

# La valutazione delle competenze matematiche: alcuni risultati di un progetto europeo sulle matricole universitarie

**Stefania Mignani, Paola Monari, Roberto Ricci**

*Dipartimento di Scienze Statistiche "P. Fortunati"*

*Università di Bologna*

**Riassunto.** La valutazione delle competenze presenta molteplici aspetti che possono essere trattati con la metodologia Item Response Theory (IRT) sviluppata per trattare problemi di misurazione di abilità. In questo lavoro, che si inserisce in un progetto europeo, tale metodologia è impiegata per valutare, sulla base di un questionario con domande a risposta multipla, la competenza in matematica delle matricole universitarie. I modelli considerati permettono di giudicare la coerenza delle domande proposte e di stimare per ciascun studente il livello di conoscenza in questa disciplina.

**Parole chiave:** Item Response Theory, Variabili latenti, Proprietà psicometriche di un item, Misura di una competenza

## 1. Introduzione

La valutazione delle competenze acquisite da un soggetto è un processo che può essere essenzialmente scandito in tre momenti distinti, strettamente legati tra di loro, che trovano nel metodo statistico una loro base scientifica:

- la definizione dell'oggetto di misurazione,
- la predisposizione di un adeguato strumento di misura
- l'analisi dei risultati ottenuti.

Il presente lavoro riprende i risultati di un progetto europeo – *Mathematics in Europe* - volto a promuovere la diffusione della cultura matematica in Europa, finanziato dalla Comunità europea nell'ambito del progetto Socrates, e coordinato dal Dipartimento di Matematica dell'Università di Bologna (Manaresi, 2005)

Tra i tanti obiettivi previsti nel progetto, ampio spazio è stato dato al monitoraggio e alla valutazione della preparazione matematica. Questo interesse ha trovato negli anni più recenti una crescente motivazione a causa della crisi di vocazioni scientifiche nei giovani che escono dalle scuole medie superiori e del progressivo af-

fievolimento della preparazione matematica, scientifica e tecnologica degli studenti che entrano nell'università italiana.

Il mondo occidentale è consapevole che la rinuncia al sapere scientifico mette in crisi lo sviluppo della società e che questa rinuncia trova le sue radici prima di tutto nei percorsi formativi della scuola in tutti i livelli che precedono gli studi universitari.

Per dare una risposta fondata alle preoccupazioni emergenti mondo della scienza, il progetto *Mathematics in Europe* si è proposto di misurare la preparazione matematica di un ampio e rappresentativo campione di studenti iscritti al primo anno delle università che hanno aderito al progetto, attraverso l'analisi delle risposte a un questionario a scelta multipla su argomenti matematici di base previsti nei programmi delle scuole medie superiori.

In questa ricerca, l'oggetto della misura voleva essere la preparazione matematica intesa come insieme di competenze in grado di risolvere problemi con livelli di difficoltà specificati.

A fronte di un questionario a risposta multipla, esistono adeguati metodi statistico-psicometrici che permettono di analizzare in modo rigorosamente fondato i risultati del questionario e offrono elementi oggettivi per una valutazione delle competenze coerenti con le specificità del processo di insegnamento-apprendimento.

Il lavoro ripropone alcune metodologie statistiche che si sono sviluppate nella seconda metà del secolo scorso col preciso intento di valutare simultaneamente due aspetti logicamente distinti in ambito psicologico-educazionale:

- le competenze dello studente misurate sulle risposte al questionario;
- le caratteristiche delle domande proposte nel questionario.

## **2. La valutazione delle competenze: i modelli di Item Response Theory**

Il processo di valutazione per accertare il livello di preparazione di un individuo prevede l'analisi dei risultati ottenuti somministrando un questionario contenente domande (*item*) opportunamente definite rispetto alle competenze da accertare.

Agli inizi degli anni '60, in ambito psicometrico-educativo, si è sviluppata una nuova impostazione metodologica, nota come *Item Response Theory* (Lord e Novick, 1968), che tratta, in un'unica soluzione, le caratteristiche delle domande e la competenza (in questo contesto chiamata abilità) del soggetto, intesa come dimensione latente sottostante al processo di apprendimento. Nell'ambito di questa teoria vengono definiti opportuni modelli che misurano la probabilità di rispondere esattamente ad una domanda in funzione delle competenze dello studente e delle caratteristiche della domanda stessa, traducendole in un'unica scala.

