ALMA MATER STUDIORUM
UNIVERSITA' DI BOLOGNA

I MODELLI DELLE VALUTAZIONI COMPARATIVE IN MATEMATICA

Tesi di Laurea in Didattica della Matematica

Relatore:
Chiarissimo Professore
Giorgio Bolondi

Laureanda:
Roberta Righetti

Sessione Terza
Anno Accademico 2013/2014

La matematica è una scienza in cui non si sa di che cosa si parla,
nè se ció che si dice è vero.
Bertrand Russel

Contents

Chapter 1

Introduzione

Il dibattito internazionale sugli obiettivi dell'insegnameto della matematica evidenzia sempre piú una concezione delle competenze matematiche come un complesso di processi basati sulla "*matematizzazione*", ovvero, processi fondati sulla modellizzazione matematica di situazioni problematiche, soprattutto del mondo reale.
Il riferimento al reale oggi è ritenuto molto importante a livello internazionale. E' necessario possedere un certo livello di comprensione della matematica, degli strumenti matematici e del ragionamento matematico, per poter comprendere a fondo un numero sempre crescente di problemi e situazioni della vita quotidiana e farvi fronte. E' dunque imporante riuscire a comprendere se il livello di preparazione dei ragazzi che giungono al termine dell'obbligo scolastico sia tale da consentire loro di essere adeguatamente preparati ad applicare la matematica per comprendere temi importanti e risolvere problemi reali.
In questo contesto si inseriscono i Sistemi di Valutazione Internazionali della Matematica, come PISA e TIMSS, rispetto ai quali, sebbene non sia possibile individuare relazioni di causa-effetto tra le politiche/pratiche dell'istruzione e i risultati degli studenti è peró possibile ottenere una rappresentazione delle maggiori differenze e similitudini tra i diversi sistemi educativi, con le relative implicazioni di quello che puó significare per gli studenti.
Oltre a ció, la possibilitá di monitorare nel tempo i risultati dei processi di acquisizione di conoscenze e abilitá, in contesti nazionali diversi, offre spunti di riflessione all'azione pubblica.
D'altra parte il rischio che una prospettiva di insegnamento della matematica incentrata fondamentalmente su ció che "è concreto, vicino, familiare" faccia perdere di vista gli oggetti di natura teorica e filosofica è una preoccupazione altrettanto ragionevole, per gli effetti che questo potrebbe avere nell'esplicarsi della ricerca matematica o nella sua evoluzione epistemologica.

La sfida e la scelta che nel contempo si pone è di un'equilibrata compresenza delle due componenti: modellizzazione matematica di situazioni problematiche all'interno di una teoria sempre piú sistematica.
La formazione del curricolo scolastico non puó prescindere dal considerare sia la funzione strumentale sia quella culturale della matematica: strumento essenziale per una comprensione quantitativa della realtá da un lato e dall'altro sapere logicamente coerente e sistematico, caratterizzato da una forte unitá culturale.

Chapter 2

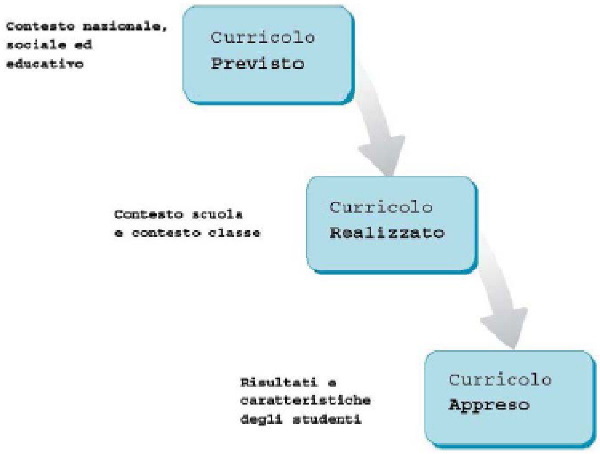
TIMSS e PISA Strumenti di Indagine Internazionale

2.1 Intenti e modello teorico di riferimento

La valutazione dell'efficacia dell'istruzione e piú in generale dei processi formativi è un problema di forte interesse. In questo contesto si inseriscono due, tra le principali indagini valutative internazionali, nelle quali l'interesse è quello di effettuare valutazioni comparative dei livelli e delle capacitá di apprendimento per la popolazione studentesca.
Le indagini TIMSS (Trends in International Mathematics and Science Study) e PIRLS (Progress in International Reading Literacy Study) si basano sul quadro teorico della IEA, un'associazione internazionale indipendente di enti nazionali di ricerca educativa e di enti governativi di ricerca che si occupano del miglioramento dell'istruzione. L'obiettivo della IEA è fornire informazioni di alta qualitá sui risultati del rendimento degli studenti e sui contesti educativi in cui gli stessi raggiungono tali risultati. L'indagine TIMSS, in particolare, ha come obiettivo la rilevazione degli apprendimenti in matematica e scienze al quarto e ottavo anno di scolaritá. Noi ci occuperemo solo dell'indagine relativa all'ottavo anno di scolaritá (corrispondente al termine della scuola secondaria di primo grado). TIMSS utilizza il "curricolo" nel senso piú ampio del termine, come principale concetto organizzatore per comprendere le strategie didattiche utilizzate e individuare i fattori che possono influenzarne l'efficacia.

L'indagine è strutturata tenendo conto di tre distinte nozioni di curricolo: curricolo previsto (il piano di studi stabilito a livello nazionale) il curricolo realizzato (quello che è effettivamente insegnato) e curricolo appreso (quello che gli studenti hanno imparato), lungo due dimensioni: una relativa ai processi cognitivi coinvoli e una relativa al contenuto degli insegnamenti in matematica e scienze.

Figura 1 : Il modello di curricolo di TIMSS.



Partendo da questo modello, TIMSS utilizza le prove congnitive per rilevare i livelli di rendimento degli studenti, nei vari Paesi, in matematica e scienze. Alle prove cognitive si accompagnano alcuni questionari che consentono di raccogliere informazioni sulle variabili di contesto che possono essere utili per interpretare i risultati conseguiti dagli studenti nelle prove cognitive.

L'indagine PISA (Programme for International student Assesment) rappresenta l'indagine per la rilevazione delle competenze matematiche relativamente ai paesi OCSE, su studenti di 15 anni, focalizzata sulla literacy matematica piuttosto che sul curricolo. Il quadro di riferimento di PISA 2012 definisce literacy matematica come la capacitá di una persona di formulare, utilizzare e interpretare la matematica in svariati contesti. In questo senso si tratta della capacitá individuale di ragionare matematicamente e di usare concetti matematici, procedure, fatti e strumenti per descrivere, spiegare e prevedere fenomeni. La formulazione in forma matematica include il capire quali sono le occasioni in cui applicare e usare la matematica. L'utilizzo della matematica prevede l'applicazione del ragionamento matematico e l'utilizzo di concetti, procedimenti, fatti e strumenti matematici per arrivare ad una soluzione matematica. L'interpretazione matematica prevede la riflessione su soluzioni o risultati matematici e la conseguente interpretazione nel contesto del problema.

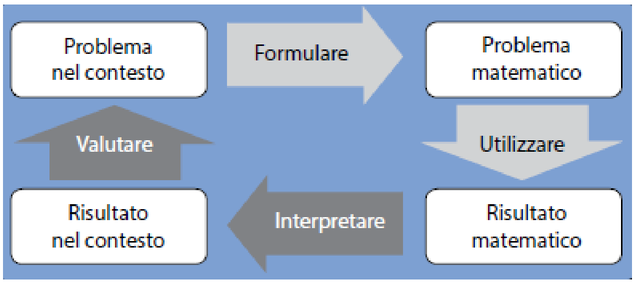


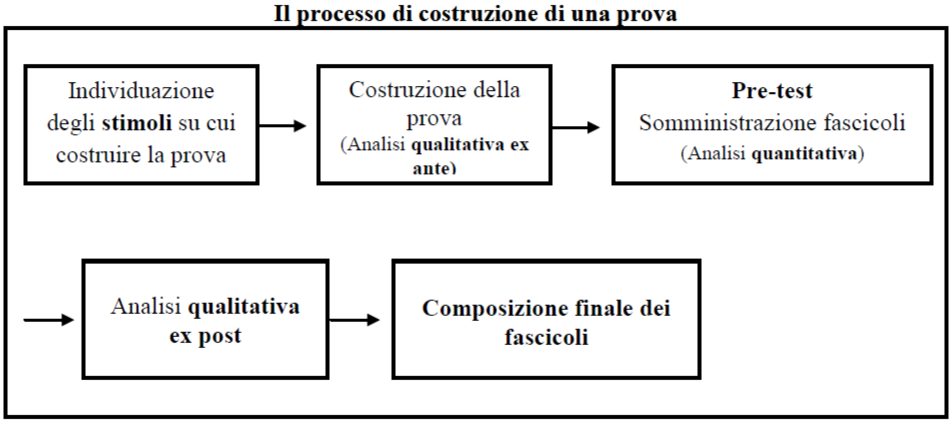
Figure 2.1: L'applicazione pratica del modello di literacy

E' importante concepire la literacy matematica non come un attributo individuale che si possiede oppure no, ma come una capacitá che puó essere sviluppata lungo tutto l'arco della vita.
Quindi gli intenti di TIMSS e PISA sono, in modo esplicito, differenti. Mentre TIMSS si concentra sull'estensione con la quale gli studenti padroneggiano la matematica (e le scienze) per come appaiono nei curriculum scolastici, l'intento di PISA è "catturare l'abilitá di usare conoscenza e capacitá nell'affrontare le sfide ricavate dalla *vita vera*" (OECD, 2001). E ancora, per citare Barry McGaw: TIMSS è interessato a scoprire "che cosa della scienza ti è stato insegnato e quanto hai imparato", mentre PISA: " che cosa puoi fare con la scienza che ti è stata insegnata".

2.2 Gli strumenti utilizzati

Le prove standardizzate e gli strumenti utilizzati per la misurazione degli apprendimenti forniscono misure sufficientemente attendibili del grado di padronanza di quelle conoscenze e di quei processi che sono l'oggetto della prova stessa. La costruzione di prove standardizzate (comune alle due indagini) prevede come base di partenza la predisposizione di un quadro teorico di riferimento in cui sono descritti gli oggetti della misurazione e le caratteristiche delle prove. La costruzione di una prova standardizzata è il risultato di un'attivitá di ideazione reperimento di materiali, stesura, verifica, correzione e altro ancora avente le caratteristiche di un percorso di ricerca sperimentale (vedi schema).
L'informazione statistica che scaturisce è generalmente di tipo dicotomico (nel senso che gli items sono corretti in temini di risposta "giusta" o "sbagliata") e l'obiettivo che si persegue è quello della misurazione delle capacitá sottostanti (o dimensioni latenti) il rendimento del soggetto nella prova. Tale fase riveste un'importanza notevole nell'intero processo di costruzione della prova perché è il momento in cui si hanno i primi riscontri empirici rispetto al lavoro realizzato. Ma preliminare a tale misurazione si pone il problema di calibratura del test, volto ad escludere quelle domande che risulterebbero inefficaci al fine della misurazione e di scoring, una tecnica per la stima delle abilitá degli individui. Nella prossima sezione entreremo nel dettaglio del modello di costruzione del test.

Il processo di costruzione di una prova



Per la costruzione dei framework (prima fase) si utilizza (per entrambe le indagini) un campionamento a matrice che comporta la suddivisione di tutto l'insieme dei quesiti di matematica (e di scienze) in una serie di fascicoli cognitivi. I fascicoli di prova delle indagini internazionali sugli apprendimenti sono costruiti a partire da insiemi di item che coprono l'intero spettro di contenuti e processi che di volta in volta si intendono misurare e che sono in un secondo momento suddivisi in blocchi o sottogruppi, equivalenti per gamma di difficoltá delle domande. I blocchi vengono quindi variamente combinati tra loro in modo da ottenere diversi fascicoli di prova con lo stesso numero totale di domande. In questo modo i fascicoli di prova utilizzati ad ogni tornata delle indagini non sono identici per tutti gli studenti oggetto di rilevazione, anche se il livello di difficoltá complessivo di ogni fascicolo rimane all'incirca il medesimo e la distribuzione dei fascicoli agli studenti avviene con rotazione sistematica, cosí da assicurare che ciascun blocco di domande all'interno dei fascicoli sia assegnato ad uno stesso numero di studenti.
A tale proposito si applicano tecniche avanzate di test equating (rinviamo alla sezione che segue).
Relativamente a TIMSS i contenuti matematici, su cui vertono i quesiti, sono classificati in: *Numero*, *Algebra*, *Geometria*, *Dati e Probabilitá*. La dimensione cognitiva, lungo la quale si articola il framework, è classificata in: *Conoscenza*, *Applicazione* e *Ragionamento*. Ciascun dominio di contenuto include quesiti sviluppati per valutare gli studenti in ciascuno dei tre domini di contenuto.
Ai fini della rilevazione PISA, la literacy matematica puó essere analizzata tenendo conto della struttura organizzativa dei processi matematici coinvolti. Pertanto le categorie cognitive che costituiscono il fondamento teorico delle prove sono: *Formulazione* di situazioni in forma matematica, *Utilizzo* di concetti, procedimenti e ragionamento matematico, *Interpretazione* in termini di applicazione e valutazione dei risultati matematici.
Le categorie di contenuto che hanno orientato lo sviluppo degli item per il ciclo 2012 sono: *Cambiamento e Relazioni, Spazio e Forma, Quantitá, Incertezza e Dati*. Tale scelta, piuttosto che riproporre la ripartizione scolastica dei diversi argomenti, come aritmetica, geometria e algebra, riflette piuttosto gli ambiti di suddivisione storica della matematica considerando che, nella realtá extra-scolastica, una sfida o una situazione non si presenta quasi mai accompagnata da regole o prescrizioni su come affrontarla.

2.3 Domini Cognitivi

Per rispondere correttamente ai quesiti dell'indagine gli studenti devono avere una certa familiaritá con i contenuti di matematica oggetto di rilevazione, ma devono anche dimostrare di avere un certo numero di abilitá cognitive. Ciascun quesito della rilevazione quindi è associato a un dominio di contenuto e a un dominio cognitivo e in tal modo esso dá un'idea sia dal punto di vista del contenuto sia dal punto di vista cognitivo del livello di apprendimento di uno studente in matematica. Vediamo come, compatibilmente allo sviluppo dell'indagine TIMSS, sono descritte queste abilitá. Il primo dominio, *Conoscenza*, riguarda i fatti, i concetti e le procedure che gli studenti devono conoscere; il secondo dominio *Applicazione* è incentrato sull'abilitá degli studenti di applicare nozioni e conoscenze concettuali; il terzo dominio *Ragionamento*, va oltre la soluzione di problemi di toutine per includere situazioni non familiari.
La facilitá nell'uso della matematica o del ragionamento in determinate situazioni, dipende dalle conoscenze matematiche e dalla familiaritá con i concetti matematici. I fatti comprendono la conoscenza effettiva che fornisce il linguaggio base della matematica, i fatti matematici essenziali e le proprietá che costituiscono il fondamento del pensiero matematico.
I procedimenti costituiscono un ponte tra le conoscenze di base e l'uso della matematica per risolvere i problemi. Il dominio *Applicazione* prevede l'applicazione degli strumenti matematici ad una varietá di contesti. I problemi di routine sono esercizi svolti normalmente in classe e possono essere ambientati in situazioni di vita reale o possono riguardare soltanto questioni matematiche. Saper risolvere i problemi è un punto centrale del dominio *Applicazione*, ma i contesti dei problemi sono piú familiari rispetto a quelli del dominio *Ragionamento*, essendo ben radicati nei curricoli.

