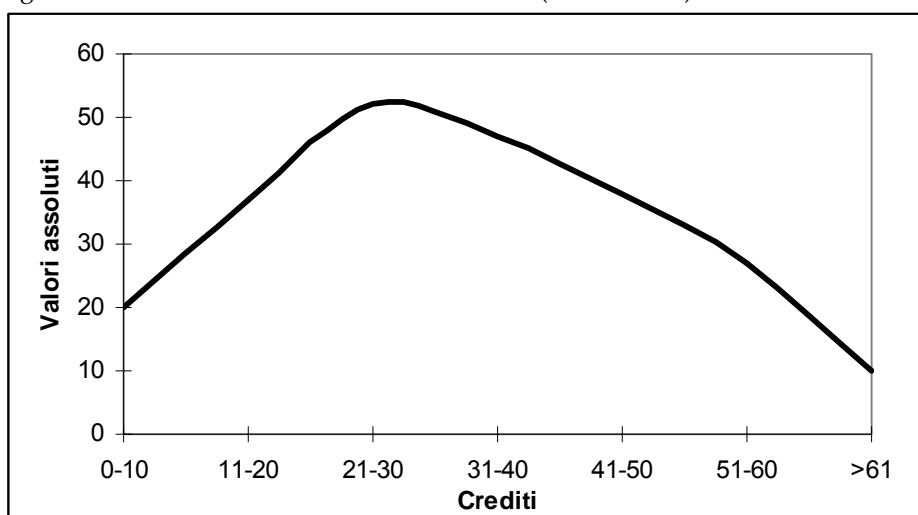
Statistiche riassuntive crediti I anno di corso			
Crediti	Totali	Solo prove sostenute	Solo insegnamenti
Monte crediti	7675	7517	7072
Media crediti	33,2 (1,1)	32,5 (1,1)	30,6 (1,0)
Dev std. crediti	16,4	16,2	15,5
C.d.V. crediti	49,4	49,8	50,7
Min crediti	2	2	2
Max crediti	86	81	81
Asimmetria	0,38	0,31	0,34
Curtosi	-0,18	-0,37	-0,17

Anche se l'indicatore normalmente usato per definire il successo accademico è il numero dei crediti ottenuto superando esami con voto, si ritiene interessante affiancare questo dato anche con quello che integra le idoneità, distinguendo quelle conseguite attraverso il superamento di prove. Gli altissimi coefficienti di variazione denotano una forte dispersione dei crediti, per cui risulta utile rappresentare graficamente la loro distribuzione procedendo a una serie di aggregazioni. Si è proceduto al raggruppamento per fasce di 10 crediti, considerando che nell'anno accademico considerato il numero di crediti per la maggior parte degli insegnamenti era di 5, per cui la seconda fascia (11-20) raccoglie gli studenti che hanno sostenuto più di 2 esami standard, la terza fascia più di 4 esami standard (o equivalenti) e così via.

In primo luogo (Figura 6.2) vediamo l'andamento relativo ai crediti totali ottenuti, comprensivi quindi di tutte le idoneità.



Figura 6.2 – Curva di distribuzione crediti totali (231 studenti)

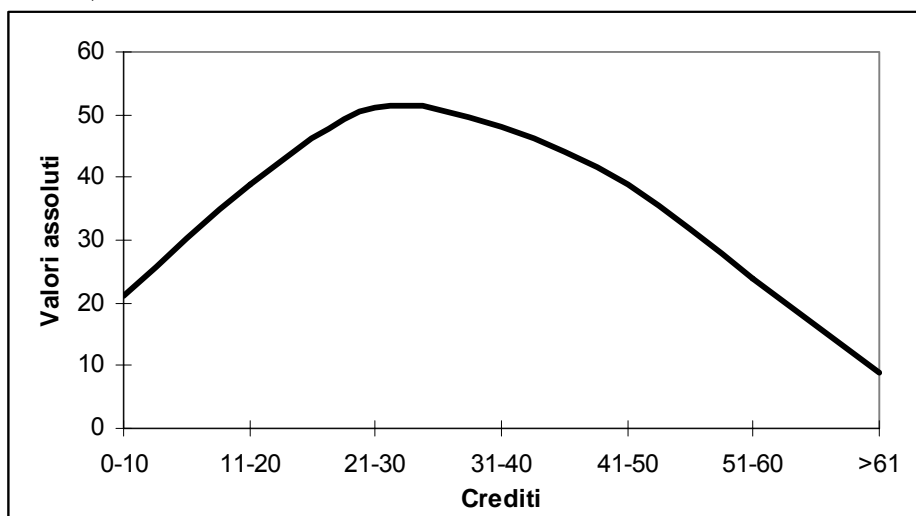


Si può osservare che l'asimmetria resta chiaramente positiva, ma con un dato (0,14) meno marcato rispetto alla distribuzione completa dei punteggi. Il gruppo modale è quello costituito dagli studenti che hanno conseguito 21-30 crediti, seguito dalla fascia immediatamente superiore. Insieme questi due gruppi rappresentano circa il 43% del totale. Il raggruppamento in fasce comporta un relativo schiacciamento della curva (la curtosi sale a  $-0,78$ ), con da una parte i 20 studenti (8,7%) che hanno conseguito un massimo di 10 crediti, e dall'altra un solido 16% che ha quasi raggiunto o addirittura superato i 60 crediti stabiliti istituzionalmente per il primo anno di corso.

Nella Figura 6.3 la curva di distribuzione tiene conto solo dei crediti conseguiti con il superamento di una prova.

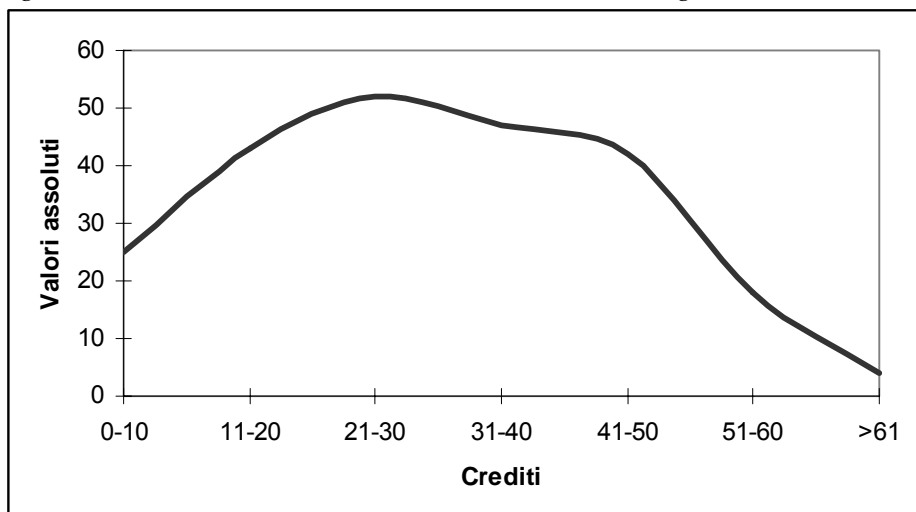
L'andamento risulta nel complesso molto simile (ricordiamo che i crediti ottenuti grazie ad attestazioni erano solo il 2,1% del totale), ma l'addolcimento del versante destro della curva mostra che la possibilità di ottenere crediti per competenza già certificate è stata sfruttata da diversi studenti già in possesso di un buon numero di crediti.

Figura 6.3 – Curva di distribuzione crediti relativi solo a prove sostenute (231 studenti)



Infine vediamo la distribuzione degli studenti prendendo in considerazione esclusivamente i crediti conseguiti in esami con voto.

Figura 6.4 – Curva di distribuzione crediti relativi solo a insegnamenti (231 studenti)

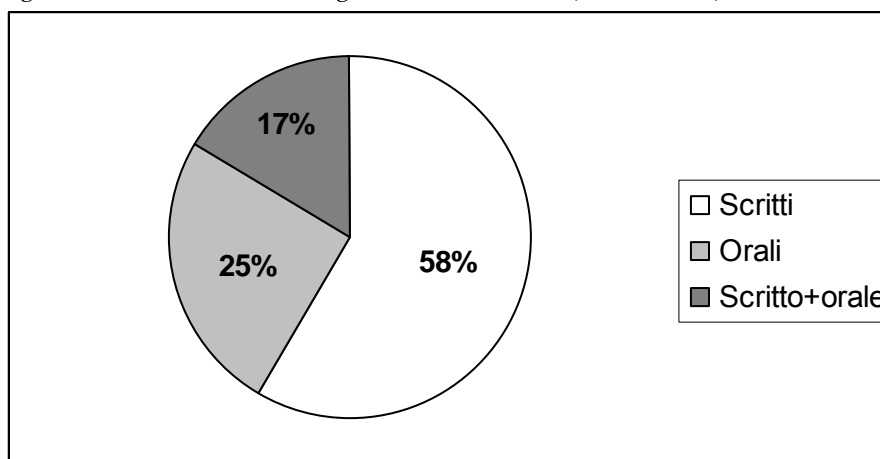


L'asimmetria risulta ancora una volta positiva ma prossima alla normalità (0,14), con un'ulteriore contrazione della fascia 51-60 che però viene assorbita del tutto dalla fascia contigua a sinistra, in tal modo i tre gruppi centrali superano nel complesso il 60%. Questa crescita centrale, unita all'incremento della fascia più bassa, appiattisce ancora di più la curva, ed infatti la curtosi riferita a questa distribuzione per fasce cresce a -0,78.

Un'altra classificazione resa possibile dai dati raccolti attraverso la scheda di rilevazione degli esiti riguarda le modalità di superamento degli esami (escluse le idoneità), distinguendo quelli che si sono esauriti in una prova scritta, quelli superati solo con un'interrogazione orale, quelli che hanno comportato entrambi i tipi di verifica.

Poiché 16 studenti non hanno compilato questa sezione della scheda, i dati si riferiscono a 6639 crediti su 7072.

Figura 6.5 – Modalità di conseguimento dei crediti (215 studenti)



Come si può notare la percentuale complessiva di crediti ottenuti grazie al superamento di uno scritto, nella maggior parte senza il riscontro di una prova orale, è ben del 75%. Si tratta di un dato forse sorprendente, ma ricordiamo che il Corso di laurea in Scienze dell'Educazione e della Formazione oggetto di indagine era giunto al terzo anno di vita, con un numero complessivo di iscritti avviato verso il migliaio e supportato ancora dalle limitate risorse umane e strutturali che ne avevano permesso l'apertura nell'A.A. 2001-02, per cui l'assiduo utilizzo delle

prove di verifica scritte è apparsa subito come una scelta strategica per garantire un controllo fluido della didattica.