Nell'IRT il livello di abilità di una persona viene messo in relazione con le risposte ad un test attraverso un modello matematico che permette di individuare sia il livello della variabile latente che esprime detta abilità, sia le proprietà delle domande. Più precisamente la funzione matematica descrive la relazione tra la probabilità di fornire una certa risposta ad un item ed il livello di abilità individuale. Ciascun modello è caratterizzato da uno o più parametri i cui valori numerici apportano informazioni circa le proprietà psicometriche delle domande stesse, che permettono di mettere in evidenza alcune peculiarità dei quesiti, come la loro difficoltà, la loro capacità di differenziare rispondenti con preparazione diversa, ecc.

L'apparato metodologico-concettuale su cui si sviluppa l'IRT presenta molti vantaggi rispetto alla soluzione più comunemente adottata nella teoria classica dei test (Hambleton e Swaminthan, 1985). In particolare, una volta che il questionario è stato validato a seguito di una accurata fase di calibrazione degli item, ovvero dopo che sono stati testati su un numero adeguato di soggetti verificandone la coerenza col costrutto da misurare, le caratteristiche di un item non dipendono dal campione di soggetti in base al quale sono state determinate. Ciò permette di costruire gruppi di domande equivalenti in termini di capacità di misurazione da cui attingere per la formulazione di questionari standardizzati in funzione della competenza che si vuole misurare. Inoltre, adottando un modello di IRT, si può valutare l'equivalenza, in termini di performance, di gruppi di rispondenti con caratteristiche diverse.

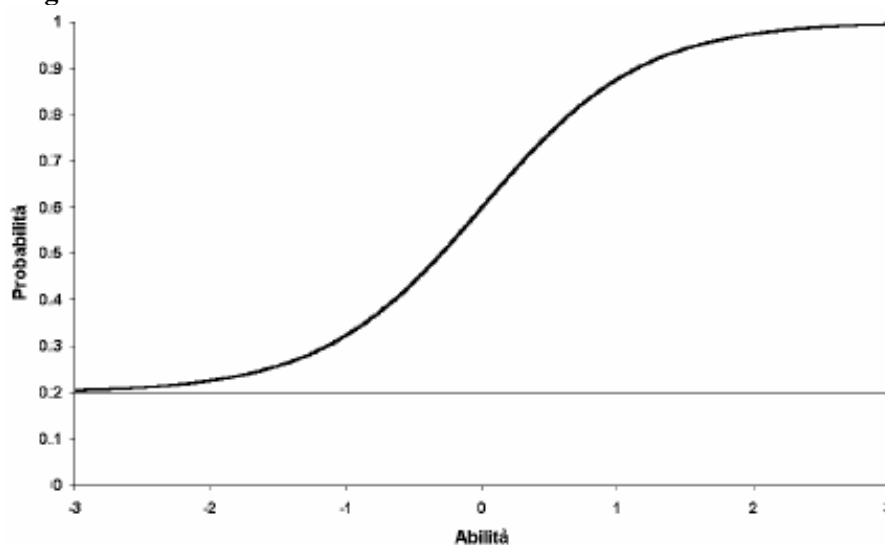
La scelta di un particolare modello dipende da diverse considerazioni sulla natura dei dati e sulle finalità conoscitive per cui si è predisposto il test. Un aspetto rilevante riguarda infatti la tipologia delle domande, le cui modalità di risposta sono strettamente legate alle caratteristiche del processo di valutazione. Un altro elemento fondamentale nella selezione del modello più adeguato è strettamente legato alla fase di valutazione che si sta affrontando, occorre infatti distinguere il momento della calibrazione del questionario dal momento più tecnico della misurazione.

### 3. I modelli per dati dicotomici

Il questionario può essere composto da item che prevedono due sole possibili risposte (ad esempio vero o falso, sì o no, ...); in tal caso i dati che vengono considerati nel modello si definiscono dicotomici. La probabilità  $P_i(\theta)$  di dare la risposta definita come corretta ad una particolare domanda  $i$  è espressa in funzione dell'abilità, indicata con  $\theta$ . Per ognuna delle domande si può quindi costruire una funzione non decrescente nota come Curva Caratteristica dell'item (CCI).