Alla base del dominio *Applicazione* sono riconosciute le seguenti capacitá:

Selezionare	selezionare un'efficiente/appropriata operazione o una strategia per risolvere problemi che prevedono: una procedura conosciuta, un algoritmo un metodo di soluzione
Rappresentare	disporre informazioni matematiche e dati in: diagrammi, tabelle o grafici e generare equivalenti rappresentazioni per una data relazione o entita' matematica
Modellizzare	generare un modello appropriato, come ad esempio un'equazione, una figura geometrica o un diagramma per risolvere un problema di routine
Implementare	implementare un set di istruzioni matematiche, ad esempio: ricavare forme e diagrammi, da specificazioni fornite
Risoluzione di problemi di routine	risolvere problemi standard simili a quelli incontrati in classe. I problemi possono essere contestuali a situazioni familiari o strettamente matematici

(sic)

Il *Ragionamento* matematico riguarda la capacitá di pensare in modo logico e sistematico. Include il ragionamento intuitivo e induttivo basato su schemi e regolaritá che si possono usare per arrivare alla soluzione di problemi non di routine. Entrambi i tipi di quesiti, problemi di routine e non (ricavati ad esempio da situazioni di vita reale) includono il trasferimento di conoscenze e abilitá a nuove situazioni e spesso é necessaria l'interazione fra diverse abilitá di ragionamento a causa della novitá del contesto, della complessitá della situazione.

Nella tabella che segue sono indicate le capacitá sottese al dominio *Ragionamento*:

Analizzare	determinare, descrivere o usare relazioni fra variabili e oggetti in situazioni matematiche e fare valide inferenze da informazioni date
Generalizzare/ Specializzare	estendere il dominio di applicabilita' del pensiero matematico e del problem solving
Integrare/ Sintetizzare	fare connessioni fra differenti elementi della conoscenza e rappresentazioni tra loro correlate, e creare dei link fra le idee matematiche corrispondenti. Combinare fatti matematici, concetti e procedure per stabilire risultati e combinare tali risultati per produrre ulteriori risultati
Giustificare	saper fornire una giustificazione tramite riferimenti, a risultati o a proprieta' matematiche conosciute
Risoluzione di problemi di non routine	risolvere problemi contestualizzati alla Matematica o a situazioni di vita reale e applicare fatti, concetti e procedure matematica ad una certa varieta' e complessita' di casi

(sic)

Nella tabella che segue sono riportati, per ogni dominio cognitivo, il numero di quesiti per ciascuna tipologia (cioè distinguendo le domande a scelta multipla da quelle a risposta aperta). Come si puó vedere il dominio cognitivo maggiormente rappresentato è quello dell'Applicazione mentre quello meno rappresentato è quello del Ragionamento.

Domande della prova TIMSS	**Domande a scelta multipla**	**Domande a risposta aperta**	**Numero totale**	**Percentuale del punteggio**
Dominio cognitivo				
Conoscenza	53 (53)	27 (30)	80 (83)	36%
Applicazione	47 (47)	38 (44)	85 (91)	39%
Ragionamento	18 (18)	34 (40)	52 (58)	25%
Totale	118 (118)	99 (114)	217 (232)	100%
Percentuale del punteggio	51%	49%		

Figure 2.2: Distribuzione del numero di quesiti TIMSS (2011) per domini cognitivi e tipologia di item

Gli item PISA sono costruiti in modo da analizzare tre aspetti interconnessi:

- i processi matematici che descrivono ció che fanno gli studenti per collegare il contesto del problema alla matematica e quindi risolvere il problema
- il contenuto matematico da utilizzare negli item
- i contesti

La literacy matematica, che ricordiamo è la capacitá di formulare, utilizzare e interpretare la matematica in svariati contesti, puó essere analizzata sulla base di alcune capacitá matematiche che sono alla base di ciascuno dei processi riferiti. Quando il livello di literacy matematica di uno studente migliora, questi sará in grado di fare affidamento in misura sempre maggiore sulle capacitá matematiche fondamentali. Il quadro di riferimento (2012) di PISA utilizza una formulazione di tali competenze sulla base delle analisi relative al loro utilizzo attraverso item somministrati in precedenza. Le sette capacitá matematiche fondamentali sono: la comunicazione, la matematizzazione, la rappresenazione, il ragionamento e l'argomentazione, l'elaborazione di strategie, l'utilizzo di un linguaggio simbolico e l'utilizzo di strumenti matematici. Di seguito vediamo come tali capacitá intervengono a vari livelli di ciascuno dei tre processi.

	Formulazione di situazioni in forma matematica	Utilizzo di concetti, fatti, procedimenti e ragionamenti matematici	Interpretazione, applicazione e valutazione dei risultati matematici
Comunicazione	Leggere, decodificare e interpretare affermazioni, domande, compiti, oggetti, immagini o animazioni (nella rilevazione computerizzata) al fine di creare un modello mentale della situazione	Articolare una soluzione, illustrare il lavoro necessario per arrivare alla soluzione e/o riassumere e presentare i risultati matematici intermedi	Elaborare e comunicare spiegazioni e argomentazioni nel contesto del problema
Matematizzazione	Identificare le variabili e le strutture matematiche soggiacenti al problema reale e formulare delle ipotesi per poterle utilizzare	Utilizzare la comprensione del contesto per orientare o organizzare il processo matematico di risoluzione, ad es. lavorare con un livello di precisione adeguato al contesto	Comprendere la portata e i limiti di una soluzione matematica derivanti dall'impiego di un determinato modello matematico
Rappresentazione	Creare una rappresentazione matematica delle informazioni del mondo reale	Dare un senso, mettere in relazione e usare una molteplicità di rappresentazioni nell'interazione con il problema	Interpretare i risultati matematici in diversi formati in relazione alla situazione o all'utilizzo; confrontare o valutare due o più rappresentazioni in relazione a una situazione
Ragionamento e argomentazione	Spiegare, difendere o giustificare la rappresentazione della situazione reale elaborata o individuata	Spiegare, difendere o giustificare il processo e i procedimenti usati per determinare un risultato o una soluzione di natura matematica Collegare le informazioni per giungere a una soluzione matematica, elaborare generalizzazioni o creare argomentazioni a più livelli	Riflettere sulle soluzioni matematiche ed elaborare spiegazioni e argomentazioni che supportino, confutino o qualifichino una soluzione matematica a un problema contestualizzato
Elaborazione di strategie per la risoluzione dei problemi	Selezionare o elaborare un piano o una strategia per inquadrare i problemi contestualizzati in forma matematica	Attivare meccanismi di controllo efficaci nel corso delle varie fasi del procedimento che porta a una soluzione, conclusione o generalizzazione matematica	Elaborare e mettere in atto una strategia finalizzata a interpretare, valutare e convalidare una soluzione matematica a un problema contestualizzato
Utilizzo di un linguaggio simbolico, formale e tecnico e di operazioni	Utilizzare variabili, simboli, diagrammi e modelli standard adeguati al fine di rappresentare un problema reale attraverso un linguaggio simbolico/formale	Comprendere e utilizzare costrutti formali basati su definizioni, regole e sistemi formali; utilizzare algoritmi	Comprendere la relazione esistente tra il contesto del problema e la rappresentazione della soluzione matematica.Usare tale comprensione per orientare l'interpretazione della soluzione nel contesto e determinarne la plausibilità e le possibili limitazioni
Utilizzo di strumenti matematici	Usare strumenti matematici per riconoscere le strutture matematiche o per delineare relazioni matematiche	Conoscere e saper utilizzare adeguatamente i diversi strumenti che possono essere utili durante i processi e i procedimenti finalizzati alla ricerca delle soluzioni	Usare strumenti matematici per accertare la plausibilità di una soluzione matematica ed eventuali sue limitazioni e restrizioni in base al contesto del problema

Figure 2.3: Relazione tra processi matematici e capacitá matematiche fondamentali

La facilitá con cui gli studenti applicano la matematica a una molteplicitá di problemi e situazioni dipende dalle competenze inerenti a ciascuno dei tre processi. Anche se il processo cognitivo a cui si da maggiore peso é: utilizzo di concetti, fatti, procedimenti e ragionamento, l'obiettivo è di arrivare ad un equilibio che presenti una ponderazione all'incirca equivalente tra i due processi che comportano il fare un collegamento tra il mondo reale e quello matematico (la formulazione, l'interpretazione) e il processo che chiede agli studenti di dimostrare la capacitá di lavorare su un problema formulato matematicamente (il ragionamento)

Categoria di processo	Punteggio percentuale
Formulazione di situazioni in forma matematica	**Approssimativamente 25**
Utilizzo di concetti, fatti, procedimenti e ragionamenti matematici	**Approssimativamente 50**
Interpretazione, applicazione e valutazionedei risultati matematici	**Approssimativamente 25**
TOTALE	**100**

Figure 2.4: Distribuzione approssimativa del punteggio per categoria di processi nel ciclo PISA 2012

In tale processo (utilizzo di concetti, fatti, procedimenti e ragionamenti matematici) gli studenti mettono in atto tutti i procedimenti necessari per ottenere i risultati e giungere a una soluzione matematica. Lavorano su un *modello* della situzione problematica, stabiliscono delle *regolaritá*, identificano *collegamenti* tra le entitá matematiche e creano argomentazioni matematiche. Torneremo su queste precise categorie o procedure del pensiero nella sezione delle classi di coordinazione.
PISA valuta la misura in cui gli studenti di 15 anni sono in grado di utilizzare la matematica quando si trovano davanti a determinare situazioni e problemi, pertanto, un aspetto importante della literacy matematica è che la matematica serve a risolvere problemi situati in un contesto. Il contesto è l'aspetto del mondo reale di una persona in cui si situa il problema e la scelta di appropriate strategie e rappresentazioni matematiche spesso dipende dal contesto in cui insorge il problema. Per il quadro di riferimento PISA 2012 per la matematica sono state definite quattro categorie di contesti utilizzati, personale, professionale, sociale e scientifico. Infatti le categorie selezionate (relative ai contesti): personale, professionale, sociale e scientifico, riflettono un'ampia gamma di situazioni nelle quali gli studenti potrebbero incontrare opportunitá matematiche e sono quindi utilizzate come quadro contestuale nella costruzione degli items.

Gli items selezionati per le prove dell'indagine (TIMSS e PISA) devono presentare un'ampia gamma di difficoltá per corrispondere alle diverse abilitá degli studenti partecipanti. I gradi di difficoltá sono determinati nel quadro di una prova sul campo condotta prima della selezione degli items (fase: analisi qualitativa *ex ante*).

2.4 Quadro internazionale dei risultati: TIMSS e PISA a confronto

In questa sezione prenderemo in esame i risultati che gli studenti, in particolare dell'Italia e della Finlandia, hanno conseguito all'indagine TIMSS 2011 e PISA 2012. Nello specifico verranno analizzate le medie e le distribuzioni dei risultati in matematica dei diversi Paesi e la distribuzione delle prestazioni degli studenti. La scala dei risultati di TIMSS (come per PISA) è stata stabilita, fin dalla prima rilevazione del 1995 (2003), con una media uguale a 500 e una deviazione standard uguale a 100 e tale media è rimasta costante anche nelle rilevazioni successive. Nella parte sinistra della figura, relativamente a TIMSS 2011, i Paesi sono disposti in ordine decrescente di punteggio ottenuto e quelli che hanno ottenuto risultati significativamente, dal punto di vista statistico, superiori o inferiori alla media TIMSS presentano accanto al punteggio una freccia rivolta rispettivamente o verso l'alto o verso il basso. L'Italia ha ottenuto un punteggio di 498, che dal punto di vista statistico non è significativamente diverso dalla media di TIMSS. Tra i primi nove migliori Paesi, tutti con punteggio significativamente superiore alla media TIMSS, troviamo la Finlandia che ha ottenuto un punteggio di 514.
Oltre ai punteggi di ogni singolo Paese nella parte destra della figura sono riportate le barre orizzontali che rappresentano la distribuzione dei risultati della prova di matematica. Considerare l'intera distribuzione consente di avere un quado molto piú completo dei diversi livelli degli studenti che hanno sostenuto la prova e consente anche di focalizzare l'attenzione da un lato sulle eccellenze dall'altro su soggetti piú in difficoltá. La lunghezza complessiva di ciascuna barra dá un'idea della dispersione dei punteggi, dispersione maggiore quindi corrisponde ad un divario maggiore tra gli studenti. Le zone in colore celeste chiaro e scuro corrispondono ai punteggi in termini percentili, la fascetta nera al centro rappresenta l'intervallo di confidenza entro cui cade la media della popolazione studenti. Se si immagina di tracciare due rette verticali in corrispondenza delle due estremitá di ogni barra, si puó vedere quale sia il punteggio dello studente che si situa al 5^o o al 95^o percentile in ciascun Paese.

Ad esempio per quanto riguarda l'Italia, lo studente che si trova al 5° percentile consegue un risultato di 372 punti, mentre lo stesso studente in Finlandia ha un punteggio di 405 punti. Ció significa, in termini comparativi, che gli studenti finlandesi nella parte piú bassa della distribuzione (e quindi piuttosto lontano dalla media della popolazione studenti) conseguono risultati superiori di 60 punti rispetto agli studenti italiani: gli studenti "meno bravi" della Finlandia sono quindi meno poveri di competenze dei loro omologhi in Italia.