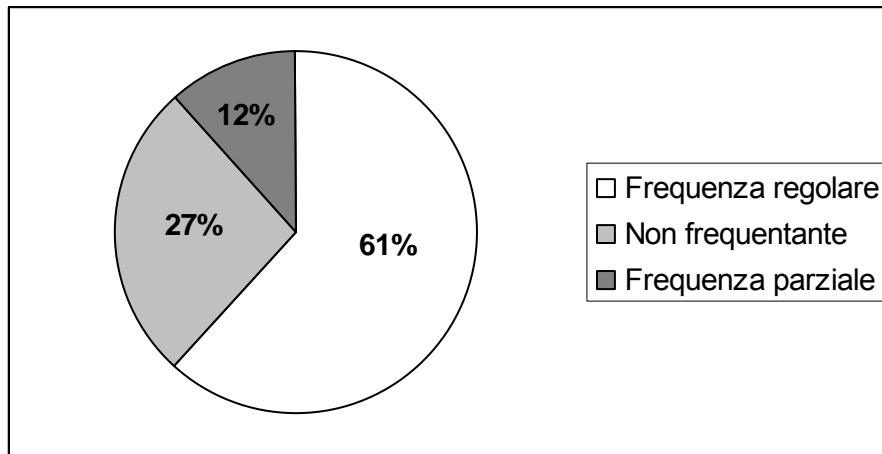
Nel complesso sono stati solamente 8 gli studenti che hanno conseguito crediti esclusivamente con prove orali (7 di questi si trovano nella fascia più bassa dei crediti ottenuti). Ma anche se la stragrande maggioranza degli studenti ha guadagnato crediti superando le prove scritte, tre quarti di loro ha anche affrontato almeno una prova esclusivamente orale, e più della metà è stata impegnata anche nella modalità mista. Nella Tabella 6.16 sono considerate le percentuali di studenti che hanno affrontato le tre tipologie di esame.

Tabella 6.16 – Riepilogo modalità di conseguimento dei crediti (215 studenti)

	Frequenza	%
Scritti	201	93,5
Orali	161	74,9
Scritto + orale	120	55,8

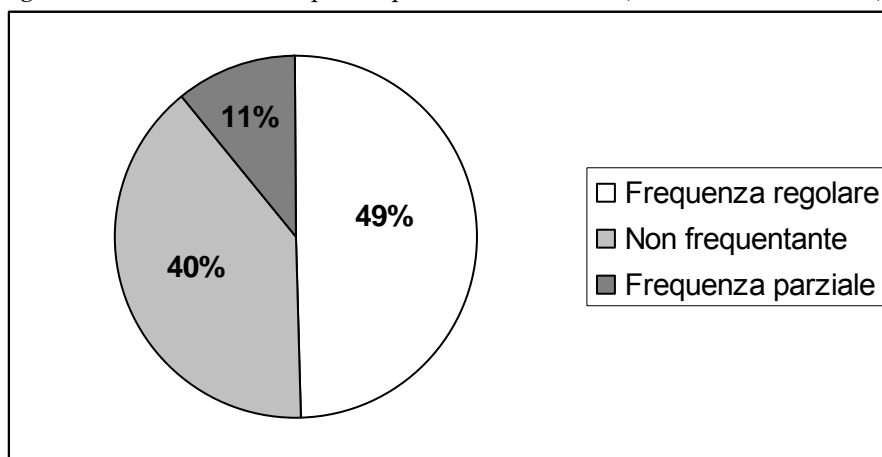
Un altro elemento da considerare per comprendere l'andamento del primo anno di corso riguarda il rapporto fra crediti e frequenza delle lezioni. Una sezione della scheda di rilevazione degli esiti richiedeva agli studenti di dichiarare, per ogni esame sostenuto, se fosse stato seguito il relativo corso, prevedendo anche la possibilità di una frequenza parziale (tra il 40% e il 60% delle lezioni). In questo caso hanno compilato la sezione 214 studenti, per cui l'analisi del rapporto crediti-frequenza riguarda 6629 crediti sui 7072 complessivi ottenuti in esami con voti.

Figura 6.6 – Modalità di frequenza per i crediti ottenuti (214 studenti)



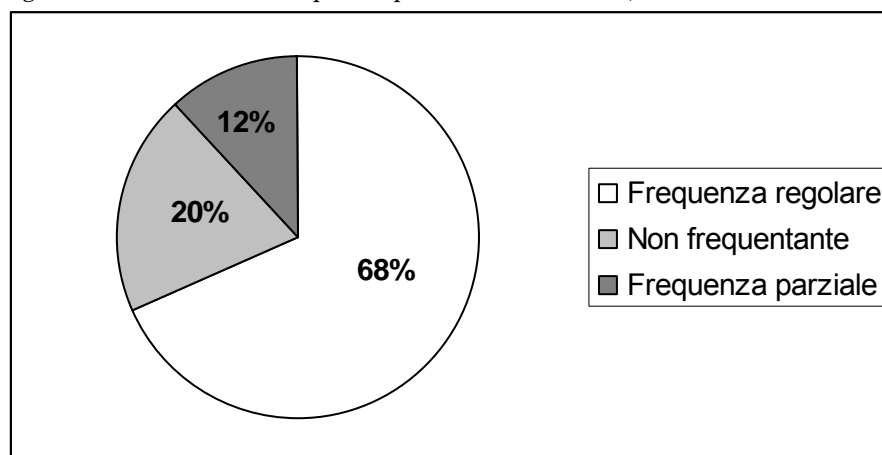
La percentuale di lezioni frequentate, che hanno comportato crediti conseguiti risulta più bassa rispetto a quella rilevata nello stesso corso di laurea nel 2001-02 (72,8%), ma la differenza sembra aumentare ulteriormente considerando che allora la percentuale di studenti lavoratori (61,5%) era notevolmente più alta del gruppo preso in esame ora (38,1%). Se verifichiamo la frequenza degli studenti lavoratori vediamo infatti che la percentuale dei crediti ottenuti con frequenza regolare scende sotto il 50%.

Figura 6.7 – Modalità di frequenza per i crediti ottenuti (80 studenti lavoratori)



Naturalmente le percentuali relative alla frequenza degli studenti non lavoratori crescono, ma non raggiungono comunque il dato rilevato in precedenza, per cui si può affermare che la generazione di nuovi iscritti considerata tende a diminuire l'impegno, almeno per quanto riguarda la frequenza delle lezioni, e in particolare risulta vincolante l'impegno lavorativo.

Figura 6.8 – Modalità di frequenza per i crediti ottenuti (134 studenti non lavoratori)



Anche in questo caso è opportuna una tabella riassuntiva per capire quanti studenti hanno seguito almeno un corso da frequentante regolare. Solo poco più del 10% non ha seguito alcun corso in maniera regolare, ma di contro si può anche osservare che quasi due terzi del gruppo osservato ha sostenuto almeno un esame da non frequentante.

Tabella 6.17 – Riepilogo modalità di frequenza per i crediti (214 studenti)

	Frequenza	%
Frequenza regolare	191	89,3
Non frequentante	138	64,5
Frequenza parziale	94	43,9

Il controllo preventivo del gruppo definitivo di indagine, determinato dal numero di studenti che ha effettivamente compilato la scheda di rilevazione del percorso accademico, ha evidenziato una buona rappresentatività rispetto al gruppo originale di partecipanti alla prova di ingresso. Inoltre per le principali variabili si è accertata un'articolazione

delle frequenze che assicura una accettabile varianza e conseguentemente favorisce la successiva analisi per gruppi.

#### **6.4 Relazioni esistenti fra i risultati della prova di ingresso e gli esiti accademici**

In questo paragrafo affrontiamo il riscontro più importante relativo all'ipotesi formulata in apertura del lavoro di ricerca, cioè la capacità predittiva delle graduatorie di accesso rispetto ai risultati conseguiti nel corso degli studi universitari.

Dopo aver esposto le modalità di costruzione delle graduatorie (cfr. 5.6) e la definizione degli indicatori di successo accademico (cfr. 6.3), possiamo verificare la qualità delle relazioni esistenti, considerando come fattore di riscontro anche una variabile già rilevata nel questionario di ingresso, il voto conseguito nell'Esame di Stato.

Si ritiene opportuno raccogliere tutta l'analisi delle correlazioni in una sola tabella riassuntiva, considerando insieme tutte le diverse variabili in gioco. In tal modo sarà possibile un confronto immediato. La complessità della tabella prodotta rende però necessaria una adeguata esplicitazione preventiva delle etichette.

Nella Tabella 6.18 sono illustrate nella colonna di destra le successive etichette relative all'analisi delle correlazioni, nella colonna di sinistra le descrizioni complete delle variabili utilizzate, riprendendo le definizioni già adottate in precedenza. Il primo gruppo di variabili fa riferimento al numero di crediti conseguiti nel corso del primo anno, subito seguito dall'unica variabile di sfondo in ingresso (il voto nell'Esame di Stato). Il secondo gruppo di variabili fa riferimento ai risultati della prova di comprensione TCL, secondo le tre diverse modalità di analisi utilizzate. Nello stesso modo è articolato il successivo gruppo di tre variabili riferite alla prova di comprensione Cloze. Infine sono stati considerati anche i punteggi complessivi nelle due prove, sommando i punti T di ogni singola prova, distinti sempre secondo le tre modalità di analisi.