A seconda della forma funzionale e dei parametri della curva si hanno modelli che si caratterizzano per le diverse proprietà psicometriche considerate. La solu-

**Figura 1.** *Curva caratteristica di un item*



zione più frequentemente adottata considera come forma funzionale quella esponenziale. In tale modo, tracciando un grafico nel quale si posizionano sull'asse delle ascisse i valori di  $\theta$  e sull'asse delle ordinate i valori della probabilità  $P_i(\theta)$ , ovviamente compresi tra 0 ed 1, la curva assume una forma ad S, in cui a bassi livelli di abilità corrispondono basse probabilità di rispondere correttamente alla domanda, mentre ad alti valori di  $\theta$  corrispondono probabilità elevate (vedi figura 1).

Il numero diverso di parametri che possono essere inclusi nel modello esponenziale definiscono altrettanti modelli con caratteristiche differenti. Il modello più generale considera tre parametri e assume quindi la seguente forma funzionale:

$$P_i(\theta) = c_i + (1 - c_i) \frac{\exp[a_i(\theta - b_i)]}{1 + \exp[a_i(\theta - b_i)]} \quad (1)$$

dove  $b_i$  è il parametro di difficoltà dell'item  $i$ , ossia il valore dell'abilità che nel soggetto corrisponde ad una probabilità uguale a 0,5 di fornire una risposta corretta e  $a_i$  rappresenta il parametro di discriminazione. Dal punto di vista matematico questo parametro  $a_i$  risulta proporzionale alla pendenza della curva nel punto  $\theta = b_i$  e quindi è definito su tutto  $\mathbb{R}$ ; tuttavia sotto il profilo fenomenico è ristretto all'intervallo  $[0; 2,0]$ , in quanto un valore negativo rappresenterebbe situazioni in cui soggetti con abilità alte hanno una probabilità di scegliere l'opzione corretta minore rispetto a soggetti con abilità più basse.

Il terzo parametro  $c_i$  (parametro di *guessing*) rappresenta l'asintoto orizzontale inferiore della curva, ovvero la probabilità di un soggetto con abilità molto bassa di dare la risposta esatta alla domanda  $i$ . Infatti è possibile immaginare che, indipendentemente dal livello di preparazione, un soggetto abbia sempre una probabilità non

nulla di scegliere a caso la risposta corretta. Tale probabilità è ovviamente inversamente proporzionale al numero delle opzioni di risposta proposte per ciascun item. Dal punto di vista applicativo  $c_i$  viene interpretato come probabilità di scegliere a caso la risposta corretta, e pertanto assume un ruolo più rilevante quando si dispone di diverse risposte tra cui scegliere, situazione descritta nel paragrafo successivo.

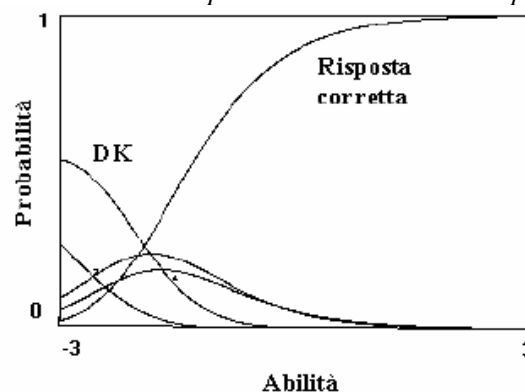
Se il parametro di *guessing* è posto uguale a zero, e si intende che ogni domanda abbia lo stesso potere discriminante del modello (1), si ritrova il più noto modello introdotto da Rasch nel 1960 (Rasch, 1960) che rappresenta la soluzione più spesso adottata per la sua semplicità computazionale e interpretativa.

#### 4. Il modello *multiple-choice* per risposte politomiche

Il modello *multiple-choice* è stato sviluppato per domande a risposta multipla caratterizzate da un insieme di opzioni di cui una corretta e le restanti errate. Queste ultime vengono comunemente denominate “distrattori” poiché costituiscono una sorta di effetto di disturbo per i rispondenti, specie per coloro che non conoscono la risposta corretta.

Il modello *multiple-choice* è descritto da funzioni di risposta che consentono di effettuare un’analisi grafica delle performance della domanda considerata. Una modellizzazione adeguata deve tener conto che i distrattori di un item, ovvero le opzioni non corrette, hanno una probabilità di essere selezionati che può essere crescente o decrescente per abilità medio-basse, ma certamente decrescente per abilità alte. In altri termini, la formulazione delle risposte errate deve essere tale che esse siano via via meno in grado di attirare le preferenze dei rispondenti più preparati. Il modello *multiple-choice* è caratterizzato da una funzione parametrica che permette di tracciare un grafico come quello della figura 2.