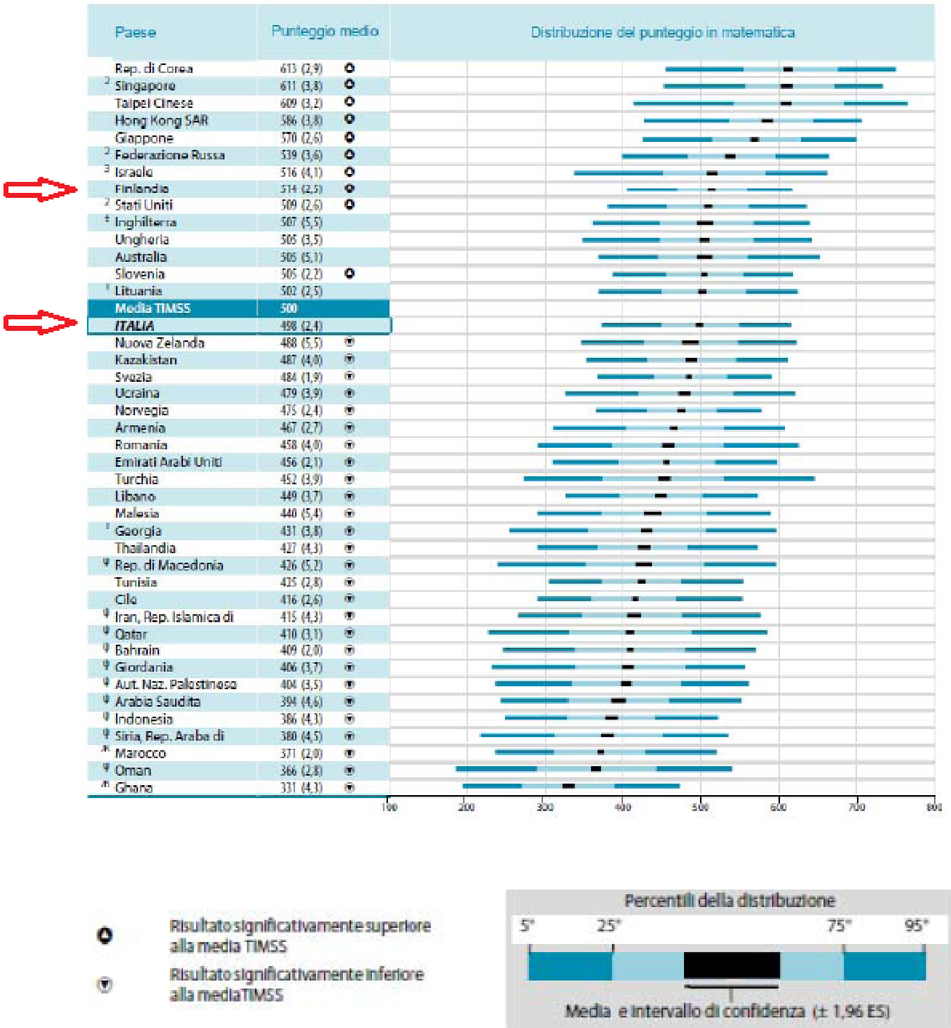


Figure 2.5: Media e dispersione nella scala complessiva di matematica - ottavo anno di scolaritá-TIMSS 2011

Come è stato descritto in precedenza, ogni quesito dell'indagine TIMSS è associato a un dominio di contenuto e a un dominio cognitivo. Ora presenteremo alcuni risultati ottenuti dall'Italia in rapporto ai quattro indici di posizione internazionale (*benchmark*) che corrispondono ai quattro punti della scala complessiva di matematica. Gli indici sono: livello Avanzato, livello Alto, livello Intermedio, livello Basso. Il riferimento ai quattro livelli permette di avere una visione non solo quantitativa ma anche qualitativa delle prestazioni degli studenti, in quanto ciascuno di essi corrisponde alla capacitá di compiere determinate operazioni cognitive sui contenuti matematici proposti, di complessitá via via piú elevata man mano che si passa dal livello basso a quello avanzato. Gli studenti che raggiungono un dato livello dovrebbero essere piú in grado di padroneggiare tutti i processi di pensiero tipici di quel livello rispetto a quanti si trovano sui livelli piú bassi della scala gerarchica. Inoltre il livello piú alto si caratterizzerá per una compresenza dei domini cognitivi (Conoscenza, Applicazione, Ragionamento) che saranno via via meno persistenti con il decrescere del livello stesso.
Di seguito sono riportate le descrizioni:

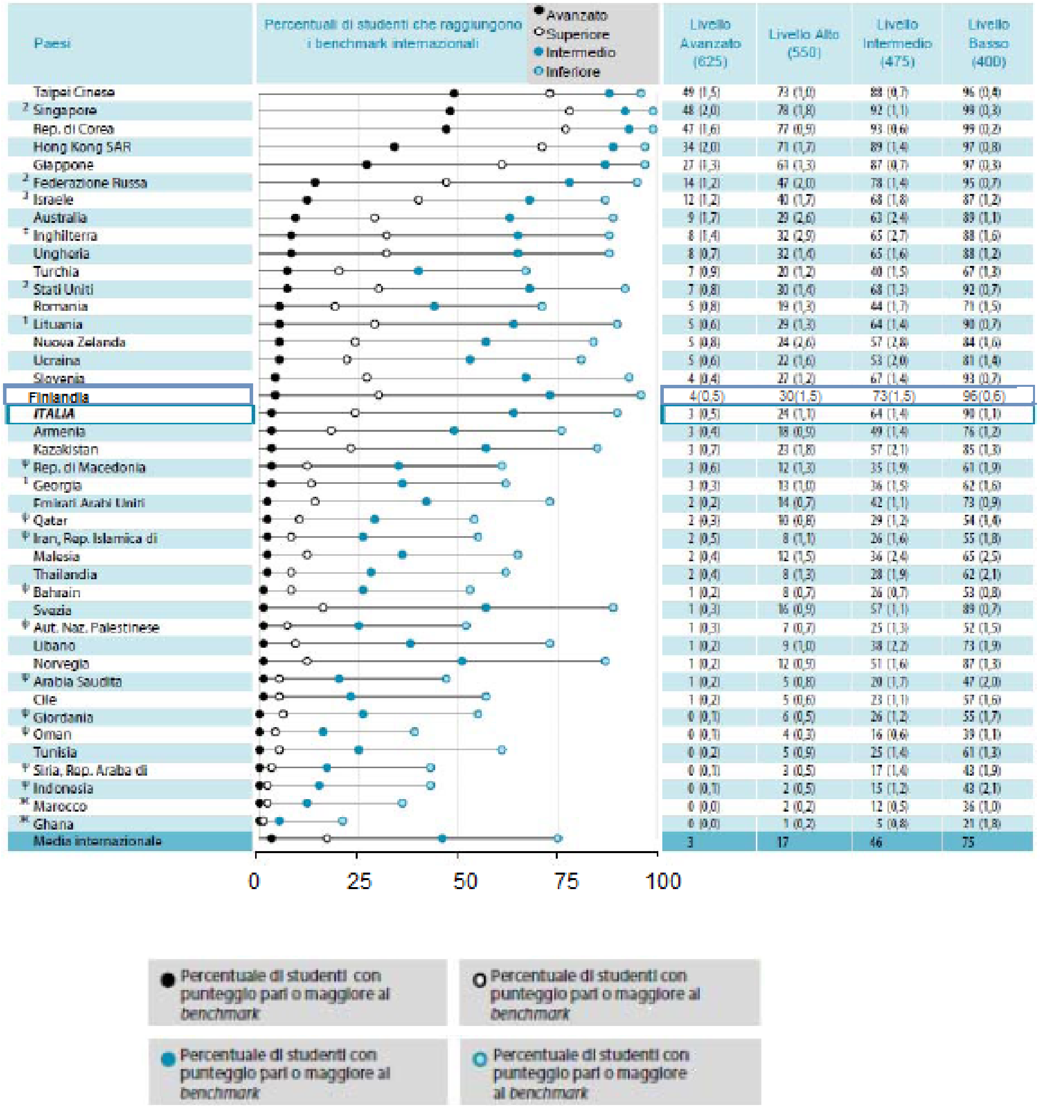
- **livello avanzato**: gli studenti sono in grado di argomentare se hanno informazioni a disposizione, di trarre conclusioni, di fare generalizzazioni e di risolvere equazioni lineari

- **livello alto**: gli studenti sono in grado di applicare conoscenze e concetti in una varietá di situazioni relativamente complesse

- **livello intermedio**: gli studenti sono in grado di applicare conoscenze matematiche di base in una molteplicitá di situazioni

- **livello basso**: gli studenti hanno alcune conoscenze relative ai numeri decimali, alle operazioni e ai grafici di base

Nella figura che segue sono riportate per ogni Paese le percentuali di studenti che raggiungono ciascuno dei livelli internazionali. I risultati sono riportati in ordine decrescente rispetto alla percentuale di studenti che raggiunge il livello piú alto. La percentuale di studenti che raggiunge il livello avanzato è indicata con un pallino nero. Poiché gli studenti che raggiungono i livelli piú alti ovviamente raggiungono anche i livelli piú bassi, le percentuali fornite nelle colonne di destra sono cumulate. Ad esempio se in Italia la percentuale di studenti a livello Intermedio è del 64% ció vuol dire che il 64% degli studenti italiani dell'ottavo anno raggiunge almeno il livello Intermedio.



Paesi	Livello Avanzato (625)	Livello Alto (550)	Livello Intermedio (475)	Livello Basso (400)
Taipei Cinese	49 (1,5)	73 (1,0)	88 (0,7)	96 (0,4)
[2] Singapore	48 (2,0)	78 (1,8)	92 (1,1)	99 (0,3)
Rep. di Corea	47 (1,6)	77 (0,9)	93 (0,6)	99 (0,2)
Hong Kong SAR	34 (2,0)	71 (1,7)	89 (1,4)	97 (0,8)
Giappone	27 (1,3)	61 (1,3)	87 (0,7)	97 (0,3)
[2] Federazione Russa	14 (1,2)	47 (2,0)	78 (1,4)	95 (0,7)
[3] Israele	12 (1,2)	40 (1,7)	68 (1,8)	87 (1,2)
Australia	9 (1,7)	29 (2,6)	63 (2,4)	89 (1,1)
[†] Inghilterra	8 (1,4)	32 (2,9)	65 (2,7)	88 (1,6)
Ungheria	8 (0,7)	32 (1,4)	65 (1,6)	88 (1,2)
Turchia	7 (0,9)	20 (1,2)	40 (1,5)	67 (1,3)
[3] Stati Uniti	7 (0,8)	30 (1,4)	68 (1,3)	92 (0,7)
Romania	5 (0,8)	19 (1,3)	44 (1,7)	71 (1,5)
[1] Lituania	5 (0,6)	29 (1,3)	64 (1,4)	90 (0,7)
Nuova Zelanda	5 (0,8)	24 (2,6)	57 (2,8)	84 (1,6)
Ucraina	5 (0,6)	22 (1,6)	53 (2,0)	81 (1,4)
Slovenia	4 (0,4)	27 (1,2)	67 (1,4)	93 (0,7)
Finlandia	4(0,5)	30(1,5)	73(1,5)	96(0,6)
ITALIA	3 (0,5)	24 (1,1)	64 (1,4)	90 (1,1)
Armenia	3 (0,4)	18 (0,9)	49 (1,4)	76 (1,2)
Kazakistan	3 (0,7)	23 (1,8)	57 (2,1)	85 (1,3)
[Ψ] Rep. di Macedonia	3 (0,6)	12 (1,3)	35 (1,9)	61 (1,9)
[1] Georgia	3 (0,3)	13 (1,0)	36 (1,5)	62 (1,6)
Emirati Arabi Uniti	2 (0,2)	14 (0,7)	42 (1,1)	73 (0,9)
[Ψ] Qatar	2 (0,3)	10 (0,8)	29 (1,2)	54 (1,4)
[Ψ] Iran, Rep. Islamica di	2 (0,5)	8 (1,1)	26 (1,6)	55 (1,8)
Malesia	2 (0,4)	12 (1,5)	36 (2,4)	65 (2,5)
Thailandia	2 (0,4)	8 (1,3)	28 (1,9)	62 (2,1)
[Ψ] Bahrain	1 (0,2)	8 (0,7)	26 (0,7)	53 (0,8)
Svezia	1 (0,3)	16 (0,9)	57 (1,1)	89 (0,7)
[Ψ] Aut. Naz. Palestinese	1 (0,3)	7 (0,7)	25 (1,3)	52 (1,5)
Libano	1 (0,2)	9 (1,0)	38 (2,2)	73 (1,9)
Norvegia	1 (0,2)	12 (0,9)	51 (1,6)	87 (1,3)
[Ψ] Arabia Saudita	1 (0,2)	5 (0,8)	20 (1,7)	47 (2,0)
Cile	1 (0,2)	5 (0,6)	23 (1,1)	57 (1,6)
[Ψ] Giordania	0 (0,1)	6 (0,5)	26 (1,2)	55 (1,7)
[Ψ] Oman	0 (0,1)	4 (0,3)	16 (0,6)	39 (1,1)
Tunisia	0 (0,2)	5 (0,9)	25 (1,4)	61 (1,3)
[Ψ] Siria, Rep. Araba di	0 (0,1)	3 (0,5)	17 (1,4)	43 (1,9)
[Ψ] Indonesia	0 (0,1)	2 (0,5)	15 (1,2)	43 (2,1)
[※] Marocco	0 (0,0)	2 (0,2)	12 (0,5)	36 (1,0)
[※] Ghana	0 (0,0)	1 (0,2)	5 (0,8)	21 (1,8)
Media internazionale	3	17	46	75

Figure 2.6: Percentuale di studenti a ciascun livello nella scala internazionale

La distribuzione dei risultati in Italia e in Finlandia è molto simile, infatti soltanto il 3-4% degli studenti raggiunge il livello piú alto (livello **avanzato**), precisamente il 3% in Italia e il 4% in Finlandia, ma la quasi totalitá di essi o almeno il 90% raggiunge il livello piú basso. Mentre per quanto riguarda il livello **alto**, la Finlandia si assesta intorno al 30% contro il 24% dell'Italia. Vengono di seguito forniti alcuni quesiti esplicativi (rivolti a studenti all'ottavo anno di scolaritá) che permetteranno di comprendere meglio la descrizione di ciascun livello e di rilevare, coerentemente a questo, la posizione dell'Italia.

I quesiti di livello **basso** dimostrano che gli studenti hanno una comprensione elementare dei numeri interi e decimali e sono in grado di eseguire calcoli di base. Possono inoltre collegare tabelle a grafici a barre o a ideogrammi e leggere un grafico con una sola retta. Ad esempio il quesito nel riquadro sotto, richiede di eseguire una somma di due numeri decimali. A livello internazionale, il 72% degli studenti ha risposto in maniera corretta, per l'Italia invece l'88%, differenza statisticamente significativa rispetto a quella internazionale.

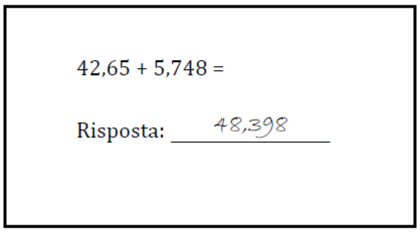


Figure 2.7: **Dominio di contenuto: Numero**
Dominio cognitivo: Conosenza

Gli studenti di livello **intermedio** sono in grado di risolvere problemi che riguardano numeri decimali, frazioni, proporzioni e percentuali in diversi contesti. Essi comprendono semplici relazioni algebriche, sono capaci di individuare e interpretare dati presentati sotto forma di tabelle, grafici... L'esempio nel riquadro riporta un quesito del dominio geormetria. Uno degli argomenti di questo dominio è proprio la relazione tra le forme tridimensionali e quelle bidimensionali e infatti questo richiede agli studenti di riconoscere una piramide dal suo sviluppo in piano e poi disegnarla come si vede vista dall'alto. In media le percentuali di risposte corrette a livello internazionale sono del 58% invece per l'Italia del 70%

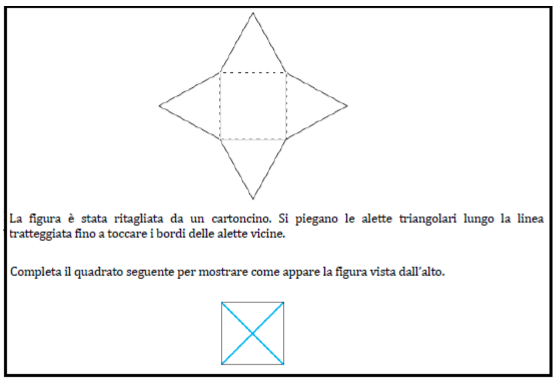


Figure 2.8: **Dominio di contenuto: Geometria**
Dominio cognitivo: Conosenza

Gli studenti di livello **alto** sono capaci di risolvere problemi con frazioni, proporzioni e percentuali e comprendere il senso di diverse espressioni e formule operando opportune manipolazioni. Possono anche identificare espressioni algebriche che corrispondono a semplici situazioni. Per quanto riguarda la geometria sono in grado di riconoscere e utilizzare proprietá. Un quesito esemplificativo di questo livello e appartenente al dominio algebra è quello riportato nel riquadro. Allo studente è richiesto di identificare la quantitá che soddisfi le due disuguaglianze rappresentate con due bilance all'interno di una situazione problematica. I risultati dimostrano che in diversi Paesi gli studenti mostrano una certa familiaritá con l'algebra: 11 Paesi hanno ottenuto percentuali superiori al 60%, l'Italia ha riportato una percentuale pari al 51%.