Tabella 6.18 – Riepilogo delle variabili considerate per l'analisi delle correlazioni

Etichetta	Descrizione della variabile
Crediti (solo esami)	Numero dei crediti riferiti a esami di insegnamenti
Cred+Ido (con prova)	Somma di crediti di esami e idoneità conseguite attraverso il superamento di una prova
Cred+Ido	Somma complessiva dei crediti (comprese tutte le idoneità)
Voto Esame	Voto riportato nell'Esame di Stato
TCL Classica	Punteggio riportato nel TCL secondo l'item analisi Classica
TCL IRT	Punteggio riportato nel TCL secondo l'item analisi IRT tradizionale
TCL IRT Om	Punteggio riportato nel TCL secondo l'item analisi IRT alternativa (penalizzazione di tutte le omissioni)
Cloze Classica	Punteggio riportato nel Cloze secondo l'item analisi Classica
Cloze IRT	Punteggio riportato nel Cloze secondo l'item analisi IRT tradizionale
Cloze IRT Om	Punteggio riportato nel Cloze secondo l'item analisi IRT alternativa (penalizzazione di tutte le omissioni)
TCL+Cloze Classica	Somma dei punteggi riportati nelle due prove secondo l'item analisi Classica
TCL+Cloze IRT	Somma dei punteggi riportati nelle due prove secondo l'item analisi IRT tradizionale
TCL+Cloze IRT Om	Somma dei punteggi riportati nelle due prove secondo l'item analisi IRT alternativa (penalizzazione di tutte le omissioni)

Si tratta evidentemente di tutte variabili riferite a scale numeriche sulle quali è possibile effettuare controlli di correlazione. Per favorire la lettura la Tabella 6.19 è stata parzialmente tagliata rispetto al consueto formato a matrice, evidenziando le colonne relative ai dati di successo accademico e alla variabile di sfondo in ingresso. Ricordiamo che le correlazioni fra le diverse graduatorie erano già state verificate nel Paragrafo 5.5.

La quasi totalità delle correlazioni si riferisce a tutti i 242 studenti del campione definitivo, con l'eccezione delle correlazioni riguardanti l'Esame di Stato, da cui sono stati eliminati 8 studenti che non avevano indicato il voto conseguito.



Tabella 6.19– Riepilogo delle correlazioni fra esiti e prove di ingresso

		Crediti (solo esami)	Cred+Ido (con prova)	Cred+Ido	Voto Esame
Crediti (solo esami)	Corr. di Pearson	1	,987(**)	,986(**)	,256(**)
	Sig. (2-code)		,000	,000	,000
	N	242	242	242	234
Cred+Ido (con prova)	Corr. di Pearson	,987(**)	1	,995(**)	,247(**)
	Sig. (2-code)	,000		,000	,000
	N	242	242	242	234
Cred+Ido	Corr. di Pearson	,986(**)	,995(**)	1	,253(**)
	Sig. (2-code)	,000	,000		,000
	N	242	242	242	234
Voto Esame	Corr. di Pearson	,256(**)	,247(**)	,253(**)	1
	Sig. (2-code)	,000	,000	,000	
	N	234	234	234	234
TCL Classica	Corr. di Pearson	,157(*)	,173(**)	,167(**)	,275(**)
	Sig. (2-code)	,014	,007	,009	,000
	N	242	242	242	234
TCL IRT	Corr. di Pearson	,101	,114	,118	,249(**)
	Sig. (2-code)	,117	,076	,067	,000
	N	242	242	242	234
TCL IRT Om	Corr. di Pearson	,130(*)	,145(*)	,141(*)	,274(**)
	Sig. (2-code)	,043	,024	,028	,000
	N	242	242	242	234
Cloze Classica	Corr. di Pearson	,040	,049	,044	,236(**)
	Sig. (2-code)	,534	,450	,491	,000
	N	242	242	242	234
Cloze IRT	Corr. di Pearson	,038	,047	,045	,246(**)
	Sig. (2-code)	,556	,464	,489	,000
	N	242	242	242	234
Cloze IRT Om	Corr. di Pearson	,041	,050	,047	,250(**)
	Sig. (2-code)	,523	,439	,468	,000
	N	242	242	242	234
TCL + Cloze Classica	Corr. di Pearson	,120	,134(*)	,128(*)	,306(**)
	Sig. (2-code)	,063	,037	,046	,000
	N	242	242	242	234
TCL + Cloze IRT	Corr. di Pearson	,084	,097	,098	,296(**)
	Sig. (2-code)	,193	,131	,128	,000
	N	242	242	242	234
TCL + Cloze IRT Om	Corr. di Pearson	,106	,120	,116	,313(**)
	Sig. (2-code)	,101	,062	,073	,000
	N	242	242	242	234

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

\* La correlazione è significativa al livello 0,05 (2-code).

Le tre modalità di articolazione dei crediti, che distinguono quelli conseguiti solo in esami riferiti a insegnamenti, quelli comunque ottenuti attraverso il superamento di una prova (idoneità comprese) e quelli complessivi, mostrano forti correlazioni significative fra loro.

Allo stesso modo mostrano una significativa correlazione con il voto dell'Esame di Stato, anche se la forza delle relazioni non risulta particolarmente marcata, oscillando intorno a 0,25.

Più articolato il quadro delle correlazioni con le prove di comprensione. I punteggi TCL Classica presentano una correlazione significativa ( $<0,01$ ) con i crediti complessivi e i crediti ottenuti con superamento di una prova. Si tratta in quest'ultimo caso della migliore correlazione che riguarda le prove di ingresso, pur trattandosi di un dato sostanzialmente debole (0,17). La correlazione con i crediti dei soli esami scende leggermente anche come significatività ( $<0,05$ ) ma si conferma la migliore tra quelle rilevate per le diverse prove di comprensione.

I punteggi TCL derivati dall'item analisi IRT tradizionale risultano invece tutti non significativi, anche se le correlazioni con crediti complessivi e crediti con prova sfiorano il primo livello di significatività.

I punteggi TCL ottenuti con l'analisi IRT alternativa risultano invece avere una correlazione debolmente significativa ( $<0,05$ ) con tutte e tre le articolazioni dei crediti, con valori non troppo lontani (0,13 - 0,14) rispetto a quelli rilevati per i punteggi TCL Classica.

Nessuna correlazione, neanche debolmente significativa, rilevata invece per le graduatorie Cloze, a conferma delle criticità rilevate in precedenza (cfr. 5.5.2).

Le correlazioni relative ai punteggi sommati delle due prove risultano appena significative ( $<0,05$ ) solo per le graduatorie Classiche, ma non rispetto ai soli crediti conseguiti con esami legati a insegnamenti.

Da notare che il voto nell'Esame di Stato risulta correlato in modo significativo anche con tutte le graduatorie, confermando sostanzialmente i valori già rilevati per l'intero gruppo dei partecipanti alla prova di ingresso (cfr. Tabella 5.43). E' interessante notare che però il voto nell'esame correla maggiormente con le graduatorie che sommano le due prove di comprensione e, tra le modalità di analisi, con quella IRT alternativa.

### 6.5. Confronto fra i sottogruppi determinati dalla prova di ingresso

Come già fatto a proposito del confronto delle graduatorie per l'intero gruppo che ha partecipato alla prova di ingresso, risulta opportuno far seguire all'analisi delle correlazioni fra graduatorie di ingresso ed esiti accademici un controllo dei due gruppi che si sono creati in seguito alla soglia simulata per l'accesso. Vediamo come si distribuiscono i due gruppi considerando solo i 242 studenti di cui si sono rilevati gli esiti in uscita dal primo anno. Come già fatto a proposito del confronto delle graduatorie (cfr. 5.6), vengono considerati solo i risultati relativi alla prova TCL.

Tabella 6.20 – Riepilogo composizione gruppi in uscita secondo le diverse graduatorie (242 studenti)

	TCL Classica		TCL IRT		TCL Om	
	Frequenza	%	Frequenza	%	Frequenza	%
Ammessi	89	36,8	86	35,5	90	37,2
Esclusi	153	63,2	156	64,5	152	67,8
Totale	242	100,0	242	100,0	242	100,0

Non si riscontrano sostanziali differenze fra i gruppi determinati dalle diverse graduatorie, con circa due terzi del numero complessivo di studenti in uscita composto da studenti che risultavano esclusi secondo i risultati del TCL. Di fatto viene mantenuto per i due gruppi lo stesso rapporto rilevato fra i sottogruppi in ingresso, composto da 150 studenti ammessi e 269 esclusi.

Nella Tabella 6.21 sono riportate le percentuali rilevate in uscita rispetto ai sottogruppi originali.

Tabella 6.21 – Riepilogo composizione gruppi in uscita rispetto ai gruppi in entrata (242 studenti)

	TCL Classica		TCL IRT		TCL Om	
	Frequenza	%	Frequenza	%	Frequenza	%
Ammessi(150)	89	59,3	86	57,3	90	60,0
Esclusi (269)	153	56,9	156	58,0	152	56,6

Far parte del gruppo degli ammessi non ha quindi comportato particolari vantaggi in termini di completamento del primo anno di corso,

con percentuali simili di abbandono rispetto al gruppo dei potenziali esclusi.

Passiamo alla verifica dei risultati conseguiti dai due gruppi in termini di crediti accademici, attraverso la procedura di confronto delle medie con relativa analisi della varianza. Nella Tabella 6.22 sono state mantenute le stesse etichette definite in precedenza. Cominciamo dai gruppi determinati dalla graduatoria Classica.

Tabella 6.22 – Confronto delle medie dei gruppi in uscita – TCL Classico (242 studenti)

		Crediti (solo esami)	Cred+Ido (con prova)	Cred+Ido
Esclusi	Media	27,68	29,18	29,87
	N	153	153	153
	Deviazione std.	16,316	16,946	17,368
Ammessi	Media	31,88	34,29	34,89
	N	89	89	89
	Deviazione std.	16,446	17,248	17,433
Totale	Media	29,22	31,06	31,71
	N	242	242	242
	Deviazione std.	16,455	17,200	17,524

Tabella 6.22bis – ANOVA relativa al confronto delle medie dei gruppi in uscita – TCL Classico (242 studenti)

		Somma dei quadrati	df	Media dei quadrati	F	Sig.
Crediti (solo esami) TCL Classica	Fra gruppi	991,003	1	991,003	3,701	,056
	Entro gruppi	64266,948	240	267,779		
	Totale	65257,950	241			
Cred+Idoprova (con prova) TCL Classica	Fra gruppi	1468,790	1	1468,790	5,048	,026
	Entro gruppi	69827,280	240	290,947		
	Totale	71296,070	241			
Cred+Ido TCL Classica	Fra gruppi	1417,064	1	1417,064	4,685	,031
	Entro gruppi	72594,262	240	302,476		
	Totale	74011,326	241			

Lo scarto rilevato risulta a favore del gruppo degli ammessi con qualsiasi modalità di calcolo dei crediti conseguiti, ma risulta significativo solo per le due ultime categorie (crediti conseguiti con un prova e totale

dei crediti), con un valore di F sufficientemente alto e una significatività  $<0,05$ . In valore assoluto significano oltre cinque crediti di più in media conseguiti dagli studenti potenzialmente ammessi. Vanno però sottolineati gli alti valori di deviazione standard di tutte le misure medie, segno che i due gruppi, al pari del gruppo complessivo degli studenti in uscita, è caratterizzato da una accentuata varianza interna.