Figura 2. Funzioni di risposta di un modello *Multiple-choice*



L'analisi della performance di un item (Wainer, 1989), specie nella fase di calibrazione della domanda stessa, può essere realizzata efficacemente proprio utilizzando grafici del tipo in figura 2. In particolare le funzioni di risposta per ciascuna delle  $m$  opzioni di un item multiple-choice sono espresse dalla seguente relazione (Thissen, Steinberg, 1984):

$$P(X_i = h | \theta) = P_i(\theta) = \frac{\exp[a_h \theta + c_h] + d_h \exp[a_0 \theta + c_0]}{\sum_{k=0}^m \exp[a_k \theta + c_k]} \quad (2)$$

dove  $\theta$  rappresenta la variabile latente. Più precisamente la (2) indica la probabilità che la risposta dell'individuo ricada sull'opzione  $h$ , con  $h = 1, 2, \dots, m$ . La categoria 0 indica, invece, un'opzione fittizia di risposta indicata con *DK* (*don't know*). L'idea è che gli individui che non conoscono la risposta esatta formino una classe teorica la cui funzione di risposta è:

$$P(X_i = 0 | \theta) = \frac{\exp[a_0 \theta + c_0]}{\sum_{k=0}^m \exp[a_k \theta + c_k]} \quad (3)$$

I soggetti che appartengono a questa classe forniscono una risposta scegliendo a caso una determinata opzione ed il parametro  $d$  della (2) rappresenta la proporzione di coloro che scelgono quella risposta in modo casuale.

La lettura dei parametri si differenzia invece nettamente da quella dei modelli che prevedono domande a risposta dicotomica. Il parametro  $a$  rispecchia l'ordinamento delle opzioni: per domande opportunamente calibrate le opzioni con bassi valori di  $a$  (solitamente negativi) mostrano su quasi tutto il dominio di  $\theta$  un andamento decrescente, mentre quella con il valore maggiore di  $a$  (opzione corretta) un andamento sostanzialmente monotono non decrescente. Le alternative con valori intermedi hanno invece una funzione di risposta non monotona.

Il parametro  $c$  rispecchia invece la frequenza relativa di selezione di ciascuna opzione.

## 5. Matematica e Cultura in Europa: il progetto ed il questionario

Come si è accennato nell'Introduzione, la crisi della vocazione scientifica nelle più giovani generazioni ha allarmato i paesi occidentali che hanno cominciato a rivolgere la loro attenzione alle conseguenze di questo preoccupante fenomeno. Si sono così aperti interessanti fonti di ricerca. Il progetto "Matematica e Cultura in Europa" si proponeva di mettere in luce la valenza culturale della matematica nella formazione

del cittadino che è sempre più spesso chiamato ad interpretare in modo rigoroso la realtà che lo circonda, in un contesto culturale che va al di là dei confini nazionali e che richiede di sapersi confrontare in un quadro di riferimento europeo.

Uno degli obiettivi fondamentali del progetto “Matematica e Cultura in Europa” riguardava la misurazione delle competenze matematiche degli studenti che scelgono un percorso formativo con almeno un esame di contenuto matematico. A tale riguardo è stato predisposto un questionario, somministrato alle matricole di alcuni corsi di studio dei cinque stati partecipanti al progetto: Cipro, Francia, Germania, Gran Bretagna e Italia. Il questionario è composto da 14 domande a scelta multipla (multiple choice item) con cinque modalità di risposta di cui una sola esatta. Le domande sono state classificate in tre gruppi tematici omogenei: 1) equazioni e disequazioni, 2) calcolo differenziale ed integrale, e 3) logica, trigonometria e geometria (Mignani e Ricci, 2005).

Il questionario è stato somministrato a 3441 studenti all’inizio dell’anno accademico 2003-2004. La tabella 1 riporta il numero dei partecipanti di ciascun ateneo coinvolto nel progetto.