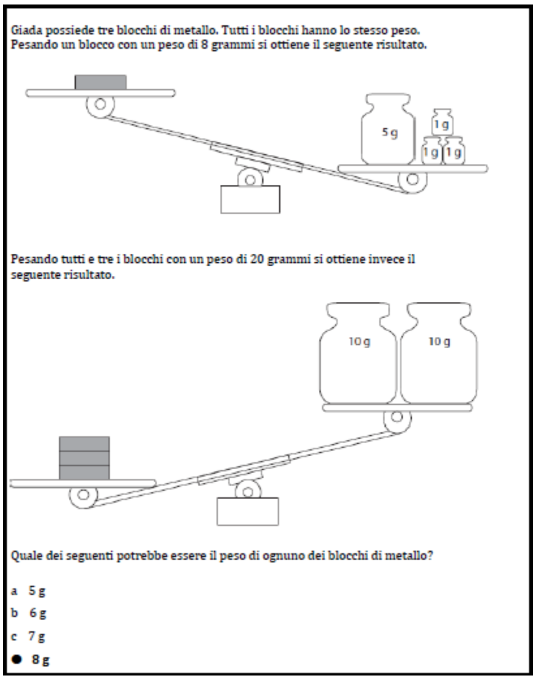


Figure 2.9: **Dominio di contenuto: Algebra**
Dominio cognitivo: Ragionamento

Gli studenti che raggiungono il livello **avanzato** sono in grado di interagire con diversi tipi di numeri (interi, negativi, frazionari, percentuali) anche in situazioni non di routine e di giustificare le loro conclusioni. Sono in grado di generalizzare utilizzando l'algebra o le parole. Sono capaci di risolvere diversi problemi che richiedono l'uso di equazioni, formule e funzioni. In geometria sono in grado di ragionare su figure geometriche per risolvere problemi che richiedono l'uso di proprietá o nozioni come il Teorema di Pitagora. Per dati e probabilitá gli studenti di questo livello dimostrano di comprendere il significato di media e sono in grado di estrapolare dati da un grafico. Nell'esempio che segue si chiede di ragionare sulle frazioni in una situazione non di routine: dati due punti P e Q, sulla retta dei numeri, che identificano due frazioni non specificate; lo studente deve identificare la posizione del prodotto.

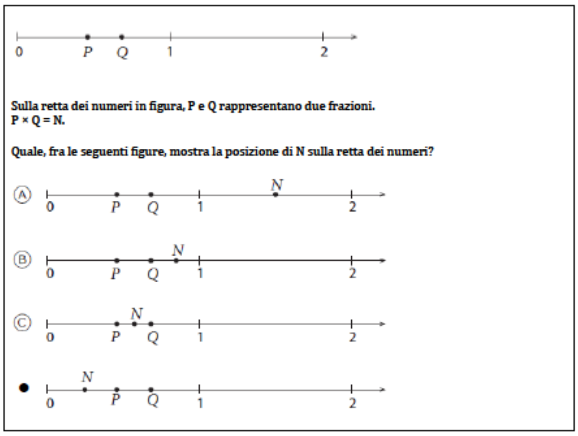


Figure 2.10: **Dominio di contenuto: Numero**
Dominio cognitivo: Ragionamento

Sebbene il formato della domanda sia a scelta multipla, solo il 23% degli studenti, a livello internazionale è stato in grado di rispondere correttamente. L'Italia ha avuto una delle percentuali piú basse di risposte corrette (16%). Un dato interessante riguarda la percentuale di studenti italiani che ha scelto l'opzione A come risposta corretta (44%) dato che conferma la presenza della misconcezione secondo la quale qualsiasi moltiplicazione fornisce un risultato maggiore dei fattori.

In sintesi, mentre i Paesi OCSE soprattutto grazie al contributo di quelli dell'Estremo Oriente, conseguono nel 2011 risultati significativamente piú elevati rispetto alla media (riferita a 500 punti), i Paesi UE complessivamente considerati, ottengono risultati significativamente piú bassi. I risultati italiani, meritano comunque un'attenzione particolare poiché, sebbene non si discostino significativamente dalla media, segnano comunque un'incremento rispetto all'edizione precedente di TIMSS (2007).
Al di lá dei numeri noi siamo interessati a scoprire quali siano i punti di forza o di debolezza al di sotto dell'esito complessivo e rispetto ad un determinato dominio di contenuto o cognitivo. Le figure che seguono mettono in luce se e in quale misura si manifestano.

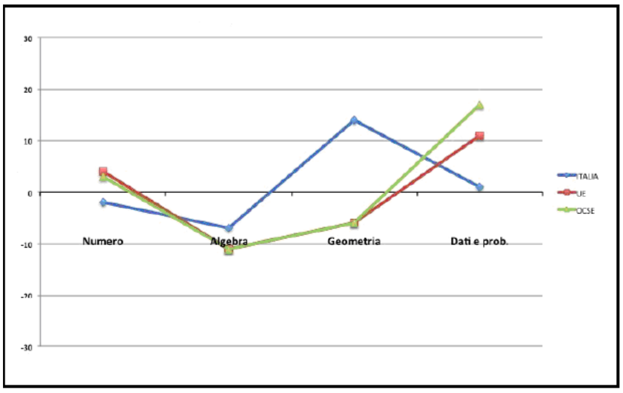


Figure 2.11: **Differenza dai punteggi medi di TIMSS 2011 rispetto ai domini di contenuto**

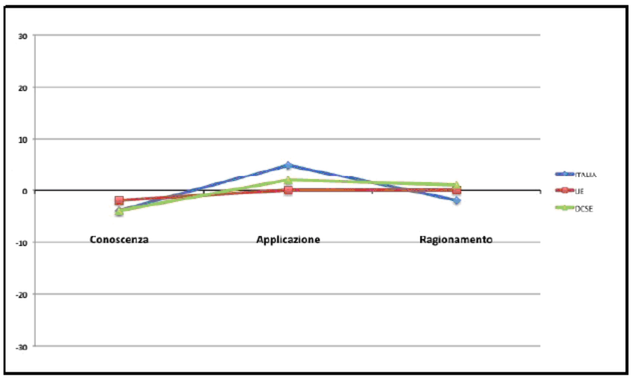


Figure 2.12: **Differenza dai punteggi medi di TIMSS 2011 rispetto ai domini cognitivi**

Le figure permettono di individuare alcune regolaritá dei risultati italiani rispetto a quelli dell'OCSE e della UE e anche alcune differenze. In particolare sembra emergere una certa debolezza degli studenti italiani nel processo cognitivo ragionamento.

Che dire di PISA? Si osserva che alcuni Paesi ottengono un punteggio piuttosto similare rispetto alle due indagini, nel senso che ottengono un punteggio superiore alla media sia in TIMSS che in PISA (come la Finlandia) o un punteggio inferiore alla media in entrambe le indagini (come l'Italia), altri Paesi invece punteggi piuttosto contrastanti.

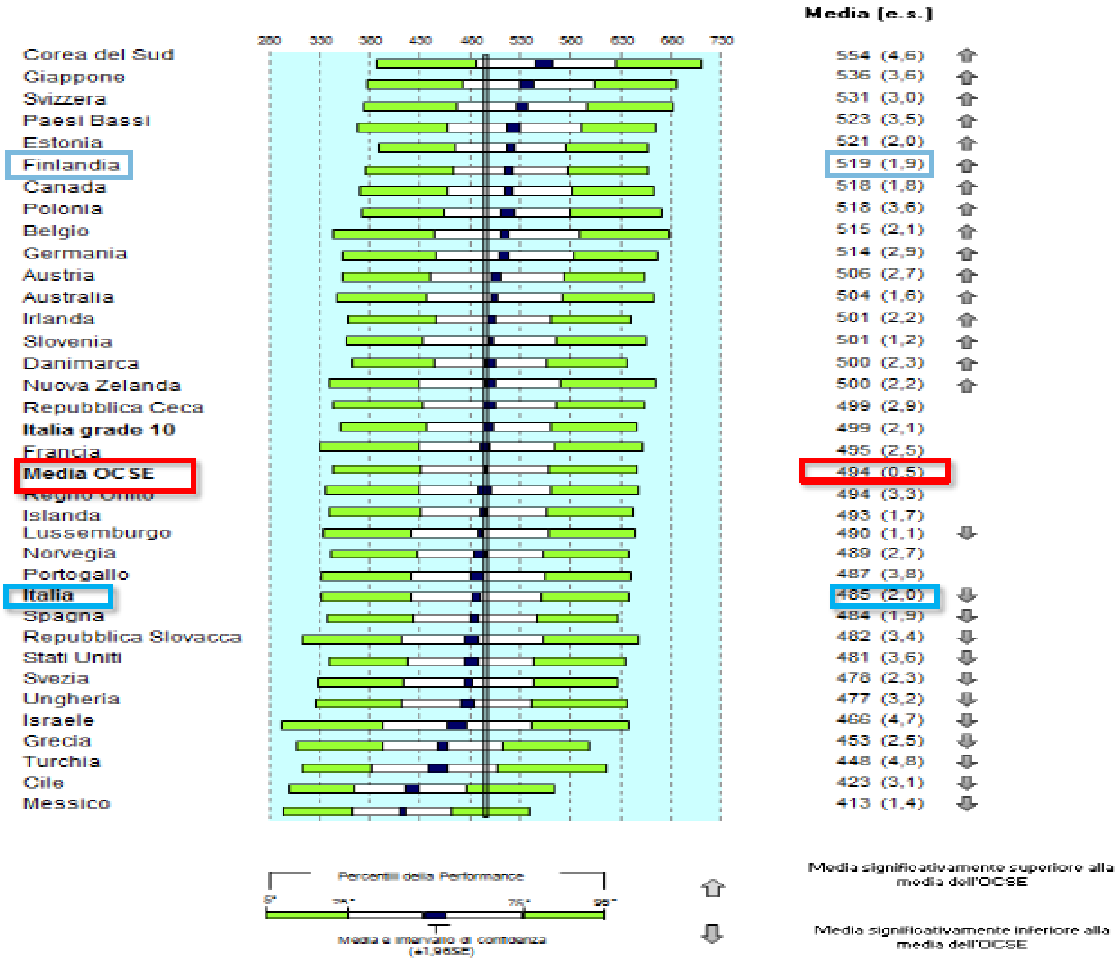


Figure 2.13: Distribuzione della perfomance in matematica nei paesi OCSE-PISA 2012

La media fornisce un'indicazione del livello di competenza complessivo di ciascun Paese partecipante all'indagine ma non indica la distribuzione delle prestazioni tra gli studenti. Per inquadrare questo aspetto, PISA ha definito 6 *livelli* di competenza che consentono di descrivere le capacitá degli studenti in ciascun ambito sottoposto al test.

Livello	
6	Al livello 6 gli studenti sono in grado di concettualizzare, generalizzare e utilizzare informazioni sulla base delle loro proprie ricerche e della modellizzazione di situazioni problematiche complesse. Sanno mettere in relazione diverse fonti di informazione e rappresentazioni muovendosi al loro interno senza difficoltà. Possono fare riflessioni e ragionamenti matematici difficili. Sanno applicare la loro comprensione e intuizione con una padronanza delle operazioni e relazioni matematiche simboliche e formali che permette loro di elaborare approcci e strategie di attacco in presenza di situazioni mai incontrate prima. A questo livello, gli studenti sono in grado di formulare e comunicare con precisione le loro azioni e riflessioni in relazione ai loro risultati, interpretazioni, argomentazioni e alla loro pertinenza rispetto alle situazioni iniziali.
5	Al livello 5 gli studenti possono mettere a punto e utilizzare modelli per situazioni complesse, identificando i vincoli e specificando i presupposti. Sanno selezionare, raffrontare e valutare le strategie appropriate alla soluzione di problemi complessi in relazione a tali modelli. Sono in condizione di affrontare situazioni da un'angolazione strategica, ricorrendo a competenze ben sviluppate di ragionamento e riflessione, alle relative rappresentazioni, caratterizzazioni simboliche e formali con una comprensione approfondita di tali situazioni. Riflettono sulle loro azioni e sanno formulare e comunicare le loro interpretazioni e il percorso seguito nel ragionamento.
4	Al livello 4 gli studenti sanno lavorare efficacemente con modelli espliciti di complesse situazioni concrete che possono comportare vincoli e richiedere deduzioni. Possono selezionare e integrare rappresentazioni di tipo diverso, anche simboliche, collegandole direttamente a situazioni del mondo reale. A questo livello, in questi contesti, gli studenti sono in grado di utilizzare competenze ben sviluppate e flessibilità di ragionamento, con un certo grado di profondità di comprensione. Possono elaborare ed esprimere argomentazioni e spiegazioni delle loro interpretazioni e azioni.
3	Al livello 3 gli studenti sono in grado di seguire procedure chiaramente descritte, comprese procedure che richiedono decisioni in sequenza. Sanno selezionare e applicare semplici strategie per risolvere problemi. A questo livello possono interpretare e utilizzare rappresentazioni a partire da fonti di informazione diverse e costruire un ragionamento su questa base. Sono in grado di elaborare comunicazioni succinte per presentare i loro risultati, interpretazioni e ragionamenti.
2	Al livello 2 gli studenti sono in grado di interpretare e riconoscere situazioni in contesti che richiedono al massimo deduzioni dirette. Possono ricavare informazioni pertinenti da un'unica fonte e utilizzare una sola modalità di rappresentazione. Sono in grado di utilizzare algoritmi, formule, procedure o convenzioni elementari. Possono fare un ragionamento diretto e fornire un'interpretazione letterale dei loro risultati.
1	**Al livello 1 gli studenti sanno rispondere a domande inserite in un contesto familiare dove tutte le informazioni sono esplicite e ciò che si chiede loro è chiaramente definito. Sono in grado di identificare informazioni ed eseguire procedure di routine seguendo istruzioni dirette in situazioni esplicite. Possono eseguire azioni ovvie derivanti direttamente dallo stimolo somministrato.**

Figure 2.14: Descrizione dei livelli di competenza matematica (PISA 2012)

I quesiti che si trovano ai livelli piú alti della scala delle competenze richiedono, da parte dello studente, un certo grado di riflessione, pensiero e creativitá. Di solito le situazioni descritte non fanno riferimento a situazioni familiari e necessitano quindi di piú alti livelli d'interpretazione. A questi alti livelli di competenza le domande tendono ad avere piú elementi che devono essere collegati dagli studenti e la soluzione in genere richiede un approccio strategico attraverso diversi passaggi interconnessi.
Al livello intermedio della scala di competenza, i quesiti richiedono un'interpretazione sostanziale, gli studenti sono tenuti a utilizzare rappresentazioni diverse della stessa situazione, comprese anche le rappresentazioni matema-

tiche piú formali. Ció comporta una catena di ragionamento o una sequenza di calcoli. Attivitá tipiche, a questo livello, includono: l'interpretazione di grafici, l'interpretazione del testo, sulla base di informazioni ricavabili in una tabella o in un grafico. Nella parte inferiore della scala delle competenze i quesiti vengono posti in modo semplice e fannno riferimento a contesti familiari. Viene richiesta solo l'interpretazione piú semplice della situazione e l'applicazione diretta di concetti matematici ben noti.
Vediamo come si distribuiscono gli studenti nei diversi livelli rispetto ai vari Paesi partecipanti.