Passiamo ora al confronto fra i gruppi definiti dall'item analisi IRT tradizionale.

Tabella 6.23 – Confronto delle medie dei gruppi in uscita – TCL IRT (242 studenti)

		Crediti (solo esami)	Cred+Ido (con prova)	Cred+Ido
Esclusi	Media	28,28	29,94	30,52
	N	156	156	156
	Deviazione std.	16,785	17,395	17,813
Ammessi	Media	30,94	33,09	33,88
	N	86	86	86
	Deviazione std.	15,791	16,748	16,874
Totale	Media	29,22	31,06	31,71
	N	242	242	242
	Deviazione std.	16,455	17,200	17,524

Tabella 6.23bis – ANOVA relativa al confronto delle medie dei gruppi in uscita – TCL IRT (242 studenti)

		Somma dei quadrati	df	Media dei quadrati	F	Sig.
Crediti (solo esami) TCL Classica	Fra gruppi	394,094	1	394,094	1,458	,228
	Entro gruppi	64863,857	240	270,266		
	Totale	65257,950	241			
Cred+Idoprova (con prova) TCL Classica	Fra gruppi	550,334	1	550,334	1,867	,173
	Entro gruppi	70745,737	240	294,774		
	Totale	71296,070	241			
Cred+Ido TCL Classica	Fra gruppi	627,547	1	627,547	2,052	,153
	Entro gruppi	73383,780	240	305,766		
	Totale	74011,326	241			

In questo caso le differenze esistenti fra i due gruppi non risultano significative per nessuna modalità di calcolo dei crediti. Le medie risultano sempre a vantaggio del gruppo degli ammessi, ma con scarti ridotti a

tre punti circa. La minore capacità predittiva della graduatoria formulata secondo l'item analisi tradizionale viene confermata anche dal confronto fra i gruppi.

Completiamo questa analisi con i gruppi determinati dall'item analisi IRT con penalizzazione indistinta delle omissioni.

Tabella 6.24 – Confronto delle medie dei gruppi in uscita – TCL IRT Om (242 studenti)

		Crediti (solo esami)	Cred+Ido (con prova)	Cred+Ido
Esclusi	Media	27,61	29,16	29,85
	N	152	152	152
	Deviazione std.	16,261	16,912	17,340
Ammessi	Media	31,94	34,28	34,87
	N	90	90	90
	Deviazione std.	16,514	17,296	17,478
Totale	Media	29,22	31,06	31,71
	N	242	242	242
	Deviazione std.	16,455	17,200	17,524

Tabella 6.24bis – ANOVA relativa al confronto delle medie dei gruppi in uscita – TCL IRT Om (242 studenti)

		Somma dei quadrati	df	Media dei quadrati	F	Sig.
Crediti (solo esami) TCL Classica	Fra gruppi	1061,130	1	1061,130	3,967	,048
	Entro gruppi	64196,821	240	267,487		
	Totale	65257,950	241			
Cred+Idoprova (con prova) TCL Classica	Fra gruppi	1481,804	1	1481,804	5,094	,025
	Entro gruppi	69814,266	240	290,893		
	Totale	71296,070	241			
Cred+Ido TCL Classica	Fra gruppi	1423,407	1	1423,407	4,706	,031
	Entro gruppi	72587,920	240	302,450		
	Totale	74011,326	241			

Lo scarto fra i due gruppi torna a crescere, dagli oltre 4 punti di differenza per i soli crediti da esami, agli oltre 5 per le altre due modalità di calcolo, ma soprattutto la differenza delle medie risulta sempre significativa (<0,05), laddove anche per i gruppi determinati dall'item analisi classica il confronto scendeva sotto la soglia di significatività.

Certo si tratta di differenze minime (la mancata significatività relativa ai crediti con esami era solo a 0,056), ma denotano in questo caso una leggera migliore capacità selettiva della graduatoria IRT Om<sup>113</sup>, riconducibile alla migliore capacità discriminativa di tale modalità di analisi in corrispondenza della fascia di taglio dei punteggi.

## 6.5. Verifica dell'ipotesi di ricerca

Riportiamo di seguito l'ipotesi di ricerca che ha guidato il lavoro di ricerca.

L'attribuzione dei punteggi di una prova di ingresso svolta secondo il modello di item analisi IRT permette di ottenere una graduatoria degli ammessi a un corso di laurea più precisa nella scala dei punteggi e maggiormente predittiva del successo accademico rispetto a una graduatoria ottenuta con il modello di item analisi classica.

La prima parte dell'ipotesi, relativa alla maggiore precisione dei punteggi in relazione alla funzione selettiva della prova risulta confermata. Le graduatorie ottenute attraverso le stime della competenza di lettura degli studenti svolte con l'analisi IRT dei quesiti risultano maggiormente discriminanti, riducono fortemente il fenomeno dei punteggi *ex aequo*, costringendo al ricorso a criteri esterni alla prova solo in rarissimi casi, laddove la graduatoria costruita secondo l'item analisi classica, che considera solo le risposte corrette attribuendo a tutte lo stesso peso, crea una scala di valori discreti, con affollamenti di punteggi uguali, soprattutto nelle fasce di distribuzione centrali, dove generalmente vengono definite le soglie di accesso. Questo può costringere al ricorso massiccio a criteri esterni alla prova (voto riportato nell'Esame di Stato, età anagrafica) per decidere dell'ammissione a un corso di studi. Bisogna segnalare però la differenza esistente fra i due metodi IRT utilizzati: la graduatoria ottenuta con l'analisi IRT tradizionale modifica sostanzialmente la posizione di alcuni studenti (caratterizzati dall'aver rinunciato a rispondere agli ultimi quesiti della prova), mentre la modalità IRT alternativa utilizzata

<sup>113</sup> I due gruppi di ammessi sono molto simili, con un solo studente degli 89 ammessi secondo la Classica che non compare fra i 90 ammessi nella graduatoria IRT Om, e due studenti ammessi di quest'ultima graduatoria che risultano esclusi da quella Classica. Si tratta però in tutti e tre i casi proprio di studenti ammessi o esclusi considerando criteri esterni alla prova.

nell'indagine (che prevede il trattamento indistinto delle omissioni) produce una graduatoria che ricalca maggiormente l'andamento della graduatoria da item analisi classica

La seconda parte dell'ipotesi invece, secondo i dati raccolti, non può essere del tutto confermata, in quanto le graduatorie costruite secondo l'item analisi classica, e in particolare quella relativa alla tradizionale prova di comprensione della lettura, risultano essere ancora quelle più correlate con gli esiti accademici, rappresentati dai crediti conseguiti al termine del I anno di corso. Tuttavia, anche in questo caso, va operata una distinzione fra i due modelli di item analisi IRT adottati: mentre la graduatoria IRT tradizionale non risulta correlata in modo significativo con gli esiti accademici, quella costruita tenendo conto del trattamento indistinto delle omissioni risulta debolmente significativa, ma di poco inferiore alla correlazione rilevata per la graduatoria dell'item analisi classica.

Se però consideriamo il confronto fra i due gruppi originati dalla procedura di selezione simulata, il confronto delle medie per i due gruppi definiti secondo la graduatoria IRT Om attesta le differenze più significative, anche se con scarti ridotti rispetto ai due gruppi definiti secondo l'item analisi classica. Confermati invece anche in questo caso i dati di significatività nulla per la graduatoria IRT tradizionale.



## Capitolo settimo

### Conclusioni e prospettive di ricerca

#### 7.1. Per una selezione consapevole, se proprio necessaria

Concludiamo questo lavoro ricollegandoci alla suggestione da cui era partito, impersonata dall'esperienza pedagogica di ricerca di Aldo Visalberghi. Piero Bertolini nel saggio *Aldo Visalberghi tra esperienza e valutazione*<sup>114</sup>, per sottolineare l'attualità del pensiero dello studioso triestino, riprende integralmente un personale saggio di commento<sup>115</sup> al volume *Esperienza e valutazione*, che aveva segnato profondamente il dibattito intorno alla pedagogia come scienza alla fine degli anni Cinquanta in Italia. L'approccio della ricerca al tema della valutazione diventa, secondo Bertolini, un paradigma di riferimento per impostare la ricerca in qualsiasi campo del sapere, in particolare per il ruolo decisivo riconosciuto all'esperienza diretta.

Non si tratta, infatti, di assumere come valido un particolare concetto di esperienza o addirittura un particolare sistema filosofico in base al quale dedurre poi tutta una serie di implicazioni nei singoli settori della ricerca filosofica; ma si tratta piuttosto di adottare il criterio operativo per il quale si parte sempre *in medias res*, cioè "da situazioni storiche reali, si tenta di operare in esse, e solo in seguito, sulla base delle esigenze emerse in quel lavoro concreto, si possono rettificare gli strumenti impiegati, per collaudarli poi ancora in situazioni culturali specifiche"

Il brano citato da Bertolini a completamento del suo pensiero è tratto proprio da *Esperienza e valutazione* di Visalberghi (1958), e rappresenta efficacemente lo spirito che ha guidato questo lavoro di ricerca, originato da una urgenza materiale rilevata nel Corso di laurea in Scienze

<sup>114</sup> Tratto dal volume *Evaluation – Studi in onore di Aldo Visalberghi* (Cives, Corda Costa, Fattori e Siciliani De Cumis 2002)

<sup>115</sup> Apparso sulla rivista *Aut Aut* nel settembre 1958.

dell'Educazione e della Formazione dell'Università di Roma, e destinato a riflettere immediatamente i suoi esiti nella pratica didattica quotidiana dello stesso Corso di laurea, cercando di gestire nel miglior modo possibile una situazione che qualsiasi docente vorrebbe non affrontare: la selezione di accesso all'università, che finisce inevitabilmente per diventare una forma di limitazione del diritto allo studio.