**Tabella 1.** Studenti partecipanti

<b>Università</b>	<b>Studenti</b>
Bayreuth (D)	221
Bochum (D)	316
Bologna (I)	1648
Bordeaux (F)	60
Catania (I)	275
Cipro (CY)	196
Durham (GB)	392
Freiburg (D)	350
Paris VII (F)	46

## 6. La calibrazione delle domande

L’analisi delle caratteristiche di ciascuna domanda del questionario e conseguentemente la misurazione dell’abilità dello studente è stata concettualmente preceduta dalla verifica della coerenza degli item con il costrutto da misurare. Questo momento è rappresentato dalla fase di calibrazione che ha permesso di selezionare i quesiti adeguatamente formulati sia sotto il profilo della domanda sia delle risposte.

A tal fine la calibrazione è stata effettuata impiegando il modello di Thissen sopra descritto. Secondo questo modello è necessario valutare l'andamento della curva caratteristica di ciascuna opzione di risposta corretta o errata. L'analisi condotta sul questionario<sup>1</sup> ha mostrato una performance accettabile per la maggior parte delle domande. Più in dettaglio, la formulazione dell'opzione corretta della maggior parte delle domande è adeguata per soggetti con livello di abilità medio-alta. Al contrario, per i rispondenti con abilità più bassa la probabilità di selezionare l'opzione corretta non cresce al crescere dell'abilità, anzi per taluni item si registra un andamento decrescente. In alcune domande si evidenzia un comportamento non adeguato, ovvero diverso da quello rappresentato nella figura 2: per livelli di abilità medio-bassi è risultato troppo ampio l'intervallo di valori di abilità in cui la probabilità di scegliere una risposta sbagliata è maggiore di quella di selezionare la modalità giusta. Tuttavia non si è ritenuto opportuno escludere nessuna domanda anche se l'analisi successiva volta allo studio delle abilità non può prescindere da quanto emerso nella fase di calibrazione.

## 7. Le proprietà psicometriche delle domande

Il momento logico successivo a quello della calibrazione è rappresentato dalla determinazione numerica (stima) delle caratteristiche psicometriche di ciascun item e dell'abilità. La realizzazione di questa fase è effettuata utilizzando il modello a tre parametri, più agile dal punto di vista computazionale e più esplicativo sotto il profilo statistico. La scelta di questo modello trova anche giustificazione nel fatto che le opzioni di risposta errate sono tutte equivalenti dal punto di vista del contenuto. Alla luce di ciò si sono ridotte a due le modalità di risposta: corretta e sbagliata. Le tabelle seguenti riportano una classificazione dei tre gruppi di domande in base ai valori dei tre parametri stimati per ciascun item<sup>2</sup>.

**Tabella 2.** Parametro di difficoltà

<b>Livello di <math>b_i</math></b>	<b>Gruppi tematici</b>
Basso	Logica, Geometria e Trigonometria
Medio	Equazioni e Disequazioni
Moderatamente alto	Calcolo integrale e differenziale

<sup>1</sup> L'elaborazione non comprende i dati dei due atenei francesi. Per l'elaborazione è stato utilizzato il software MULTILOG 7.0.3.

<sup>2</sup> Anche in questo caso l'elaborazione è stata effettuata con MULTILOG 7.0.3.



**Tabella 3.** Parametro di discriminazione

Livello di $a_i$	Gruppi tematici
Basso	Logica, Geometria e Trigonometria
Moderatamente basso	Equazioni e Disequazioni
Medio	Calcolo integrale e differenziale

**Tabella 4.** Parametro di *guessing*

Livello di $c_i$	Gruppi tematici
Moderatamente basso	Eq. e Diseq., Calcolo Integrale e Diff.
Moderatamente alto	Logica, Geometria e Trigonometria

Dalla tabella 2 si può rilevare che nessun item ha un elevato parametro di difficoltà. Le domande relative al calcolo integrale e differenziale sono comunque quelle risultate più difficili, ovvero che richiedono un più elevato livello di abilità per avere una probabilità di fornire una risposta corretta pari a 0,5. Un'analisi più approfondita richiede di considerare congiuntamente anche i valori dei parametri di guessing e di discriminazione. Infatti come si deduce dalle tabelle 3 e 4 gli item più facili sono quelli che meno discriminano soggetti con abilità diverse e che presentano guessing più elevato. Questo può determinare una ridotta capacità del questionario di graduare adeguatamente gli studenti con abilità più basse.