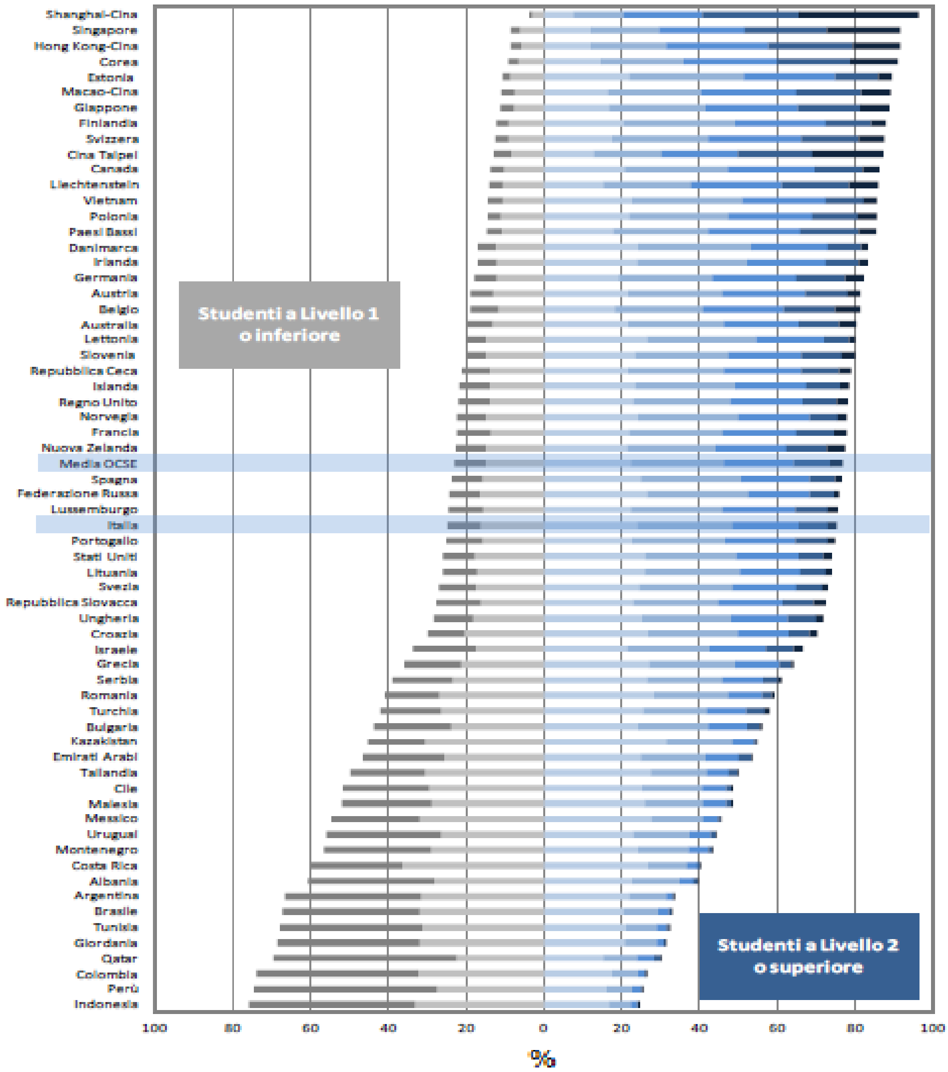


Figure 2.15: Percentuale di studenti a ciascun livello della scala di literacy matematica

Restringiamo il focus su alcuni Paesi tra i quali Italia e Finlandia e vediamo quali sono le percentuali nei vari livelli di competenza

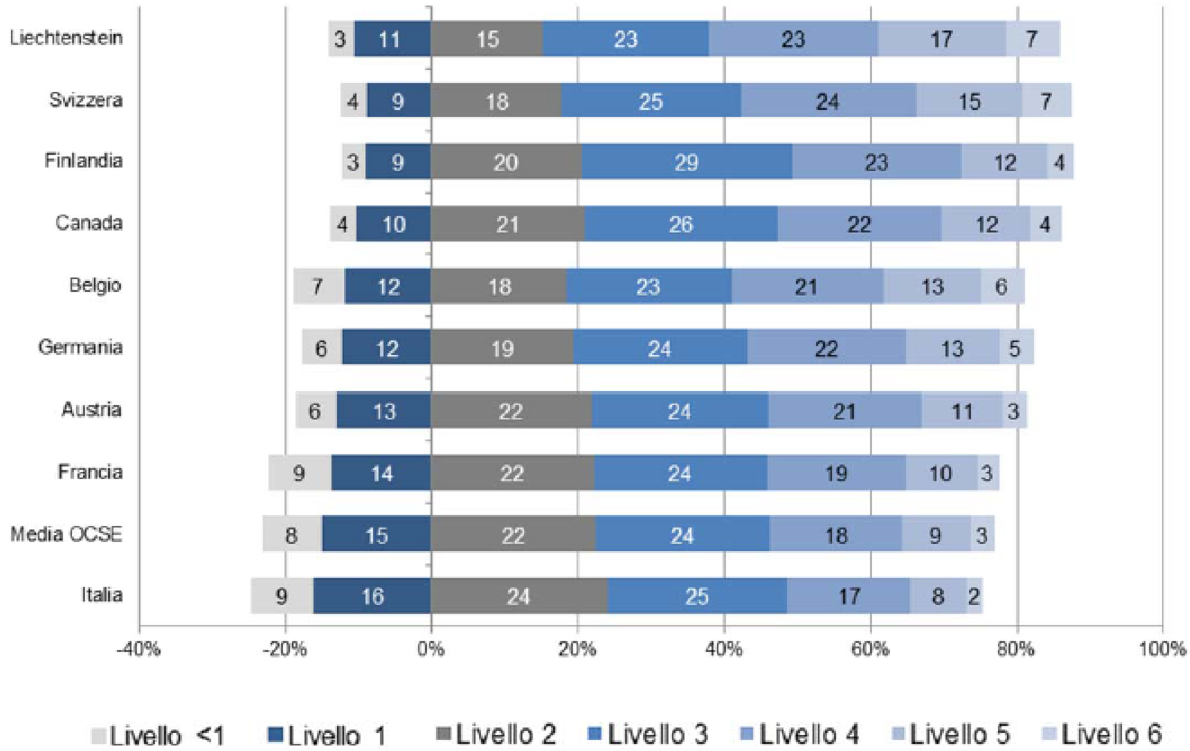


Figure 2.16: Prestazioni in matematica, secondo i livelli di competenza PISA 2012

Gli ideatori dell'indagine stimano che il livello 2 corrisponda al livello minimo di competenza per partecipare efficacemente alla vita quotidiana.
Al contrario, si considera che gli studenti che raggiungono i livelli 5 o 6 siano molto competenti. Si rileva quindi che, mediamente, nei paesi OCSE il 23% di essi si colloca al di sotto del livello 2. All'estremitá opposta, il 13% rientra nei livelli 5 o 6. In particolare il 3,3% degli studenti raggiunge il livello 6. In generale è possibile riscontrare che laddove la prestazione di un paese è migliore, la percentuale di allievi che ottiene risultati inferiori al livello 2 tende a diminuire. In Finlandia si riscontra una percentuale di studenti molto deboli (livello ≤ 1) pari al 12% contro il 25% dell'Italia, quindi percentualmente parlando, gli studenti italiani che non hanno le competenze minime sono circa il doppio rispetto agli studenti finlandesi. Mentre la percentuale di studenti molto competenti (livello 5 o 6) in Finlandia è pari al 16% (con una ripartizione approssimativa del 12% nel livello 5 e del 4% nel livello 6) contro il 10% dell'Italia (con una ripartizione approssimativa

dell'7.8% nel livello 5 e del 2.2% nel livello 6).
In particolare si osservano le seguenti differenze, in termini di prestazione, tra le due indagini TIMSS e PISA relativamente a Italia e Finlandia (le stime PISA non sono cumulate):

	avanzato	alto	intermedio	basso
TIMSS $\rightarrow$ Italia	3%	24%	64%	90%
TIMSS $\rightarrow$ Finlandia	4%	30%	73%	96%

	livello 6	livello 5	livello 4	livello 3	livello 2	livello 1
PISA $\rightarrow$ Italia	2.2%	7.8%	16.7%	24.6%	24%	16%
PISA $\rightarrow$ Finlandia	4%	12%	23%	29%	20%	9%

Di seguito si osservano la distribuzione delle prestazioni migliori degli studenti Italiani negli aspetti di competenza: **Quantitá** e nell'aspetto cognitivo: **Interpretazione**.

		Differenza di media tra la competenza in matematica e ciascun aspetto di competenza di contenuto				Differenza di media tra la competenza in matematica e ciascun aspetto di competenza di processo		
	Media in matematica	*Variazioni e relazioni*	*Spazio e forme*	*Quantità*	*Incertezza e dati*	*Formulare*	*Applicare*	*Interpretare*
Liechtenstein	535	7	4	3	–9	0	1	5
Svizzera	531	–1	13	0	–9	7	–2	–2
Finlandia	519	2	–12	8	0	0	–3	9
Canada	518	7	–8	–3	–2	–2	–2	3
Belgio	515	–1	–6	4	–7	–2	1	–2
Germania	514	2	–6	4	–5	–3	2	3
Austria	506	1	–5	5	–7	–6	4	3
Francia	495	2	–6	1	–3	–12	1	16
Media OCSE	494	–1	–4	1	–1	–2	–1	3
Italia	485	–9	2	5	–3	–10	0	13

La media del paese nell'aspetto di competenza è tra 0 e 3 punti **al di sopra** della media della competenza in matematica
La media del paese nell'aspetto di competenza è tra 3 e 10 punti **al di sopra** della media della competenza in matematica
La media del paese nell'aspetto di competenza è più di 10 punti **al di sopra** della media della competenza in matematica

La media del paese nell'aspetto di competenza è tra 0 e 3 punti **al di sotto** della media della competenza in matematica
La media del paese nell'aspetto di competenza è tra 3 e 10 punti **al di sotto** della media della competenza in matematica
La media del paese nell'aspetto di competenza è più di 10 punti **al di sotto** della media della competenza in matematica

Figure 2.17: Risultati negli aspetti di competenza in matematica di alcuni Paesi partecipanti in PISA 2012

La Finlandia supera la media italiana relativamente alla competenza di contenuto (con 3 punti percentuale in piú) ma non rispetto alla competenza di processo posizionandosi al di sotto di ben 4 punti percentuale. Mentre la prestazione italiana peggiore, rispetto alla competenza di processo, riguarda la **Formulazione**, con una percentuale sotto la media di 10 punti contro una prestazione nella media degli studenti finlandesi.

A questo punto è naturale chiedersi se, dietro questa differenza in percentuale (significativa per quanto riguarda l'indagine PISA), tra Italia e Finlandia si nasconda una reale superioritá nella competenza matematica degli studenti finlandesi e se sí, a cosa puó essere ricondotta, o come si spiegano i diversi risultati (per l'Italia), rispetto alle due rilevazioni TIMSS e PISA, visto che ció che si misura, la competenza matematica, è ambito comune alle due indagini.
Prima di rispondere alle due domande vorremmo passare in rassegna alcune prove, somministrate da PISA, volte a testare il processo cognitivo **Formulazione**. Innanzitutto, secondo PISA la Formulazione si riferisce: alla capacitá degli studenti di riconoscere e individuare le opportunitá di usare la matematica e di fornire quindi una struttura matematica a un problema presentato in forma contestualizzata. Nel processo di formulazione delle situazioni in forma matematica, gli studenti determinano i punti da cui *estrarre* gli elementi matematici necessari per *analizzare*, *impostare* e *risolvere* il problema. Eseguono un processo di *traslazione* da un contesto reale a un ambito matematico e conferiscono al problema una *struttura*, una *rappresentazione* e una *specificitá* di tipo matematico. *Ragionano* e *interpretano* le limitazioni e le ipotesi poste dal problema.
La prova "Pizze" si basa in larga misura sulle capacitá degli studenti di formulare una situazione matematica. Gli studenti sono chiamati ad eseguire calcoli per risolvere il problema e a interpretare i risultati per comprendere quale pizza presenti il vantaggio economico maggiore, ma la vera sfida cognitiva di questa prova risiede nella capacitá di formulare un modello matematico che racchiuda il concetto di prezzo piú conveniente.

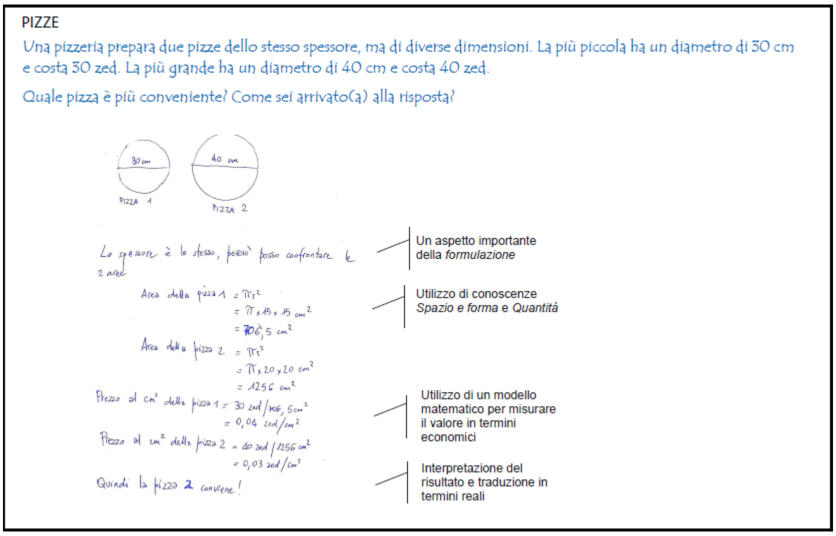


Figure 2.18: Item della prova PIZZE ed esempio di risposta ad un quesito

La prova “Concerto rock” costituisce un ulteriore esempio di prova basata in larga misura sulle capacitá degli studenti di formulare una situazione in forma matematica, infatti prevede che interpretino le informazioni contestuali fornite (dimensione e forma del campo, il fatto che il concerto rock è tutto esaurito e il fatto che i fan stanno in piedi) e le traducano in una forma matematica utile ai fini di una stima del numero di persone presenti al concerto. Inoltre l’aspetto ragionamento e argomentazione entra in gioco con la necessitá di riflettere sulla relazione tra il modello elaborato, la soluzione ottenuta e il contesto reale, prima di validare il modello e dare la risposta corretta.

CONCERTO ROCK

Per un concerto rock, è stato riservato al pubblico un campo rettangolare da 100m x 50m. Il concerto è andato tutto esaurito e il campo è pieno di fan, che stanno in piedi.

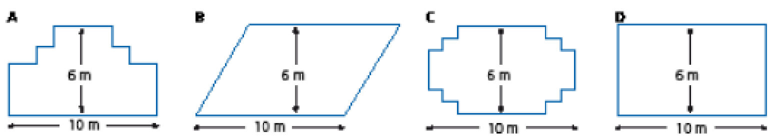
Quale di questi numeri è la stima più probabile del numero di persone che hanno assistito al concerto?

A. 2 000
B. 5 000
C. 20 000
D. 50 000
E. 100 000

Figure 2.19: Item Concerto rock

Questo item richiede di selezionare la risposta (multipla semplice) e considerato che il 28% degli studenti ha risposto correttamente (C) l’item va considerato moderatamente difficile tra quelli utilizzati nel test.
Un ultimo esempio che consente di visualizzare la tipologia degli item focalizzati sul ragionamento (è tra gli item piú difficili con una percentuale di risposte corrette (2003) di poco inferiore al 20%) e osservare la differenza sostanziale del carico cognitivo richiesto dagli omologhi di TIMSS

Un carpentiere ha 32 metri di tavole di legno e vuole fare il recinto a un giardino. Per il recinto prende in considerazione i seguenti progetti.