Se superiamo per un attimo il problema e consideriamo inevitabile il ricorso a procedure di selezione per l'accesso universitario<sup>116</sup>, appare evidente che qualsiasi strumento piegato a tale scopo risulti poco efficace. La migliore predittività riconosciuta al voto riportato nell'Esame di Stato (Andreani Dentici e Amoretti 2000) risulta tale solo perché si trova ad affrontare indicatori ancora più scadenti. Proseguire sulla strada del miglioramento delle prove di accesso risulta molto simile all'incamminarsi in una larga autostrada padana, in cui non solo non si scorgono gli ostacoli, ma si perdono anche i riferimenti laterali: se gli ostacoli rappresentano le caratteristiche psicometriche di strumenti nati per uso didattico e riconvertiti ad uso "bellico", le molte carreggiate simboleggiano l'ampiezza dei problemi di validità della prova, cioè quello che è necessario rilevare per decidere del futuro di studio dei giovani che escono dal ciclo secondario dell'istruzione.

Tanto più si cerca di risolvere il problema ampliando la quantità di competenze rilevate, secondo una logica di maggiore spettro di capacità indagate, tanto più si sbanda sul terreno scivoloso delle competenze trasversali, anche di tipo affettivo: ma è corretto selezionare (per accedere a un percorso di studio) dando ampio peso alla capacità di autostima degli studenti o alla capacità di resistere allo stress da verifica? I dubbi aumentano se si considera che probabilmente chi organizza prove di accesso complesse neanche considera l'impatto di tali variabili trasversali sul rendimento degli studenti.

D'altra parte l'affinamento tecnico delle prove, tentato anche in questa ricerca, trova presto i suoi limiti, anche in questo caso legati alla difficile definizione delle competenze necessarie per l'accesso e al complesso rapporto tra queste e l'effettivo successo accademico.

<sup>116</sup> In questa fase le esigenze di selezione risultano più dipendenti da fattori strutturali delle università che a un effettiva programmazione dei titoli di studio rispetto ai bisogni del mercato del lavoro (CNVSU 2004)

Certo sono possibili diversi approcci per migliorare la capacità selettiva delle prove di accesso. Da una parte una strumentazione mista, basata su prove standardizzate e sulla considerazione del percorso precedente di studio, sintetizzato dal voto nell'Esame di Stato; anche in questa indagine si è verificata la consueta correlazione tra questo fattore e il successo accademico, ma si pongono necessariamente problemi di peso delle diverse componenti. La raccolta regolare di dati riguardanti le prove di accesso costituisce un presupposto fondamentale per la definizione di una serie di ipotesi operative, sempre da verificare sul campo, sull'equilibrio fra esperienza di studio precedente e situazione in ingresso. Come già segnalato in premessa, si tratta proprio di uno dei filoni di ricerca su cui si muove il Dottorato di ricerca consortile in Pedagogia Sperimentale nel cui ambito si è svolto il presente lavoro.

Oppure si può tentare la strada della verifica delle motivazioni che guidano la scelta del corso di studi, attraverso strumenti prevalentemente aperti (sul modello di quello sperimentato dalla collega Truffelli per il suo lavoro di dottorato): l'indubbio fascino di questa prospettiva è rappresentato dalla forte componente di analisi qualitativa richiesta per questo tipo di strumenti.

La crescente insofferenza verso gli strumenti valutativi di tipo oggettivo è rappresentata in modo efficace da McDonnell (2004) quando analizza il forte stress introdotto dall'uso/abuso di prove oggettive di valutazione nel percorso didattico, e pur riferendosi in modo specifico alla situazione statunitense, disegna un sentire che si sta diffondendo in molti sistemi formativi occidentali<sup>117</sup>. Ma dare maggior spazio a strumenti di tipo psicologico non pone certo minori problemi di attribuzione di valore con funzione selettiva, oltre naturalmente a creare problemi operativi di gestione per la fase di raccolta delle informazioni.

Sicuramente questa prospettiva favorisce un uso maggiormente orientativo delle prove di accesso, permettendo sia ai singoli studenti, sia alla struttura universitaria, di acquisire e aggiornare informazioni per migliorare la propria preparazione e l'offerta didattica.

<sup>117</sup> Ricordiamo il sorprendente esito dell'indagine OCSE-PISA in Gran Bretagna, i cui risultati non sono stati inclusi nel confronto internazionale per problemi di copertura del campione, dovuti a una crescita dei rifiuti di scuole e studenti a partecipare alla rilevazione (OECD 2004). Una riflessione interessante sul rapporto distorto che si crea fra valutazione e didattica è quella di Graves (2002).

Nel nostro piccolo, rappresentato da questo lavoro di ricerca e dalla costante attenzione dimostrata dal Corso di laurea in Scienze dell'Educatione e della Formazione dell'Università di Roma "La Sapienza", continueremo a seguire il percorso accademico dei nuovi iscritti che hanno affrontato la prova di accesso (simulato) per l'A.A. 2003-04, rilevando via via gli esiti parziali, riferiti ai successivi anni di corso, il completamento del percorso, comprese le sue modalità (tempo di realizzazione, media dei voti, voto finale) e gli abbandoni, cercando per questi ultimi di indagare anche sui possibili esiti alternativi, in cui l'esperienza di studio, sia portata a compimento in un'altra realtà, sia interrotta per motivi personali, abbia prodotto dei risultati personali. Il controllo degli esiti accademici su tempi più lunghi rispetto al solo I anno considerato in questo lavoro, risulta fondamentale anche per verificare la migliore predittività attestata per il voto nell'Esame di Stato e sicuramente favorita in un confronto ravvicinato nel tempo<sup>118</sup>.

In tal modo sarà possibile sfruttare fino in fondo l'aspetto più originale di questo studio, cioè la possibilità di seguire i risultati di un gruppo di studenti che potevano risultare esclusi da un corso di laurea per effetto di una prova selettiva.

## **7.2. Quando la prova è diventata veramente selettiva**

A conferma dello spirito che ha animato questo lavoro di ricerca nel corso dell'anno accademico in cui si era svolta la rilevazione simulata in ingresso, il Consiglio del Corso di laurea in Scienze dell'Educatione e della Formazione di Roma "La Sapienza" ha effettivamente dovuto introdurre l'iscrizione a numero programmato per l'A.A. 2004-05.

La decisione, molto sofferta, è stata presa in considerazione dell'elevato numero di iscritti raggiunto dopo i primi tre anni di corso (oltre 1000) e considerando che ancora il numero dei laureati, al termine del primo triennio di attivazione del corso di laurea, risulta basso. Inoltre è facilmente prevedibile la conseguente crescita del numero di iscritti al Corso di laurea specialistica collegato, che si appoggia sulle stesse, scarse, strutture della facoltà di Filosofia.

Il limite fissato per il numero programmato è stato di 150 iscritti, ed è stato deciso di utilizzare la prova di comprensione lettura (TCL) già spe-

<sup>118</sup> Cfr. nota 21.

rimentata nello studio principale di questa indagine. La decisione è stata presa dopo due livelli di analisi del problema.

E' stata scartata l'utilizzazione di prove di tipo enciclopedico, utilizzate nella maggior parte delle selezioni svolte dalla Sapienza, per evidenti problemi di validità del contenuto: osservando alcuni esempi di prove svolte in altri corsi di laurea si nota subito l'assoluta arbitrarietà degli argomenti considerati dalle domande rispetto alle caratteristiche del corso di laurea; bisogna anzi aggiungere che dopo i contatti stabiliti con diversi corsi di laurea, questi ultimi si sono spesso convinti a inserire domande di comprensione della lettura (il tipo di verifica scelto per costruire la prova) anche nelle loro prove di accesso.

Una volta deciso di legare l'accesso alla capacità di lettura il problema è stato quello di scegliere le prove, considerando la necessità di utilizzare prove in qualche modo già collaudate. La soluzione più efficace è apparsa quello di riproporre gli stessi 5 testi già utilizzati nello studio principale, svolte dagli studenti iscritti nel precedente Anno Accademico. In questo modo, oltre a contare su una prova di cui già si dispone di una prima validazione, sarà possibile effettuare dei confronti fra le popolazioni coinvolte.

Di fatto le condizioni di somministrazione risultano molto simili (nello studio principale il TCL è stato svolto per primo) anche se naturalmente cambiano, e di molto, gli aspetti motivazionali dei partecipanti: dalla tranquillità dell'orientamento alla tensione della selezione. Proprio per assicurare le possibilità di confronto è stato deciso di non modificare la prova, lasciando anche i tre item che avevano presentato maggiori criticità nell'item analisi. Naturalmente sono stati considerati con la dovuta attenzione gli aspetti di riservatezza degli strumenti, garantita dal fatto che nelle precedenti somministrazioni della prova tutti i materiali utilizzati erano stati restituiti, per cui si può ragionevolmente escludere che i candidati conoscessero in anticipo i testi oggetto di indagine.

La sola decisione di attuare la procedura per il numero programmato ha prodotto una prima selezione pratica: i tempi tecnici definiti dall'amministrazione universitaria prevedevano l'iscrizione obbligatoria alla prova entro i primi giorni di settembre 2004, solo 217 candidati hanno presentato domanda, ma successivamente diverse decine di studenti hanno contattato le strutture del corso manifestando il loro disorientamento per non aver saputo in tempo utile l'esistenza di una data di scadenza. Naturalmente per loro non c'è stata nessuna possibilità di partecipare alla prova di accesso, per cui l'informazione corretta e tempestiva

è risultata di fatto il primo criterio di selezione per l'iscrizione al corso di laurea.

La prova si è svolta il 17 settembre 2004, in condizioni ottimali, considerando che l'amministrazione ha fornito con congruo anticipo l'elenco dei candidati. Questo ha permesso di utilizzare quattro aule, con una disposizione attenta dei candidati nelle aule per assicurare una compilazione corretta e valida della prova. Hanno partecipato 203 candidati su 217 iscritti, compresi i due della sessione suppletiva che si è svolta lo stesso giorno nei locali della casa circondariale di Rebibbia.