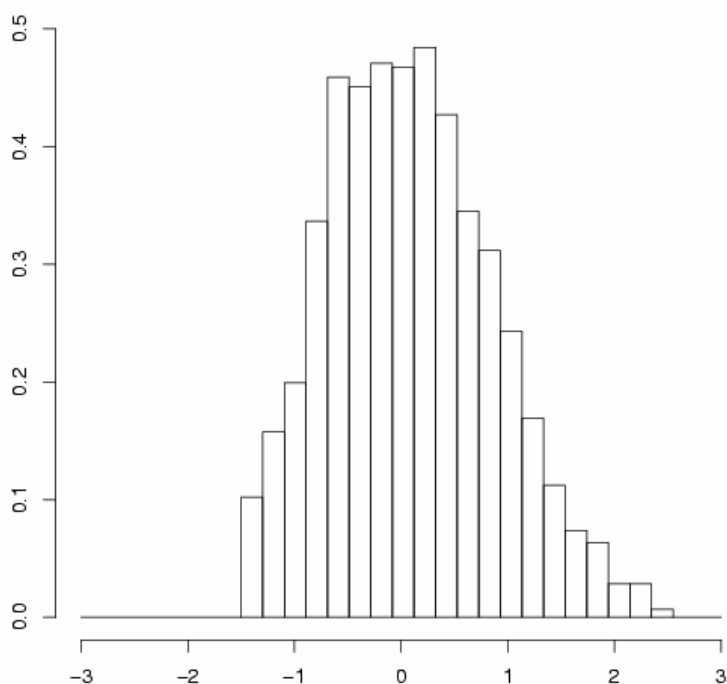
## 8. La stima delle abilità degli studenti

L'analisi basata semplicemente sul numero di risposte corrette non permette di cogliere adeguatamente le differenze tra le performance degli studenti. I risultati ottenuti con gli strumenti di IRT offrono invece la possibilità di graduare con maggiore dettaglio il livello di preparazione matematica raggiunto da ogni rispondente. Infatti una delle peculiarità principali dell'IRT è quella di determinare simultaneamente la stima dei parametri che descrivono le proprietà psicometriche di ciascuna domanda e l'abilità di ogni studente. Quest'ultima è ottenuta considerando congiuntamente le risposte a tutte le domande del questionario. In particolare l'abilità di un soggetto è calcolata associando le risposte a ciascuna domanda mediante la funzione di verosimiglianza:

$$\prod_{i=1}^n [P_i(\theta)]^{u_i} [1 - P_i(\theta)]^{1-u_i} \quad (4)$$

dove  $n$  è il numero di item che compongono il questionario (nel caso in esame 14) e dove  $u_i$  assume il valore 1 se  $X_i = 1$  e 0 se  $X_i = 0$ .