Indica per ciascun progetto se è possibile realizzarlo con 32 metri di tavole. Fai un cerchio intorno a “Sì” o a “No”.

Progetto	Utilizzando questo progetto,si può realizzare il recinto con 32 metri di tavole?
Disegno A	Sì / No
Disegno B	Sì / No
Disegno C	Sì / No
Disegno D	Sì / No

Figure 2.20: Item esemplificativo carpentiere

Dopo aver messo a confronto alcuni esempi di quesiti relativi alle due indagini e alla luce delle rispettive finalitá e intenti operativi e valutativi, appaioni piuttosto evidenti delle differenze "strutturali" tra TIMSS e PISA, ovvero nell'ideazione e architettura del framework e nella tipologia degli item. Mentre il focus di TIMSS è in larga misura sulle attivitá della scuola, in termini di insegnamenti e di apprendimenti e di conseguenza la struttura dei quesiti è di tipo curricolare, l'attenzione di PISA è principalmente rivolta sulle competenze degli studenti, attitudini (verso l'oggetto dell'apprendimento) strategie e motivazione sottese agli apprendimenti, quindi le prove sono concepite per far emergere le capacitá matematiche fondamentali. TIMSS si focalizza su quegli item che richiedono la riproduzione di fatti o algoritmi standard, mentre PISA principalmente su item che richiedono il saper fare collegamenti fra la conoscenza acquisita. Come mostrano le tabelle sottostanti, TIMSS ha un piú largo numero di item concernenti Algebra e Numero mentre gli item di PISA sono piú equamente distribuiti attraverso i vari domini di contenuto.

Domande della prova TIMSS	**Domande a scelta multipla**	**Domande a risposta aperta**	**Numero totale**	**Percentuale del punteggio**
Dominio di Contenuto				
Numero	31 (30)	30 (36)	61 (67)	29%
Algebra	37 (37)	26 (27)	70 (76)	33%
Geometria	25 (25)	18 (19)	43 (44)	19%
Dati e probabilità	25 (25)	18 (20)	43 (45)	19%
Totale	118 (118)	99 (114)	217 (232)	100%
Percentuale del punteggio	51%	49%		

Figure 2.21: Distribuzione del numero di quesiti TIMSS per domini di contenuto e tipologia di item

Categoria di contenuto	Punteggio percentuale
Cambiamento e relazioni	Approssimativamente 25
Spazio e forma	Approssimativamente 25
Quantità	Approssimativamente 25
Incertezza e dati	Approssimativamente 25
TOTALE	100

Figure 2.22: Distribuzione approssimativa del punteggio per categoria di contenuto nel ciclo PISA 2012

In altre parole, gli item di Pisa sono diretti alle "capacitá nella *vita*" quelli di TIMSS sono piú orientati alla *conoscenza*, TIMSS chiede: "che cosa sai fare con la matematica che hai imparato", PISA: "che cosa puoi fare...". Possiamo in tal caso parlare di differenze intrinseche alle due indagini.

Il grafico che segue consente di confrontare invece le prestazioni dei Paesi OCSE in matematica e l'equitá dei risultati dell'educazione (in Pisa equitá significa offrire agli studenti le stesse opportunitá di scolarizzazione indipendentemente dalla loro condizione sociale) misurata mediante l'indice socioeconomico. I Paesi che, come la Finlandia, si trovano nel quarto superiore destro del grafico presentano prestazioni superiori alla media OCSE (per quanto riguarda la prestazione in matematica) e un livello di equitá superiore (o inferiore) a tale media a seconda della loro posizione rispetto al livello medio di tale equitá (contrassegnato dalla linea verticale). Risulta inferiore alla media la relazione tra prestazione (matematica) e livello socioeconomico dell'Italia, tra i Paesi OCSE.

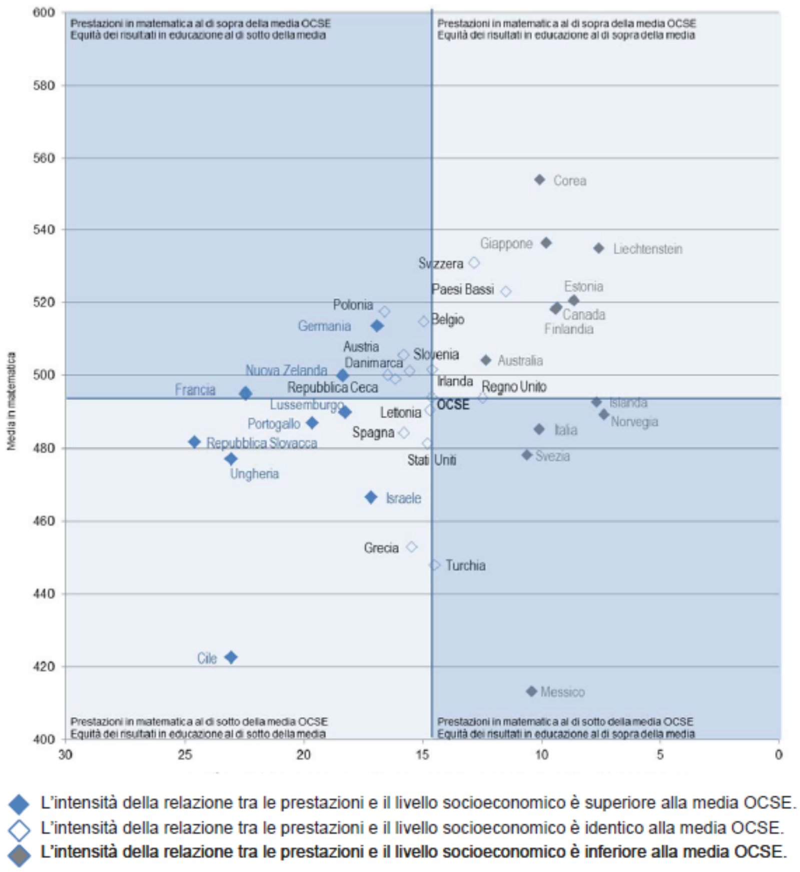


Figure 2.23: Prestazioni medie in matematica ed equitá, PISA 2012

Nota: l'asse orizzontale indica l'equitá dei risultati in educazione, misurati secondo l'indice socioeconomico, sociale e culturale di PISA; l'asse verticale indica le prestazioni in matematica sulla scala di PISA 2012; la linea orizzontale rappresenta la media dell'OCSE e la linea verticale rappresenta il livello medio di equitá dell'OCSE.

Questo ci consente di inferire delle osservazioni che consentono di trarre la seguente conclusione: sembra che l'indagine PISA possa essere meglio allineata con i sistemi educativi di quei Paesi piú altamente sviluppati, piú di quanto possa esserlo l'indagine TIMSS. Ovvero, mentre i Paesi altamente sviluppati (prevalentemente Paesi Occidentali) sottendono un sistema scolastico in grado di fornire competenze teoriche e operative al passo con la modernitá, in risposta ad un continuo bisogno di interpretazione e rielaborazione della mutevolezza della realtá circostante, i paesi meno sviluppati (fondamentalmente dell'Est) sono ancora concentrati su un piú formale e tradizionale approccio culturale, quindi piú consono ad essere testato su un piano piú strettamente curricolare. Possiamo in tal caso, parlare di differenze estrinseche alle due indagini. Di seguito un grafico (relativo al 2003) che mostra come l'allineamento PISA-TIMSS con la realtá socioculturale di alcuni Paesi partecipanti non sia sostanzialmente cambiato.

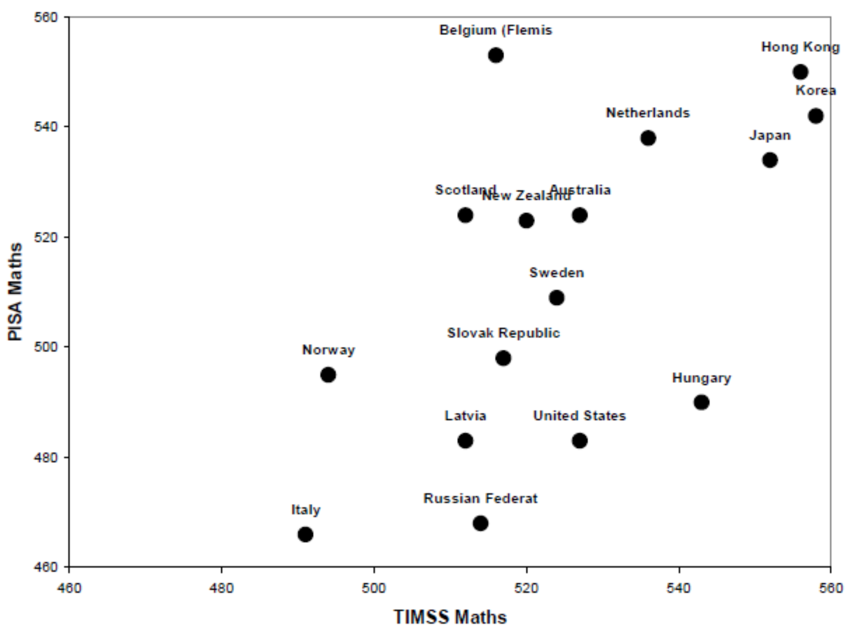


Figure 2.24: TIMSS vs PISA 2003

In conclusione, alla luce di tutto ció, risulta piuttosto evidente che le eventuali differenze di punteggio tra TIMSS e PISA non rappresentano di per sè differenze in performance assoluta.

2.5 Perché i Finlandesi sono cosí bravi in matematica

In questa sezione discuteremo il ruolo e l'impatto dell'indagine PISA in Finlandia. PISA ha creato una nuova geografia di politiche e riforme educative spostando l'interesse globale dal sistema educativo Anglo-Sassone ai Paesi Asiatici, Finlandia e Canada. Precedentemente al primo ciclo di PISA (2000) molti Paesi si facevano paladini della superioritá del loro sistema scolastico e del fatto che i loro studenti fossero tra i migliori al mondo. Tra questi Paesi possono essere annoverati: Germania, Francia, Norvegia, Svezia, Inghilterra e Stati Uniti. Molti altri, appartenenti al blocco dell'Est Socialista (come Yugoslavia, Bulgaria, Romania e Ungheria nonché Unione Sovietica) credevano di poter competere nel panorama internazionale. Indicatori di tale successo erano cospicui investimenti economici stanziati per l'istruzione o brillanti risultati in performance Internazionali come le Olimpiadi della Matematica e della Chimica. La tabella che segue mostra la posizione della Finlandia, quando per la prima volta vi ha preso parte.

	Medals			Number of participations	Number of participating students
	Gold	Silver	Bronze		
China	101	26	5	23	134
USA	80	96	29	34	216
The Soviet Union	77	67	45	29	204
Hungary	74	138	77	48	324
Romania	66	111	88	49	332
Russia	65	28	9	17	102
Bulgaria	50	89	88	49	336
Japan	23	52	30	19	114
Canada	16	37	66	28	168
Sweden	5	23	66	41	271
The Netherlands	2	21	48	38	250
Norway	2	10	24	25	142
Finland	1	5	47	35	224
Denmark	0	3	18	18	102

Figure 2.25: Posizione dei Paesi selezionati nelle Olimpiadi della Matematica dal 1959

Fino al 2001 era un pensiero comunemente condiviso che il livello della conoscenza scientifica e le competenze matematiche degli studenti Finlandesi fossero al piú modeste, rispetto agli standard internazionali. Infatti confrontando i risultati degli studenti Finlandesi, rispetto ai tre gradi di scolarizzazione (alla fine della scuola primaria, alla fine della scuola secondaria di primo grado e nel corso della scuola secondaria di secondo grado), a partire dalla prima edizione (1960) degli Studi Internazionali della Matematica fino al 1999 si osserva che le prestazioni in matematica e scienze erano piuttosto modeste.

	Population	Countries	Rank of Finland
IEA First International Mathematics Study (FIMS) 1962–67	13-year-olds and high school completion	12	Average performer
IEA First International Science Study (FISS) 1967–73	10 and 14-year-olds and high school completion	18	Average performer
IEA Study of Reading Comprehension 1967–73	10 and 14-year-olds and high school completion	14	Average performer (in one area third)
IEA Second International Mathematics Study (SIMS) 1977–81	13-year-olds and high school completion	19 (13-year-olds) 15 (high school)	Average performer
IEA Second International Science Study (SISS) 1980–87	At primary, middle and high school completion	23	10-year-olds high 14-year-olds Average performer
IEA Written Composition Study 1980–88	At primary, middle and high school completion	14	Average performer
IEA Reading Literacy Study 1988-94	9 and 14-year-olds	32	Top performer
IEA Third (later Trends in) International Mathematics and Science Study (TIMSS)	4th and 8th grade	1995: 45 1999: 38 2003: 50 2007: 59	Above average performer in 1999 (only participation)
IEA Progress in International Reading Literacy Study (PIRLS)	4th grade	2001: 35 2006: 45	Not participated
IEA International Civic and Citizenship Education Study (CIVED and ICCS)	8th grade	1999: 31 2009: 38	Top performer
OECD Programme for International Student Assessment (PISA)	15-year-olds	2000: 43 2003: 41 2006: 57 2009: 65	Top performer

Figure 2.26: Performance degli studenti Finlandesi nel sistema di valutazione internazionale dal 1960

Qualcosa di diverso invece si è registrato a partire dalla prima edizione di PISA (2000) rispetto alla quale, la Finlandia, si è classificata come una delle nazioni con migliore performance, tra i paese OCSE, mantenendo nel corso degli anni tale primato. Vale pertanto la pena di chiedersi: che cosa puó spiegare questo evidente progresso nell'apprendimento della matematica? Quali fattori stanno dietro al successo di una riforma scolastica?

In questa analisi appaiono tre possibili spiegazioni:

- l'insegnamento della matematica è fortemente integrato nel curriculum relativo alla formazione degli insegnanti della scuola primaria (ad esempio all'Universitá di Helsinki ogni anno circa il 15% di studenti orientati all'insegnamento nella scuola primaria, seguono corsi specializzanti per l'insegnamento della matematica)
- sia il curriculum formativo degli insegnanti stessi, che il curriculum realizzato (l'insegnamento della matematica a scuola) hanno un forte focus sul problem-solving e tenendo conto che, i quesiti di PISA sono fortemente basati sull'uso della matematica in situazioni nuove, ció pone gli studenti Finlandesi in vantaggio.
- la formazione di insegnanti di matematica è garantita da una stretta sinergia fra le facoltá di Matematica e di Scienze dell' Educazione

Il crescente utilizzo di PISA come indice di posizione internazionale del successo scolastico, alla fine della scuola dell'obbligo ha creato una situazione di continuo rinnovamento, orientando il sistema educativo Finlandese a focalizzare obiettivi formativi sempre piú concernenti con ció che gli studenti possono fare con quello che imparano a scuola. Ma il contributo piú importante a tale progresso deve essere ricondotto primariamente a politiche e riforme educative instaurate nel 1990 e quindi precedenti all'avvento di PISA. Il principio che ha ispirato la riforma educativa degli anni '90 è stato che il successo scolastico si sarebbe misurato con: quanto bene uno studente impara ció che si aspetta di imparare, in accordo con il fatto che l'apprendimento rappresenta una traccia (piú o meno fedele) dell'insegnamento e deve poter fornire, contestualmente al percorso, conoscenza e metaconoscenza ovvero criteri e strumenti di autovalutazione e autoregolazione. Cosí la formazione degli insegnanti di scuola primaria si è concentrata, negli scorsi due decenni, sul ridefinire l'insegnamento e l'apprendimento delle materie scientifiche in modo tale che gli studenti imparino attraverso la sperimentazione e la scienza *hands-on* e *minds-on*, secondo il principio di rendere tutti gli studenti attivi nell'insegnamento e nell'apprendimento. In una prospettiva nella quale la conoscenza disciplinare diventasse un oggetto culturale da usare per affrontare i problemi, il pensiero formale dovesse crescere con le idee e le ipotesi interpretative, attivandosi in modo funzionale ai bisogni con capacitá operative in diversi contesti. Gli studi universitari si sono incentrati sulla costruzione di contenuti pedagogici di conoscenza e comprensione di quei processi cognitivi che intervengono nella costruzione della conoscenza. Cosí il curriculum realizzato è stato trasformato da tradizionale e accademico ad un curriculum sperimentale e problem-oriented, superando l'abitudine consolidata ed auto-referente di insegnare sempre nello stesso modo e impostando la didattica sui contenuti con i metodi propri della scienza piuttosto che sui suoi prodotti.