Naturalmente la prima lettura dei risultati è risultata immediatamente funzionale agli scopi selettivi, per cui si è deciso di utilizzare, proprio alla luce dei risultati ottenuti nella prova simulata del precedente anno accademico, la metodologia IRT Om, che prevede la penalizzazione delle omissioni (non distinguendo fra omissioni e domande non raggiunte<sup>119</sup>), soprattutto per la necessità di differenziare i punteggi nella zona di "taglio" del numero programmato. In effetti, considerando i punteggi secondo l'item analisi classica, il taglio si avrebbe a 15 punti grezzi<sup>120</sup>, ma lo stesso punteggio grezzo è stato raggiunto dai candidati compresi fra il 141° e 157° posto. I punteggi IRT risultano invece perfettamente scalati e funzionali allo scopo. Da sottolineare che i candidati sono stati informati sulla procedura, insistendo in particolare sulla considerazione del valore delle risposte in relazione alla difficoltà e alla discriminatività della domanda.

La Tabella 7.1 riassume, nel consueto formato adoperato nel corso del lavoro di ricerca, le due item analisi svolte sui risultati della prova.

<sup>119</sup> Su questo punto sono state fornite le opportune spiegazioni ai candidati prima dello svolgimento della prova, esplicitando nelle istruzioni sia le modalità di formazione dei punteggi, sia l'informazione relativa alla penalizzazione delle omissioni

<sup>120</sup> Il precedente taglio simulato a 17 punti si svolgeva su un gruppo di 419 rispondenti, mentre il gruppo reale dell'A.A. 2004-05 era composto da soli 203 studenti.

Tabella 7.1- Analisi degli item – Prova di accesso A.A. 2004-05

Item	Chiave	Modello classico							Modello IRT (2 parametri)					
		A	B	C	D	Omis.	Fac.	PBs.	A	B	C	D	Omis.	PBt
1	C	30	4	43	19	4	0,43	0,33	-15	3	32	-21	-8	0,32
2	A	23	22	21	33	1	0,23	0,11	9	7	-26	10	-5	0,09
3	D	6	10	17	64	3	0,64	0,35	-11	-27	-10	34	-11	0,34
4	B	27	34	33	1	3	0,34	0,48	-31	46	-11	-7	-12	0,46
5	B	1	61	1	37	-	0,61	0,34	-15	31	-6	-27	-4	0,31
6	A	85	3	4	7	-	0,85	0,16	13	-23	-4	1	-	0,13
7	A	77	4	8	9	1	0,77	0,29	27	-12	-9	-19	-11	0,27
8	D	11	1	6	78	3	0,78	0,22	-10	-17	-9	23	-12	0,23
9	D	18	3	2	75	2	0,75	0,36	-21	-16	-15	32	-6	0,32
10	D	3	7	12	76	1	0,76	0,39	-17	-7	-28	34	-6	0,34
11	D	4	8	5	77	5	0,77	0,38	-21	-18	-14	37	-14	0,37
12	D	9	29	7	51	4	0,51	0,36	-16	-12	-16	35	-16	0,35
13	B	6	82	6	4	1	0,82	0,28	-10	26	-7	-23	-10	0,26
14	D	13	34	7	35	10	0,35	0,31	-8	-9	-11	33	-21	0,33
15	C	6	4	86	2	2	0,86	0,26	-12	-18	24	-8	-6	0,24
16	C	5	14	30	50	1	0,30	0,27	-14	-11	25	-10	0	0,25
17	D	18	54	6	21	1	0,21	0,37	-16	-12	-14	40	-7	0,40
18	B	5	62	20	9	3	0,62	0,37	-10	36	-19	-21	-8	0,36
19	A	47	35	8	9	1	0,47	0,38	37	-20	-15	-12	-13	0,37
20	B	7	46	24	18	5	0,46	0,32	-16	31	-12	-12	-7	0,31
21	D	6	6	5	81	1	0,81	0,39	-24	-11	-24	38	-9	0,38
22	D	-	19	6	68	7	0,68	0,30	0	-16	-17	36	-24	0,36
23	A	48	33	7	8	4	0,48	0,42	45	-17	-12	-25	-24	0,45
24	C	9	2	79	2	7	0,79	0,16	-5	-9	27	-10	-26	0,27
25	D	4	15	4	69	7	0,69	0,37	-14	-19	-12	44	-32	0,44
26	C	13	8	62	6	11	0,62	0,42	-23	-13	51	-12	-33	0,51
27	A	72	9	5	2	11	0,72	0,25	39	-12	-9	-8	-34	0,39
28	C	6	7	52	23	12	0,52	0,47	-20	-11	51	-20	-30	0,51
29	B	10	60	8	8	14	0,60	0,44	-27	53	-14	-5	-37	0,53
30	A	77	1	4	4	14	0,77	0,21	39	-11	-10	-7	-35	0,39

Nella parte sinistra della tabella è riportata l'analisi degli item secondo il modello classico. Nella parte destra è riportata l'analisi effettuata con il metodo IRT. Le prime colonne (A-D e Omis.) riportano le correlazioni esistenti fra ogni alternativa e il tratto latente della prova ( $\theta$ ), che dovrebbe essere positiva e il più alta possibile per l'alternativa corretta, negativa per i distrattori e le omissioni. L'ultima colonna (PBt) riporta la misura equivalente del punto biseriale per il metodo IRT, cioè la correlazione

(Pearson) fra le risposte esatte/non esatte e il tratto latente della prova (e non il punteggio totale come per il punto biseriale). Sono stati evidenziati in grigio le misure critiche.

Una prima lettura dei risultati relativi agli item della prova conferma sostanzialmente la precedente somministrazione con due sole eccezioni: l'item 1 ora risulta pienamente accettabile, mentre peggiorano i dati relativi (in particolare per il metodo IRT) sull'item 2.

Nel complesso si riducono in maniera sensibile le omissioni, soprattutto nella parte finale della prova, e questo può ragionevolmente essere attribuito alla maggiore motivazione dei candidati. Interessante notare che il punteggio medio rispetto allo studio principale è salito da 15,6 a 18,2, con una deviazione standard che si riduce (da 4,8 a 4,7), quindi miglior rendimento e maggiore omogeneità dei candidati per il primo anno di applicazione selettiva della prova. Di conseguenza è aumentata la facilità media della prova, da 0,52 a 0,61, con diversi item che si avvicinano alla soglia di 0,80. Migliora anche la consistenza interna della prova, con un indice KR20 di 0,77.

Il controllo delle domande non raggiunte conferma quanto sostenuto a proposito dei risultati della prova simulata (cfr. 5.3) circa le diverse motivazioni che guidano i rispondenti nel corso di una prova di accesso rispetto a una prova di semplice orientamento. In effetti le omissioni relative alle ultime domande crollano nella nuova somministrazione, in particolare per le omissioni superiori alle ultime tre domande.

Tabella 7.2- Riepilogo risposte prova TCL – Rispondenti che non completano la prova (NR=Non Raggiunte)

	Stud.	%NR ultima domanda	%NR1-3	%NR4-6	%NR7-9	%NR>9
2003-04	419	29,8	6,9	11,5	7,4	4,1
2004-05	203	13,3	6,4	2,5	3,4	1,0

Anche il confronto puntuale fra risposte esatte, errate e omissioni fra i due gruppi attesta il diverso profilo dei rispondenti, con un netto calo complessivo delle omissioni e una riduzione anche delle risposte errate.



Tabella 7.3 – Riepilogo risposte prova TCL – Incidenza delle omissioni

	Stud.	%Esatte	%Errate	%Omissioni
2003-04	419	52,0	36,2	11,8
2004-05	170	60,7	34,6	4,8

Naturalmente, superata la fase empirica di utilizzazione dei risultati per gli scopi selettivi, sarà possibile approfondire il confronto fra i diversi modelli di analisi anche per questo nuovo set di dati, e rapportare le due diverse somministrazioni, cercando per quanto possibile di definire anche le possibili distorsioni introdotte dai diversi scopi per cui è stata utilizzata la stessa prova.

Il confronto fra le due popolazioni misurate in ingresso con lo stesso strumento permetterà di approfondire anche l'analisi del rapporto con gli esiti accademici, attraverso la verifica periodica, per anno di corso, dei crediti conseguiti, anche se in questo caso non sarà naturalmente possibile un confronto interno alla popolazione, mancando il sottogruppo dei "non ammessi".

Concludiamo con un segnale di speranza, ch  il tema della selezione resti il pi  possibile confinato nelle stanze della ricerca, segnale tratto ancora una volta dall'esperienza didattica: nessuno studente che ha partecipato alla selezione nell'Anno Accademico 2004-05   risultato escluso per effetto di questa prova di accesso, in quanto il meccanismo incrociato delle rinunce degli ammessi (da parte di coloro che hanno scelto di non iscriversi, in diversi casi perch  ammessi anche in altri corsi di laurea) e dei subentri degli esclusi ha permesso di esaurire tutta la graduatoria determinata dalla prova di accesso.



## Bibliografia

### Impostazione della ricerca e accesso all'università

- ANDREANI DENTICI O., AMORETTI E. (2000), *Gli accessi all'università: selezione e orientamento*, Franco Angeli, Roma
- AUSUBEL D.P. (1968), *Educational Psychology*, Holt Rinehart e Winston, New York (tr.it. *Educazione e processi cognitive. Guida psicologica degli insegnanti*, Franco Angeli, Milano, 1985).
- BAILEY K.D. (1982), *Methods of Social Research*, The Free Press, New York (tr.it. *Metodi della ricerca sociale*, il Mulino, Bologna, 1995).
- BECCHI E., VERTECCHI B. (a cura di) (1984), *Manuale critico della sperimentazione e della ricerca educativa*, Franco Angeli, Milano.
- BENVENUTO G. (a cura di) (2003), *L'università possibile*, Anicia, Roma
- BINI G., BORSESE A. (1995), *Alcune considerazioni sui test per l'ammissione all'Università in Italia*, in «Scuola & Città», a. 45, n. 9, 30 set 1995, p. 384-389
- BONCORI G., BONCORI L. (2002), *L'orientamento*, Carocci, Roma.
- BONCORI L., GIANGIACOMO P., LAGHI F., (2001) *Un intervento di orientamento universitario: la previsione del successo accademico e della soddisfazione personale*, in «Università e scuola», 1, pp. 25-38.
- BOLLETTA R. (2000), *L'assegnazione dei punteggi. Una riflessione che prosegue*, in *ONES, Proposte per le terze prove 2000*, Franco Angeli, Roma.
- CATALANO G., FIGÀ TALAMANCA A. (a cura di) (2002), *EURO STUDENT. Lecondizioni di vita e di studio degli studenti universitari italiani*, Il Mulino, Bologna.
- CAVALLI A., FACCHINI C. (a cura di) (2001), *Scelte cruciali, indagine IARD su giovani e famiglie di fronte alle scelte alla fine della scuola secondaria*, Il Mulino, Bologna.
- CIVES G., CORDA COSTA M. FATTORI M., SICILIANI DE CUMIS N. (a cura di) (2002), *Evaluation – Studi in onore di Aldo Visalberghi*, Salvatore Sciascia Editore, Caltanissetta - Roma.
- CNVSU (2004), *Quinto Rapporto sullo Stato del Sistema Universitario*.