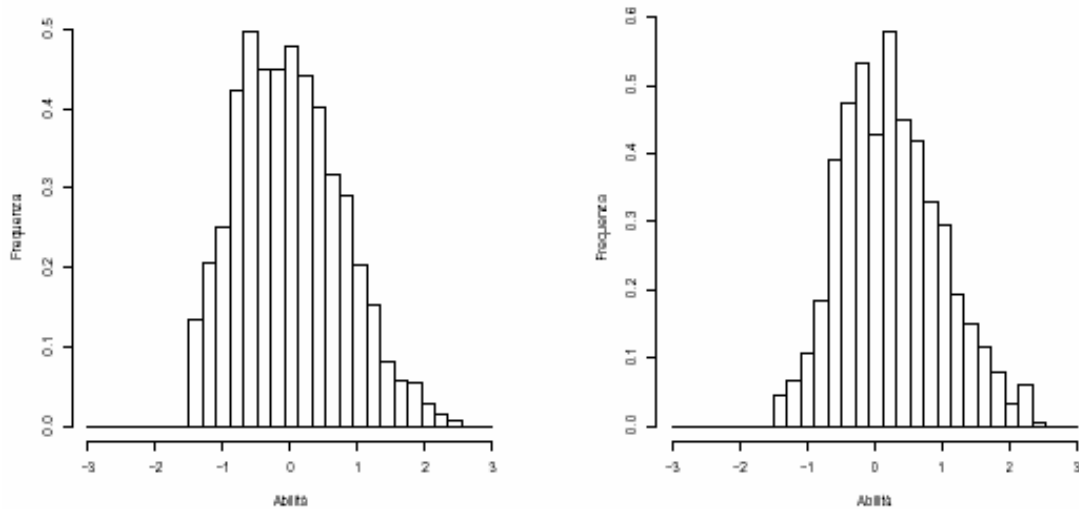
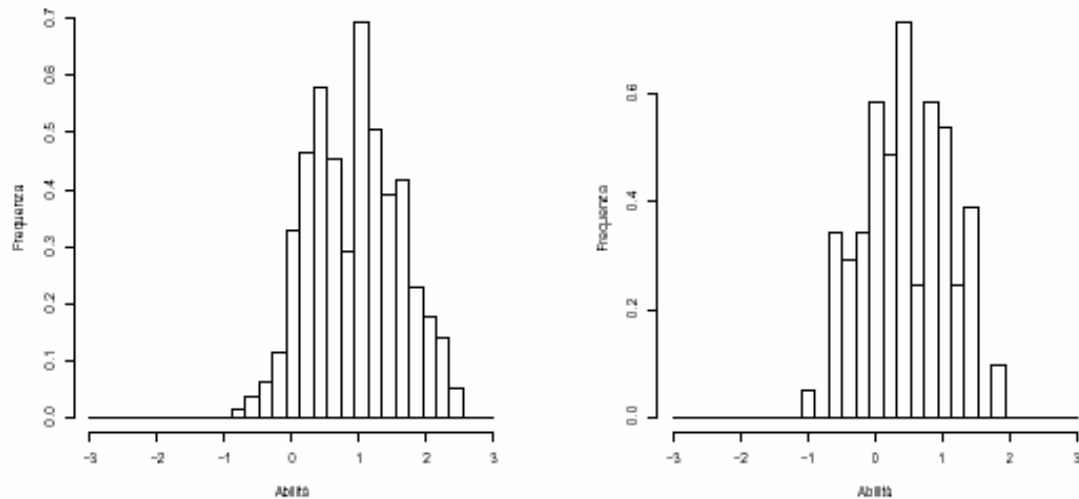
**Figura 3.** *Descrizione complessiva delle abilità*



Il valore di  $\theta$  di ogni individuo è determinato applicando alla (4) la procedura di stima *marginal maximum likelihood* e considerando il modello a tre parametri per dati dicotomici. La distribuzione delle abilità così stimate per l'intero gruppo di studenti è illustrata nella figura 3.

La distribuzione evidenzia un comportamento sostanzialmente soddisfacente dell'abilità degli studenti. Occorre sottolineare però che, come evidenziato, il questionario non gradua con sufficiente precisione gli studenti con abilità basse. Infatti la distribuzione in figura 3 presenta asimmetria positiva, ovvero la coda sui valori positivi è più lunga di quella sui valori negativi. Si nota inoltre che il massimo effettivo raggiunto dalle abilità è abbastanza vicino a quello teorico, mentre ciò non accade per il valore minimo, confermando ulteriormente la capacità del questionario di cogliere meglio le performance degli studenti con abilità alte. Secondo l'impostazione IRT, stimati i parametri di ciascuna domanda, è possibile calcolare le abilità per sottogruppi di rispondenti per individuare eventuali similarità o differenze. In particolare è possibile poi analizzare le performance di singoli gruppi di studenti. Di seguito si confrontano i risultati rispetto alla nazione di appartenenza.

La figura 4 illustra a sinistra la distribuzione dell'abilità degli studenti italiani e a destra quella degli studenti tedeschi.

**Figura 4.** *Distribuzione dell'abilità - Italia e Germania***Figura 5.** *Distribuzione dell'abilità - Gran Bretagna e Cipro*

La figura 5, a sua volta, riporta sulla sinistra la distribuzione dell'abilità degli studenti britannici e sulla destra quella degli studenti ciprioti.

Dal confronto delle distribuzioni non emergono differenze sostanziali e si conferma quanto sottolineato per valori bassi dell'abilità. Tuttavia gli studenti britannici e ciprioti presentano una performance migliore. Occorre però tenere conto che i gruppi non sono omogenei rispetto al corso di studio intrapreso, infatti le matricole della Gran Bretagna e di Cipro sono iscritte a corsi di laurea a forte contenuto scientifico.

## 9. Il caso dell'Ateneo di Bologna

Un'ulteriore analisi è stata condotta considerando esclusivamente i risultati dell'Ateneo di Bologna.