Chapter 3

Metodi e Modelli per la costruzione e l'analisi del test

3.1 Capacitá matematiche e la loro relazione con la difficoltá degli item

Per giungere alla formulazione finale di un test, le indagini TIMSS e PISA usano una forma piuttosto sofisticata di studio e analisi basata sull'Item Response Theory (IRT). L'Item Response Theory è un'approccio di studio focalizzato nello specificare la relazione tra caratteristiche degli item e capacitá dei soggetti. Tale approccio definisce un modello statistico (una funzione) che consente di prevedere probabilisticamente la risposta all'item. Si compone di due fasi:

- calibrazione (stima delle proprietá psicosometriche degli item)
- scoring (stima delle abilitá degli individui)

La calibrazione è la fase della costruzione vera e propria del test, nella quale, è necessario valutare se gli item introdotti nel test (prelevati da una banca di item) e le risposte previste sono idonee a "misurare" le competenze (o capacitá latenti). La capacitá misuratoria di ogni quesito è analizzata mediante modelli statistici in grado di stabilire la coerenza di ciascuna opzione di risposta rispetto al costrutto oggetto di valutazione, rispetto al livello di abilitá del rispondente e rispetto alla difficoltá specifica del quesito stesso. I richiamati modelli statistici permettono, inoltre, di valutare il cosiddetto potere discriminante di ciascun quesito, ossia la capacitá di distinguere adeguatamente gli allievi in termini di abilitá, in funzione della risposta fornita.
Lo scoring rappresenta la fase della quantificazione della competenza (analisi qualitativa e quantitativa). La calibrazione e lo scoring possono essere considerate come le due facce della medaglia perché un buon metodo per

analizzare la difficoltá empirica degli item è considerare quali aspetti delle capacitá matematiche fondamentali siano necessarie per formulare e applicare una soluzione. Per stabilire a priori la difficoltá è necessario considerare sia il numero di capacitá sia la complessitá della loro attivazione.
La sezione sui modelli concettuali illustrerá le caratteristiche che rendono piú o meno complessa l'attivazione di una capacitá.
La fase del pre-test è complementare sia alla calibrazione che allo scoring, in quanto dopo aver somministrato un campione di prove (ad un campione di studenti che consenta di avere una buona tenuta statistica dei dati) si procede, una volta analizzati i dati (forniti dallo studio della funzione caratteristica del test), alla scelta di quegli item che daranno vita alla prova di profitto, facendo attenzione che tutti gli obiettivi didattici individuati in fase di progettazione siano adeguatamente rappresentati. Scopo del pre-test è ottenere-confermare indicatori psicosometrici di ogni singolo item e, conseguentemente, elaborare delle scale di misurazione, sulla base dei processi matematici e relative alle categorie di contenuto.
Per comprendere l'importanza del modello ai fini meramente esplorativi e interpretativi (delle performance) possiamo osservare che se sottoponessimo gli studenti ad un test di profitto su argomenti non ancora affrontati i risultati che si otterrebbero si distribuirebbero probabilmente in modo molto simile alla curva riportata in figura 4.

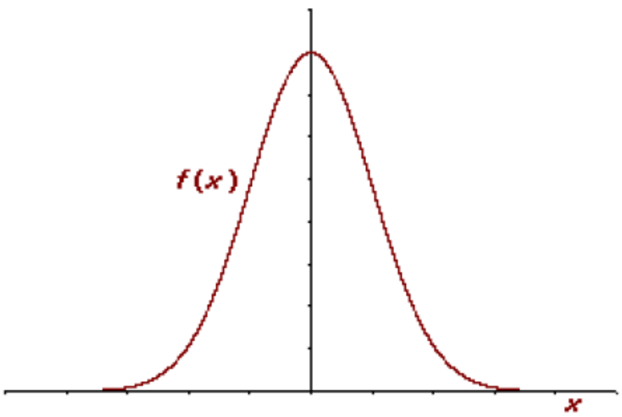


Figure 3.1: Curva gaussiana: distribuzione casuale delle performance in un test di profitto

Questo perché, in assenza di un apprendimento specifico degli argomenti proposti, gli studenti utilizzerebbero le proprie risorse intellettive per tentare di trovare delle soluzioni (e si sa che la capacitá intellettiva si distribuisce nella popolazione secondo una curva normale). La curva in questione ben descrive lo stato di partenza degli studenti ma non puó essere considerata un indice di apprendimento. Infatti lo scopo di un'attivitá di apprendimento non è mantenere invariata la situazione precedente l'intervento: in linea teorica, quando si insegna una serie di concetti ci si aspetta che tutti gli allievi apprendano in ugual misura quanto proposto. Questo obiettivo non si raggiunge perché esistono differenze individuali, ció nonostante è auspicabile poter rilevare variazioni positive nell'andamento della performance, una volta concluso un processo formativo e capire se sono state attivate delle abilitá. Pertanto se sottoponessimo agli studenti un test di profitto su argomenti insegnati, potremmo osservare, tramite la curva caratteristica del test, TCC (che normalmente è uno strumento che si usa in fase di costruzione del test e consente di ottenere informazioni circa la capacitá predditiva degli items che compongono il test stesso), una distribuzione delle prestazioni che non segue piú un andamento "normale" (come nella gaussiana vista prima) piuttosto una curva che giá a piccoli valori di abilitá assume valori positivi di prestazione, a dimostrazione del fatto che l'intervento didattico ha lasciato una traccia.

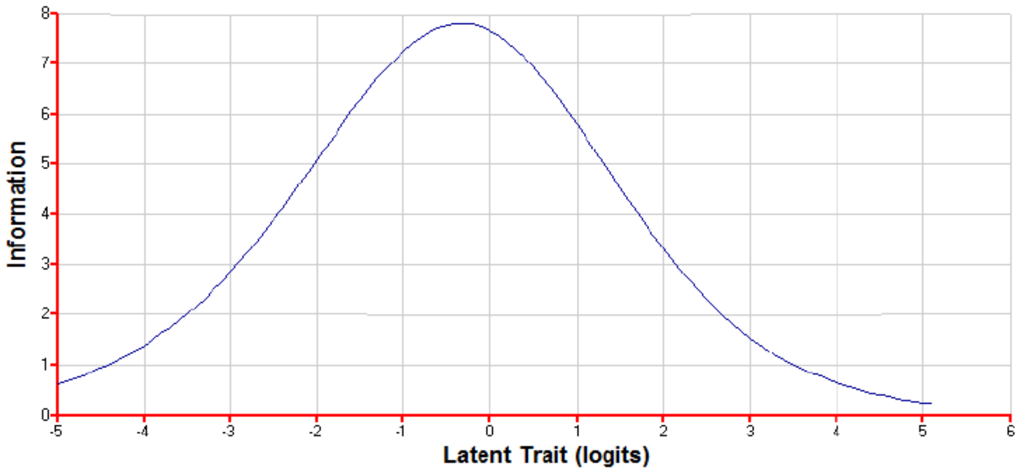


Figure 3.2: Informazione totale della prova di Matematica (Invalsi) III°secondaria, 1°grado

3.2 Caratteristiche della IRT e la scelta del modello

Tutti i modelli di risposta all'item sono basati sulle seguenti assunzioni:

1. unidimensionalitá: esiste una sola abilitá latente, sottostante il processo di risposta
2. indipendenza locale: dato un certo livello di abilitá, le risposte agli item sono tra loro indipendenti
3. universalitá dell'applicazione (comparabilitá dei risultati)

La prima di tali assunzioni richiede che gli item di un test misurino solamente un tratto latente per volta. E' evidente che, poiché un tratto latente è un aspetto della natura umana, il requisito di unidimensionalitá molto difficilmente puó essere soddisfatto; infatti esistono molteplici aspetti che possono influenzare la performance di un qualunque soggetto e che costituiscono, a loro volta, possibili tratti latenti oggetto di interesse, ad esempio: motivazione, esperienza, emotivitá ecc.
Si dice allora che, affinché il requisito della unidimensionalitá sia soddisfatto, è sufficiente che sia presente un tratto latente "dominante" che influenzi, rispetto ad altri, i risultati del test.[1]
La seconda assunzione, quella dell'indipendenza locale, richiede che, quando le abilitá sono mantenute costanti, la risposta dei soggetti a ciascun item sia indipendente dal quelle date agli altri. Affinché ció sia vero è allora necessario che la risposta fornita da un soggetto ad un item non influenzi in alcun modo quella data a qualunque altro item.
In generale, un modello di risposta all'item assume che il punteggio che un individuo ottiene ad un test possa essere spiegato in termini di *fattori* e *parametri* attraverso una funzione, definita funzione caratteristica dell'item, che esprime in termini probabilistici, la relazione esistente fra le risposte date dai soggetti agli item e i *parametri.* La relazione ipotizzata tra i predittori è una regressione di tipo lognistico, che viene rappresentata da una curva logistica. Il *fattore* (o abilitá, tratto, variabile latente) è ció che vogliamo quantificare mediante il test e ció che determina la prestazione al test. Il *parametro* corrisponde ad alcune caratteristiche degli item quali difficoltá, discriminativitá.

[1] Esistono anche casi in cui il test misura piú di un tratto latente, ad esempio l'indagine Invalsi usa anche modelli multidimensionali, ma ancora in fase sperimentale. In tali casi il test viene definito multidimensionale e richiede una diversa procedura di analisi, la quale peró, esula dai nostri scopi (Hambleton e Swaminathan 1985)

In altre parole la probabilitá di rispondere correttamente ad un item puó essere spiegata dal livello di abilitá del soggetto e dai parametri dell'item. Al crescere del livello di abilitá cresce la probabilitá di rispondere correttamente all'item; maggiore è la difficoltá dell'item, maggiore deve essere l'abilitá del soggetto per riuscire a rispondere correttamente.

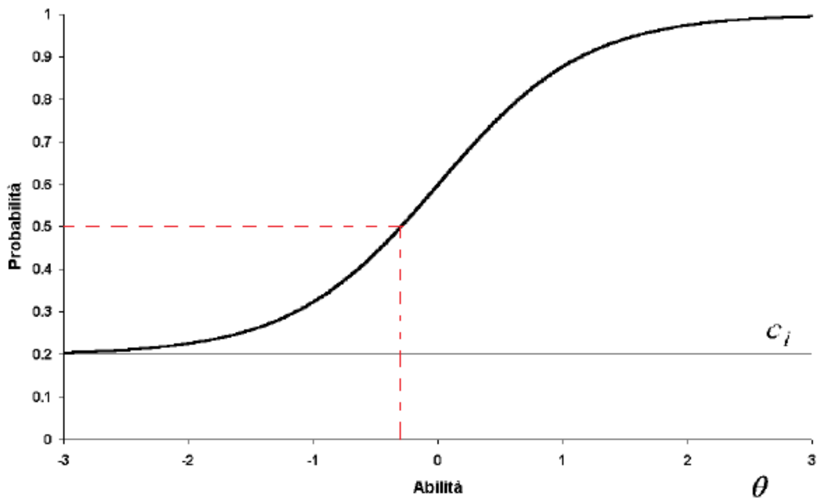


Figure 3.3: Curva Caratteristica, modello per dati binari

Un modello di risposta all'item quindi consente di:

- rappresentare ciascun item tramite una curva caratteristica che esprime l'idea di difficoltá (intesa come indice di posizione)
- ottenere una scala sulla quale vengono collocati gli item in base alla loro difficoltá e successivamente vengono posizionati i livelli di abilitá

La difficoltá (relativa all'tem) viene interpretata come quel punto nella scala di abilitá in cui la probabilitá di risposta corretta è uguale a $0,5$ (eccezione fatta per il modello a tre parametri, dove l'indice di difficoltá viene interpretato come il punto sulla scala di abilitá in cui la probabilitá di rispondere correttamente è uguale a $(1 + c)/2$, dove c rappresenta un parametro della curva, quindi ad una probabilitá superiore a $0,5$).
La stima dell'abilitá del soggetto (rappresentata dalla variabile θ) si effettua sulla base del punteggio che il soggetto ottiene al test ed è indipendente dal test che viene somministrato, ovvero il livello di abilitá attribuito al soggetto è identico sia che venga somministrato un test facile sia che venga somministrato un test difficile. Quindi modificando il test su cui tale stima viene compiuta, il livello di abilitá che viene stimato resta invariato. E' cosí possibile confrontare i punteggi ottenuti a test diversi purché i punteggi grezzi

siano trasformati in una unitá di misura comune sia per il livello di abilitá che per i parametri degli item (a tal proposito si veda la sezione dedicata all'equating). I punteggi del soggetto vengono interpretati considerando la posizione del soggetto rispetto al tratto latente.
All'interno di questo generale paradigma entrambe le indagini, PISA e TIMSS, utilizzano la stessa forma funzionale del modello, ovvero la funzione logistica, come funzione caratteristica dell'item, su cui si basano tecniche avanzate di **scaling** (trasformazione del punteggio), **linking** (comparabilitá tra test diversi) ed **equating** (comparabilitá tra punteggi ottenuti in test diversi); nonché la stessa tipologia di item e di risposta (dicotomica).
Si riscontra piuttosto un certo grado di differenza nella scelta del numero di parametri assunti per descrivere la prestazione al test, vediamo nel dettaglio.
Nell'indagine TIMSS:

- un modello (IRT) a 2/3 parametri: il livello di **discriminazione** dell'item e il **guessing** (oltre alla **difficoltá** dell'item)

- la calibrazione genera 5 scale per la matematica (e 5 per le scienze), le scale per la matematica sono relative a: numero, algebra, misurazione, geometria e dati.

Nell'indagine PISA:

- un modello (IRT) a 1 parametro (riferito al modello di Rasch): la **difficoltá** dell'item

- la calibrazione genera 4 scale per la matematica relative a: quantitá, spazio e forma, cambiamento e relazioni, incertezza e dati.