- FAVRETTO G. (a cura di) (2002), *Prove d'accesso universitario e capacità predittive: cinque anni di ricerche del Centro docimologico, DiPAV*, in «Quaderni Quadrimestrale di psicologia ed antropologia culturale», fascicolo 4, sezione monografica.
- GARDNER H. (1983), *Frames of Mind: The Theory of Multiple Intelligence*, Basic Books, New York (tr.it. *La nuova scienza della mente*, Feltrinelli, Milano, 1988).
- GRAVES D. (2002), *Testing Is Not Teaching*, Portsmouth, Heinemann.
- GUICCIARDI M., LOSTIA M. (1998), *Utilità o inutilità delle prove di ammissione in psicologia: otto anni di ricerche nel corso di laurea di Cagliari*, in «Giornale italiano di psicologia», 4, 823-846
- MCDONNELL L.M. (2004), *Politics, persuasion, and educational testing*, Cambridge, MA: Harvard University Press.
- LUCISANO P., SALERNI A. (2002), *Metodologia della ricerca in educazione e formazione*, Carocci, Roma
- OECD (2005), *Education at a Glance, OECD Indicators 2005*, OECD, Paris.
- ONES (2000), *Selezione delle terze prove. Materiali utilizzati nella sessione 1999*, Franco Angeli, Roma.
- SANTAMBROGIO M. (1999), *Chi ha paura del numero chiuso?*, Laterza, Bari.
- SCALERA V. (2002), *Competenze cross-curricolari*, in «Iter», n. 16-17.
- VISALBERGHI A. (1958), *Esperienza e valutazione*, Taylor, Torino.
- VISALBERGHI A. (1978), *Pedagogia e scienze dell'educazione*, Mondadori, Milano.
- WALBERG H.J., HAERTEL G.D. (1994), *Le implicazioni della psicologia cognitiva per la misurazione dei risultati degli studenti*, in OECD, *Valutare l'insegnamento*, Roma, Armando (tit.or. *Making Education Count*, 1994).

### **Costruzione di prove oggettive**

- ANDREANI DENTICI O. (1968), *Abilità mentale e rendimento scolastico*, La Nuova Italia, Firenze.
- BENVENUTO G., LASTRUCCI E., SALERNI A. (1995), *Leggere per capire*, Anicia, Roma.
- BENVENUTO G., SALERNI A., (1998), *Criteri per la somministrazione collettiva di prove strutturate*, in «Cadmò», 21, pp. 68-76.
- BONCORI L. (1993), *Teoria e tecnica dei test*, Bollati Boringhieri, Torino.
- CAMPBELL J., KELLY D., MULLIS I., MARTIN M., SAINSBURY M. (2001), *Framework and Specifications for PIRLS Assessment 2001*, International Study Center, Lynch School of Education, Boston College, Chestnut Hill, MA USA.

- CORDA COSTA M., VISALBERGHI A. (a cura di) (1995), *Misurare e valutare le competenze linguistiche. Guida scientifico pratica per gli insegnanti*, La Nuova Italia, Firenze.
- DE MAURO T. (2000), *Dizionario italiano*, Paravia Bruno Mondadori Editore.
- DOMINICI G. (1993), *Manuale della valutazione scolastica*, Laterza, Bari.
- GALLINA V. (a cura di) (2000) *La competenza alfabetica in Italia*, Franco Angeli, Roma
- GATTULLO M. (1968), *Didattica e Docimologia. Misurazione e valutazione nella scuola*, Armando, Roma.
- GATTULLO M., GIOVANNINI M.L. (1989), *Misurare e valutare l'apprendimento nella scuola media*, Bruno Mondadori, Milano.
- HALADYNA T.M. (2004), *Developing and validating multiple-choice test items*, Mahwah, N.J., Lawrence Erlbaum Associates.
- INValSI (2001), *Ricerche valutative internazionali 2000*, Franco Angeli, Roma
- LUCISANO P. (a cura di), (1989), *Lettura e comprensione* Loescher, Torino.
- LUCISANO P. (a cura di), (1994), *Alfabetizzazione e lettura in Italia e nel mondo. I risultati dell'indagine internazionale IEA-SAL*, Tecnodid, Napoli.
- LUCISANO P., SINISCALCO M.T. (1994) *Rassegna bibliografica delle ricerche IEA*, in «Cadmò», II, numero 5-6, pp. 164-186.
- OECD (1999), *Measuring Student Knowledge and Skills*, OECD, Paris,
- OECD (2001), *Knowledge and Skills for Life. First Results from PISA 2000*, OECD, Paris
- OECD (2003), *The PISA 2003 Assessment Framework*, OECD, Paris.
- OECD (2004), *Learning for Tomorrow World. First Results from PISA 2003*, OECD, Paris.
- OSTERLIND S.J. (2001), *Constructing Test Items*, Kluwer Academic Publishers, Boston.
- VERTECCHI B. (1984), *Manuale della valutazione*, Roma, Editori Riuniti.
- ZAMMUNER V. L. (1998), *Tecniche dell'intervista e del questionario*, Il Mulino, Bologna.

### **Item analisi e trattamento dei dati**

- ADAMS R.J., WILSON M., WU M. (1997), *Multilevel Item Response Models: an approach to errors in variables regression*, in «Journal Educational and Behavioral Statistics», Vol.22, No.1, 47-76.

- ANKEMAN R., WITT E., DUNBAR S. (1999), *An Investigation of the Power of the Likelihood Ratio Goodness-of-Fit Statistic in Detecting Differential Item Functioning*, in «Journal of Educational Measurement», Vol.36, No.4, 277-300.
- ASQUINI G. (1997), *Il foglio di calcolo nell'analisi delle prove a risposta chiusa*, in «Cadmò», V, 13-14.
- ASQUINI G., PIRIA L. (1998), *Elementi essenziali di item analisi e statistiche descrittive*, in PAGNONCELLI L. (a cura di) *Formazione e valutazione dell'apprendimento*, Roma, Anicia.
- ATTALI Y., FRAENKEL T. (2000), *The Point-Biserial as a Discrimination Index for Distractors in Multiple Choice Items: Deficiencies in Usage and an Alternative*, in «Journal of Educational Measurement», Vol.37, No.1, 77-86.
- BAKER F.B. (1987), *Methodology review: item parameter estimation under the one-, two-, and three-parameter logistic models*, in «Applied Psychological Measurement», 11, 111-141.
- BAKER, F.B. (2001), *The Basics of Item Response Theory*, (Second edition), New York, ERIC Clearinghouse on Assessment and Evaluation.
- BARTHOLOMEW D.J. (1987) *Latent variable models and factor analysis*, Griffin & C. Limited, London.
- BARTHOLOMEW, D. J. (1995), *Spearman and the origin and development of test theory*, in «British Journal of Mathematical and Statistical Psychology», 48, 211-220.
- BIRNBAUM A. (1968). *Some latent trait models and their use in inferring an examinee's ability*, in Lord F.M., Novick M.R. (eds.), *Statistical theories of mental test scores*, Addison-Wesley, Reading.
- BOCK R.D., AITKIN M. (1981) *Marginal maximum likelihood estimation of item parameters: application of an EM algorithm*, in «Psychometrika», 46, 443-459.
- BOCK R.D., GIBBONS R., MURAKI E. (1988), *Full-information Item Factor Analysis*, in «Applied Psychological Measurement», 12, 261-280.
- BOCK R.D., LIEBERMAN M. (1970), *Fitting a response model to n dichotomously scored items*, in «Psychometrika», 35, 179-197.
- BOLASCO S. (1999), *Analisi multidimensionale dei dati*, Carocci, Roma.
- BOLT D., COHEN A., WOLLACK J. (2001), *A Mixture Item Response Model for Multiple-Choice Data*, in «Journal Educational and Behavioral Statistics», Vol.26, No.4, 381-409.
- BOLT D., COHEN A., WOLLACK J. (2002), *Item Parameter Estimation Under Conditions of Test Speededness: Application of a Mixture Rasch Model With*

- Ordinal Constraints*, in «Journal of Educational Measurement», Vol.39, No.4, 331-348.
- BOOMSMA A., VAN DUIJN, M.A.J., SNIJDERS, T.A.B.(Eds.) (2000), *Essays on Item Response Theory*, New York, Springer-Verlag
- BOVE G. (2000), *Metodologie per la misura dei livelli di competenza nelle indagini educative su larga scala*, in GALLINA V. (a cura di) *La competenza alfabetica in Italia*, Franco Angeli, Roma.
- BURTON R. (2001), *Do Item-discrimination Indices Really Help Us to Improve Our Tests?*, in «Assessment & Evaluation in Higher Education», Vol.26, No.3, 213-220.
- CALONGHI L. (1978), *Statistiche d'informazione e di valutazione*, Roma, Bulzoni
- CHRISTOFFERSON A. (1975) *Factor analysis of dichotomized variables*, *Psychometrika*, 40, 5-32.
- CROCKER, L., ALGINA, J. (1986), *Introduction to classical & modern test theory*. New York, Harcourt Brace Jovanovich
- CRONBACH, L. J., WARRINGTON, W. G. (1952), *Efficiency of multiple-choice tests as a function of spread of item difficulties*, in «Psychometrika», 17, 127-147.
- DE AYALA R., PLAKE B., IMPARA J. (2001), *The Impact of Omitted Responses on the Accuracy of Ability Estimation in Item Response Theory*, in «Journal of Educational Measurement», Vol.38, No.3, 213-234.
- DE GRUIJTER D.N.M., VAN DER KAMP, L.J.T. (2003), *Statistical Test Theory for Education and Psychology*, Manuscript, © D. N. M. de Gruijter & L. J. Th. Van der Kamp.
- EMBRETSON, S. E., REISE, S. P. (2000). *Item response theory for psychologists*, Mahwah, NJ: Lawrence Erlbaum Associates.
- FABBRIS L. (1997), *Statistica multivariata: analisi esplorativa dei dati*, McGraw Hill, Milano.
- GAL P.M., GARFIELD, R.W.(1997), *The Assessment Challenge in Statics Education*, IOS Press
- GOLDSTEIN L.A., LEWIS P.J. (Eds.), (1996), *Assessment: Problems, Developments and Statistical Issue*, John Wiley & Sons
- GUILFORD J. P. (1936 - 1954), *Psychometric Methods*, New York: McGraw-Hill.
- GUILFORD J. P. (1950), *Fundamental Statistics in Psychology and Education*, New York: McGraw- Hill.
- GUTTMAN L. (1950), *The basis for scalogram analysis*, in STOUFFER ET AL. (Eds.), *Measurement and prediction*, New York: Wiley.