La tabella 5 riporta per ogni corso di studio l'abilità media stimata in base al modello di IRT e alla percentuale di risposte esatte.

Nella prima colonna della tabella i corsi di studio sono ordinati in base al livello di abilità. Nella seconda colonna la graduatoria decrescente è riportata tra parentesi. Come di può notare le differenze tra le due graduatorie sono più evidenti per i corsi di studio con performance peggiore a ulteriore conferma che la sola percentuale di risposte corrette non permette bene di cogliere le differenze soprattutto per le basse abilità.

La molteplicità degli aspetti legati alla valutazione del processo di insegnamento e apprendimento richiede il ricorso ad un apparato metodologico che sia in grado di coglierne le diverse sfaccettature. L'analisi effettuata si inserisce in questo contesto di ricerca e ha permesso di valutare in modo articolato ed unitario sia la capacità misuratoria dello strumento adottato, ovvero il questionario, sia il livello di cultura matematica degli studenti coinvolti nel progetto.

**Tabella 5.** *Parametro di guessing*

<i>Corso di laurea</i>	<b>Ordinamento in base a:</b>	
	<i>Abilità media</i>	<i>% Risp</i>
Fisica	0.78	57.2 (1)
Scienze dell'Informazione	0.70	54.3 (2)
Matematica	0.50	48.8 (3)
Chimica Industriale (BO)	0.46	48.6 (4)
Cons. Beni Culturali	0.34	44.8 (5)
Biotechnologie	0.34	43.9 (6)
Scienze Statistiche (BO)	0.25	40.9 (7)
Astronomia	0.14	39.8 (9)
Chimica, Chimica dei Materiali	0.12	40.4 (8)
Informatica	0.08	38.6 (10)
CTF	0.03	36.7 (11)
Farmacia	-0.12	33 (12)
Tecn. Chim. Amb. Gest. Rif.	-0.21	32.5 (13)
Scienze di Internet	-0.25	31.3 (14)
Economia	-0.31	29.4 (15)
Scienze Biologiche	-0.46	25.8 (17)
Scienze Naturali	-0.47	26.6 (16)
Agraria (Bologna)	-0.57	10.4 (20)
Chimica Industriale (Faenza)	-0.58	22.1 (19)
Agraria (Imola)	-0.64	22.4 (18)

Al di là delle aspettative, i risultati ottenuti sembrano indicare una preparazione matematica sostanzialmente adeguata delle nuove matricole, anche se il questionario ha rivelato alcune criticità, nei termini dei modelli statistici di validazione adottati, che non devono essere trascurate e che suggeriscono consistenti modifiche in vista di una futura replicazione della ricerca.

### Riferimenti bibliografici

- R. K. HAMBLETON, H. SWAMINTHAN (1985) *Item Response Theory*, Kluwer Nijhoff Publishing
- F.M. LORD, M.E. NOVICK (1968) *Statistical theories of mental test scores*, Addison-Wesley Publishing Co.
- M. MANARESI (ed.) (2005) *Matematica e Cultura in Europa*, Springer-Verlag, Berlin
- S. MIGNANI; R. RICCI (2005) Metodi statistici per la misurazione delle competenze, in M. Manaresi (ed.) *Matematica e Cultura in Europa*, Springer-Verlag, Berlin: 2-17
- G. RASCH (1960) *Probabilistic Models for Some Intelligence and Attainment Tests*, Danish Institute for Educational Research, Copenhagen.
- D.THISSEN, L. STEINBERG (1984) A response model for multiple choice items. *Psychometrika*, **49**: 5011- 5019
- H. WAINER (1989) The future of item analysis, *Journal of Educational Measurement*, **26**: 191- 208

***Mathematical Skill Evaluation:  
Some Results of an European Project on Freshmen***

***Summary.*** *The skill evaluation insists on several methodological fields and has to take into account many aspects. In this paper a methodology called Item Response Theory (IRT) which allows a quality content evaluation of the questionnaire and an efficient ability estimation is considered. First, the multiple-choice IRT model is proposed in order to verify the capability of each item to measure in an appropriate way the student ability. Second, the three-parameter IRT model is applied to measure the psychometric features of each question and to estimate the freshmen mathematical knowledge. Furthermore, the IRT methodology allows to break down into groups the items according to their difficulty and their discrimination capability.*

***Keywords.*** *Item Response Theory, Latent variables, psychometric properties of an item, ability evaluation*