- HAMBLETON R.K. SWAMINATHAN H., ROGERS H.J. (1991), *Fundamentals of Item Response Theory*, Sage, New York.
- HOLLAND, P.W., HOSKENS, M. (2002), *Classical Test Theory as a First-Order Item Response Theory: Application to True-Score Prediction From a Possibly Nonparallel Test*, research Report (RR-02-20), Princeton, Educational Testing Service.
- JANSSEN R., TUERLINCKX F., MEULDERS M., DE BOECK P. (2000), *A Hierarchical IRT Model for Criterion-Referenced Measurement*, in «Journal Educational and Behavioral Statistics», Vol.25, No.3, 285-306.
- KAMATA A. (2001), *Item Analysis by the Hierarchical Generalized Linear Model*, in «Journal of Educational Measurement», Vol.38, No.1, 79-93.
- KUDER, G. F., RICHARDSON, M. W. (1937), *The theory of estimation of test reliability*, in «Psychometrika», 2, 151-160.
- LAWLEY, D.N. (1943) *On problems connected with item selection and test construction*, in «Proceedings of the Royal Statistical Society of Edinburgh», 61, 273-87.
- LITTLE R., RUBIN D.B. (1987), *Statistical analysis with missing data*, Wiley, New York.
- LORD, F.M. (1952), *The relation of the reliability of multiple-choice tests to the distribution of item difficulties*, in «Psychometrika», 17, 181-194.
- LORD F.M., NOVICK M.R. (eds.) (1968), *Statistical theories of mental test scores*, Addison-Wesley, Reading.
- LORD F.M. (1980) *Application of item response theory to practical testing problems*, Hillsdale, NJ. Erlbaum
- LUCISANO P, SINISCALCO M.T. (1992), *Finalizzazione e taratura di una prova di alfabetizzazione lettura: item analisi tradizionale e item analisi di Rasch*, in «Scuola & città» 1, 31 gennaio 1992, pp.31-40.
- MISLEVY R.J. (1994) *Test Theory Reconceived - CSE Technical Report 376*, Educational Testing Service/CRESST, University of California, Los Angeles
- MISLEVY R.J., WU P. (1988), *Inferring examinee ability when some item responses are missing* (RR 88-48-ONR), Educational Testing Service, Princeton, NJ.
- MISLEVY R.J., WU P. (1996), *Missing responses and IRT ability estimation: Omits, choice, time limits and adaptive testing*, (RR 96-30-ONR), Educational Testing Service, Princeton, NJ.
- MARTIN, M.O., MULLIS, I.V.S., KENNEDY, A.M. (Eds.) (2003), *PIRLS 2001 Technical Report*, Chestnut Hill, MA: Boston College.



- MULLIS, I.V.S., MARTIN, M.O., GONZALEZ, E.J., KENNEDY, A.M. (2003), *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools* Chestnut Hill, MA: Boston College.
- NOVICK, M. R. (1966), *The axioms and principal results of classical test theory*, in «*Journal of Mathematical Psychology*», 3, 1-18.
- NUNNALLY, J.C. (1978), *Psychometrics Theory, Second Edition*, New York, McGraw Hill.
- OECD (2002), *Pisa 2000 Technical Report*, OECD, Paris.
- OECD (2005), *Pisa 2003 Technical Report*, OECD, Paris.
- RASCH, G. (1960) *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research
- RASCH G. (1977), *On Specific Objectivity on Attempt al Tormalizing the Request for Generality and Validity of Scientific Statements*, in «*Danish Yearbooks of Philosophy*», 14, 58-94.
- SPEARMAN, C. E. (1904a), 'General intelligence' objectively determined and measured, in «*American Journal of Psychology*», 5, 201-293.
- SPEARMAN, C. E. (1904b), *Proof and measurement of association between two things*, in «*American Journal of Psychology*», 15, 72-101.
- STOCKING M.L., EIGNOR D., COOK L. (1988), *Factors affecting the sample invariant properties of linear and curvilinear observed and true score equating procedures (RR-88-41)*, Educational Testing Service, Princeton, NJ
- STONE C. (2000), *Monte Carlo Based Null Distribution for an Alternative Goodness-of-Fit Test Statistic in IRT Models*, in «*Journal of Educational Measurement*», Vol.37, No.1, 58-75.
- TAKANE Y., DE LEEUW J. (1987) *On the relationship between item response theory and factor analysis of discretized variables*, «*Psychometrika*», 52, 393-408.
- THISSEN, D. (1991), *MULTILOGtm User's Guide. Multiple Categorical Item Analysis and Test Scoring Using Item Response Theory, Version 6.0*. Chicago, Illinois, Scientific Software.
- THORNDIKE R. L., HAGEN E, P. (1955 – 1977), *Meaurement and Evaluation in Psychology and Education (Fourth Edition)*, John Wiley & Sons New York
- TUCKER L R (1946) *Maximum validity of a test with equivalent items*, in «*Psychometrika*» 11(1) 1-13
- VAN DER LINDEN W.J., HAMBLETON R.K. (Eds.), (1997), *Handbook of modern item response theory*, New York, Springer-Verlag.
- VAN DER SCHOOT, F.C.J.A. (2002), *The application of an IRT-based method for standard setting in a three-stage procedure*, Paper presented at the an-

- nual meeting of the National Council on Measurement in Education, April 2, 2002, New Orleans.
- VERSTRALEN H, BECHGER T, MARIS G. (2001), *The Combined Use of Classical Test Theory and Item Response Theory*, Cito, Arnhem.
- VINCENTI A.B., CALVANI A. (1987), *Item analisi e modello di Rasch*, in «Orientamenti pedagogici», vol.34, n.2, 223-248.
- WAINER H., THISSEN D.J. (1994), *On examinee choice in educational testing*, in «Review of Educational Research», 64, 159-195.
- WOOD R. (1985), *Item Analysis*, in Husen T., Postlethwaite T.N., *The International Encyclopedia of Education*, New York, Pergamon Press.
- WRIGHT B.D. (1985), *Rasch Measurement Models*, in HUSEN T., POSTLETHWAITE T.N., *The International Encyclopedia of Education*, New York, Pergamon Press.
- WRIGHT B.D. E MASTERS G.N. (1982). *Rating Scale Analysis*, MESA Press, Chicago.
- ZANI S. (2000), *Analisi dei dati statistici*, Giuffrè, Milano.

### Test Cloze

- ABRAHAM R. (1992), *The Meaning of Cloze Test Scores: An Item Difficulty Perspective*, in «The Modern Language Journal», Vol.76, No.4, 468-478.
- BORMUTH, J. R. (1968), *Cloze Test Readability: Criterion Reference Scores*, in «Journal of Educational Measurement», 5, no.3, 189-196
- BROWN, J. D. (1980), *Relative merits of four methods for scoring cloze tests*, in «Modern Language Journal», 64, 311-317.
- BROWN, J. D. (1983), *A closer look at cloze: Validity and reliability*, in J. W. Oller, Jr. (Ed.) *Issues in Language Testing Research* (pp. 237-250). Rowley, MA: Newbury House.
- BROWN, J. D. (2002), *Do Cloze Tests Work? Or, Is It Just an Illusion?*, in «Second Language Studies», 21(1), Fall 2002, pp. 79-125.
- BUTLER J. (1991), *Cloze Procedures and Concordances: The Advantages of Discourse Level Authenticity in Testing Expectancy Grammar*, in «System», Vol.19, No.1/2, 29-38.
- CHAPPELLE C. (1990), *Cloze method: What Difference Does It Make?*, in «Language Testing», Vol.7, No.2, 121-146.
- CHIARI S. (2002), *La procedura cloze, la ridondanza e la valutazione della competenza della lingua italiana*, «Italice», 4.
- FOTOS S. (1991), *The Cloze Test as an Integrative Measure of EFL Proficiency: A Substitute for Essays on College Entrance Examinations?*, in «Language Learning», Vol.41, No.3, 313-336.

- KOBAYASHI M. (2002), *Cloze Tests Revisited: Exploring Item Characteristics with Special attention to scoring Methods*, in «The Modern Language Journal», Vol.86, No.4, 571-586.
- MARELLO C. (1984), *Fare buchi nei testi per poi riempirli*, in «Lend», 2, pp.3-16 e 3, pp.3-10.
- MARTINEZ P.S., (1978), *The Effect of Teaching Sixth Graders to Use Context Cues to Complete Cloze Deletions in Social Studies Materials*, Dissertation Abstracts International, 39, 3503° (University microfilms n.78-24, 144).
- NUCCORINI S. (2001), *Il cloze test in inglese*, Carocci, Roma.
- PATEL C., DAY R. (1996), *The Influence of Cognitive Style on the Understandability of a Professional Accounting Pronouncement by Accounting Students*, in «The British Accounting Review», Vol.28, No.2, 139-154.
- ROHLFS G. (1969), *Grammatica storica della lingua italiana e dei suoi dialetti. Sintassi e formazione delle parole*, Einaudi, Torino.
- SASAKI M. (2000), *Effects of Cultural Schemata on Students' Test-Taking Processes for Cloze Tests: a Multiple Data Source Approach*, in «Language Testing», Vol.17, No.1, 85-114.
- SPOLSKY B. (2000), *Language Testing* in «The Modern Language Journal», Vol.84, No.4, 536-552.
- TAYLOR, W.L. (1953), *Cloze procedure: A new tool for measuring readability*, in «Journalism Quarterly», 30, 415-433.



Finito di stampare nel mese di aprile 2006  
dal Centro Stampa Nuova Cultura, Roma