3.3 Il modello e la sua predittivitá

I modelli di regressione logistica nascono nell'ottica di spiegare la probabilitá di possesso di un attributo o di accadimento di un evento in relazione ad una serie di possibili determinanti, di variabili esplicative. Un ambito in cui tali modelli trovano speciale applicazione è la LST, la teoria del tratto latente (su cui si basano i sistemi di valutazione comparativa, come quelli in esame). Volendo fare un classico esempio di modelli di regressione logistica e contesto applicativo, un obiettivo potrebbe essere quello di individuare i principali fattori di rischio di una particolare malattia, oppure, per rimanere nel contesto della valutazione, individuare i fattori che sono dietro una certa prestazione. La regressione logistica è un caso particolare di analisi di regressione [2] dove la **variabile dipendente Y** è **dicotomica** invece

[2]Sono tutti modelli riconducibili alla classe dei **modelli lineari generalizzati**, non condizionati alla natura di Y e al tipo di funzione (*funzione legame*) che la lega alla

che quantitativa: convenzionalmente una modalitá viene associata al valore 0 e l'altra al valore 1, per cui la sua distribuzione è ovviamente bernulliana, con media pari alla probabilitá di successo relativo al verificarsi dell'evento. Supponendo di osservare p items è definita una variabile casuale dicotomica associata alla risposta all'item j-esimo:

$$Y_j = \begin{cases} 0 & \text{risposta errata} \\ 1 & \text{risposta corretta} \end{cases}$$

Nel modello di regressione lineare la stima della variabile dipendente Y_j è ottenuta da una combinazione lineare delle esplicative, che teoricamente, puó assumere valori in tutto l'asse reale:

$$Y_j = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \varepsilon$$

$$Y_j = \alpha + \bar{X}\bar{\beta} + \varepsilon$$

Ora il modello di risposta all'item (ITR) che esprime la relazione tra variabile dipendente Y_j, la risposta all'item j-esimo, e variabili esplicative $X_1, X_2, ..., X_n$ (o predittori), che sono i fattori che concorrono alla risposta, è di tipo probabilistico, ovvero il set di predittori é definito dal modello di regressione:

$$E[Y|\bar{X}] = \alpha + \bar{X}\bar{\beta}$$

che, dovendo modellare una probabilitá, non puó assumere un qualsiasi valore da meno infinito a piú infinito, pertanto una scelta di questo tipo non sarebbe ovviamente opportuna:

$$\lambda = P(Y_j = 1|\bar{X}) = E[Y|\bar{X}]$$

poiché perde di significato fuori dall'intervallo $[0, 1]$.
Si ovvia a questo problema ricorrendo ad una trasformazione biunivoca di quest'ultima, che estenda l'intervallo in questione a tutto il campo reale, sará poi necessario ricorrere alla relativa inversa per ricondurre le stime a valori sensati nel campo delle probabilitá. Le alternative sono molteplici [3]; tra queste la piú comune è l'utilizzo della **funzione logistica**:

$$g(\lambda) = logit(\lambda) = ln(\frac{\lambda}{1-\lambda}) = \alpha + \bar{X}\bar{\beta}$$

Utilizzando la funzione logistica resta dimostrato che il logaritmo del rapporto tra la probabilitá di successo e la probabilitá di insuccesso é una funzione legame che prende valori tra 0 $[\lambda = 0]$ e $+\infty$ $[\lambda \to 1]$ e lega la variabile di risposta (Y) alla combinazione lineare di predittori $(X_1, X_2...)$

---

combinazione lineare delle esplicative, purché monotona e differenziabile: $g(E[Y|\bar{X}]) = \alpha + \bar{X}\beta$

[3] Purché tutte funzioni continue e monotone crescenti

in maniera tale da assicurare che per ogni valore assunto dalle variabili esplicative, la probabilitá di risposta sia sempre compresa nell'intervallo $[0, 1]$. Infatti applicando l'inversa si ottiene:

$$\frac{P(Y_j = 1|\bar{X})}{P(Y_j = 0|\bar{X})} = \frac{\lambda}{1 - \lambda} = exp(\alpha + \bar{X}\bar{\beta})$$

Ovvero:

$$\lambda = \frac{exp(\alpha + \bar{X}\bar{\beta})}{1 + exp(\alpha + \bar{X}\bar{\beta})} = \frac{1}{1 + exp[-(\alpha + \bar{X}\bar{\beta})]} \tag{3.1}$$

La scelta di tale trasformazione (logit) per mettere in relazione la probabilitá del fenomeno (associato a Y_j) con le variabili predittive é sicuramente dovuta alla semplicitá di interpretazione dei parametri stimati ma anche al fatto che la sua inversa rispecchia l'andamento a "*sigmoide*" tipico della probabilitá, con un lento avvicinarsi agli estremi 0 e 1.

Per scegliere in modo opportuno le potenziali variabili esplicative da introdurre nel modello è fondamentale definire con chiarezza quali siano gli scopi dell'analisi: nel caso in cui quest'ultima sia prevalentemente orientata alla ricerca dei fattori determinanti di un evento (come fornire la risposta corretta ad un item) si presterá particolare attenzione alla natura delle variabili e si svilupperá lo studio delle loro interazioni. Alcuni accorgimenti da avere nella fase di selezione delle esplicative sono i seguenti:

- di tutte le variabili andranno specificate la natura, la scala e l'unitá di misura
- le variabili selezionate dovranno rappresentare plausibili fattori determinanti della dipendente: andrá analizzata la significativitá delle relazioni intercorrenti tra quest'ultima ed ogni singola variabile candidata ad entrare nel modello di regressione
- dovranno essere prese in considerazione combinazioni di variabili che potrebbero rivelarsi determinanti significative
- per quanto riguarda le variabili quantitative dovrá risultare accettabile l'ipotesi di dipendenza lineare tra logit(λ) ed esse, che altrimenti dovranno essere trasformate in variabili ordinali
- dovrá essere rispettata la condizione di non collinearitá tra i regressori

In accordo alla LST (Latent State-Theory) il tratto latente deve essere decomposto in due parti, il tratto e l'effetto della situazione o interazione, sul soggetto, nella quale ha luogo la valutazione:

$$\theta_t = \zeta + \zeta_t$$

dove

$$\begin{aligned} \zeta &= \text{tratto latente} \\ \zeta_t &= \text{tratto residuo} \end{aligned}$$

Ora, poiché la stima dell'abilitá (o tratto latente) si effettua sulla base del punteggio che il soggetto ottiene al test, questa stima, a sua volta, conterrá due componenti, il valore reale dell'abilitá (sconosciuto) e l'errore di misura:

$$\hat{\theta}_t = \theta_t + \varepsilon_t$$

vale pertanto la seguente:

$$\hat{\theta}_t = \zeta + \zeta_t + \varepsilon_t \tag{3.2}$$

Assumiamo l'ipotesi di invarianza nel tratto se valgono:

$$\begin{aligned} cov(\varepsilon_t, \varepsilon_{t'}) = 0 &\quad \text{gli errori di misurazione sono incorrelati, t}\neq\text{t'} \\ cov(\zeta, \zeta_t) = 0 &\quad \text{tratto e tratto residuo sono incorrelati} \\ cov(\zeta, \varepsilon_t) = 0 &\quad \text{tratto ed errore di misurazione sono incorrelati} \\ cov(\zeta_t, \zeta_{t'}) = 0 &\quad \text{i tratti residui sono incorrelati per t}\neq\text{t'} \\ cov(\varepsilon_t, \zeta_{t'}) = 0 &\quad \text{errore di misurazione e tratto residuo sono incorrelati} \end{aligned}$$

Ora valgono i seguenti fatti:

- la stima del tratto latente ha uno specifico errore standard [4], ovvero varia in funzione del livello di abilitá posseduto dal soggetto:

  $$SE(\hat{\theta}) = SE(\hat{\theta}_{t,u})$$

  (per il soggetto **u** al tempo **t**). L'errore standard tende ad essere minimo intorno al valore di θ in cui l'informazione del test è massima al contrario tende ad aumentare per quei valori di abilitá in cui l'informazione del test é minima. Collegato al concetto di errore standard vi è il concetto di attendibilitá della misurazione, ovvero la precisione della misurazione stessa. Errore standard e attendibilitá hanno un rapporto inversamente proporzionale, ovvero all'aumentare della precisione della misura diminuisce l'incidenza dell'errore commesso e viceversa. Quindi con l'aumentare della precisione di misurazione aumenta anche la probabilitá che il livello di abilitá stimato rappresenti fedelmente il livello (medio) di abilitá posseduto dalla popolazione studenti.

[4] l'errore standard è una stima di quanto la media, relativa all'abilitá, del campione si avvicinerá alla media della popolazione $SE(\hat{\theta}) = \frac{\sigma}{\sqrt{N}}$, analogamente a quanto avviene con la varianza campionaria che fornisce una stima della varianza della popolazione. In genere piú grande è il campione minore sará l'errore standard e piú la media del campione si avvicinerá alla media della popolazione.

Ora, tralasciando per semplicitá di notazione la dipendenza da t e da u, vale:

$$SE(\hat{\theta}) \approx \sqrt{var(\hat{\theta} - \theta)} = \sqrt{var(\varepsilon)} \tag{3.3}$$

e la varianza dell'errore di misurazione:

$$\sigma_\varepsilon = \sqrt{\frac{\sum_{i=1}^{p}(\varepsilon_i - \bar{\varepsilon})^2}{p}}$$

(p gli item osservati).

- non c'è correlazione tra il tratto misurato e l'errore di misurazione, ovvero l'errore di misurazione è indipendente dalla variazione del fenomeno

$$cov(\theta_t, \varepsilon_t) = 0$$

Ora elevando al quadrato l'errore standard $SE(\hat{\theta})$ otteniamo la varianza dell'errore di misurazione, che per la (3.3) essa è anche uguale a:

$$var(\varepsilon) = var(\hat{\theta} - \theta)$$

A questo punto, calcolando $E[var(\hat{\theta}_{t,u})]$ il valore atteso di $var(\hat{\theta}_{t,u})$ su un campione di soggetti (cioé al variare di t e di u) possiamo conoscere, in termini probabilistici, quanto incide la varianza dell'errore sul punteggio relativo $\hat{\theta}$ e conseguentemente, calcolando la differenza tra varianza dell'errore di misurazione e varianza dell'abilitá osservata, la varianza vera:

$$var(\theta) = var(\hat{\theta}) - var(\varepsilon)$$

Ovvero posto che:

$$\begin{aligned} var(\zeta) &= cov(\hat{\theta}_t, \theta_{t'}) && \text{varianza del tratto per t}\neq\text{t'} \\ var(\zeta_t) &= var(\hat{\theta}_t) - var(\zeta) - var(\varepsilon_t) && \text{varianza del tratto residuo} \end{aligned}$$

è possibile separare la varianza del tratto dalla varianza del tratto residuo. Osserviamo che la varianza vera di θ sará la media di θ poichè nella stima dell'abilitá sono ammesse possibili variazioni nella manifestazione del fenomeno stesso.

Dall'equazione generale:

$$\lambda = P(Y_j = 1|\bar{X}) = \frac{exp(\alpha + \bar{X}\bar{\beta})}{1 + exp(\alpha + \bar{X}\bar{\beta})} \tag{3.4}$$

si ottengono diversi modelli, sostituendo al posto della combinazione lineare delle esplicative i parametri relativi al tratto latente e alle caratteristiche degli item, tra i quali ricordiamo:

- Modello logistico di Rasch (utilizzato da PISA)
- Modello logistico a 2-3 parametri (utilizzato da TIMSS)

Nel caso specifico del modello di Rasch il parametro fondamentale, ossia l'altra "causa" del punteggio osservato oltre all'**abilitá** (denotata con θ) del soggetto è la **difficoltá** dell'item, **b**, che come abbiamo detto corrisponde al (o è inteso come il) livello di abilitá necessario perché il soggetto abbia una probabilitá di rispondere correttamente uguale alla probabilitá di non riuscirci (o abbia una probabilitá di rispondere positivamente ad un item di prestazione tipica uguale alla probabilitá di rispondere negativamente). Dall'equazione generale (3.4) si ricava il seguente modello logistico ad un parametro:

$$\lambda = \frac{exp(\theta - b)}{1 + exp(\theta - b)}$$

Pertanto, anziché parlare di variabili esplicative (modello teorico) si parlerá di indicatori categoriali (i parametri) che devono essere trasformati in "objectives measures" quantitative e continue "calibrate" lungo l'intero arco dei numeri reali.
Nel caso invece del modello a 2-3 parametri (Birnbaum 1968) utilizzato da TIMSS si aggiungono: il parametro della discriminativitá, **a** che è la capacitá dell'item di discriminare fra individui con diversi livelli nel tratto, e il guessing, **c**, un parametro che tenga conto del fatto che in un test a scelta multipla un soggetto con bassa abilitá potrebbe ottenere una risposta corretta anche tirando ad indovinare (da cui un asintoto inferiore corrispondente alla probabilitá che ha il soggetto di ottenere la risposta corretta solo per caso), otteniamo cosí il seguente modello logistico a 2-3 parametri (se **c**=0 si definisce un modello a 2 parametri):

$$\lambda = c + (1 - c) \cdot \frac{exp \cdot a(\theta - b)}{1 + exp \cdot a(\theta - b)}$$

Nel modello di Rasch si assume che non ci sia la possibilitá che il soggetto possa rispondere correttamente solo tirando a caso (c=0) e che tutti gli item abbiano la stessa discriminativitá (a=1).

Vediamo un esempio di curve caratteristiche di item secondo il modello ad un parametro:

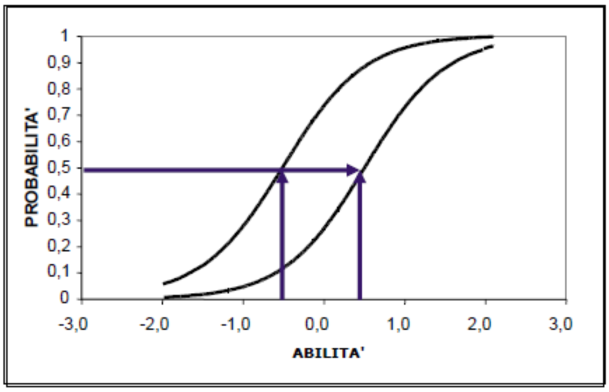


Figure 3.4: Due item con diverso livello di difficoltá ma uguale livello di discriminativitá

Si osserva che il 50% della probabilitá di rispondere correttamente implica, rispetto ai due item, livelli di abilitá diversi. Convenzionalmente l'origine delle stime di Rasch (posta a zero), per i soggetti e per gli item, è la difficoltá media degli item ed è intesa come l'abilitá di un soggetto che ha il 50% di probabilitá di rispondere correttamente. Valori negativi indicano livelli inferiori alla difficoltá media (valori positivi livelli superiori).
Graficamente, la capacitá discriminativa di un item è data da quanto è "ripida" la curva nella sua parte centrale (la parte in cui la probabilitá di rispondere correttamente vale 1/2) ossia dalla sua inclinazione: quanto piú questa inclinazione è vicina alla verticale, tanto maggiore sará la capacitá del test di distinguere fra soggetti con diversa abilitá. Quindi item con la stessa discriminatività corrispondono a curve caratteristiche parallele (con la stessa pendenza), come è evidente nella figura (3.4).

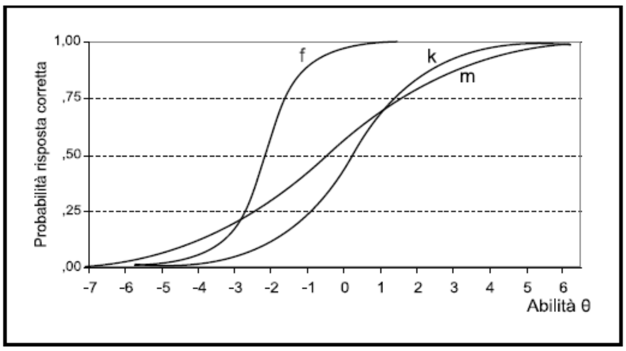


Figure 3.5: Due item con diverso livello di difficoltá e discriminativitá

Se osserviamo la figura che segue notiamo come tutti e tre gli item abbiano lo stesso livello di difficoltá (l'abilitá necessaria per avere almeno il 50% di probabilitá di rispondere corrispondente a -0.5) ma siano diverse le inclinazioni

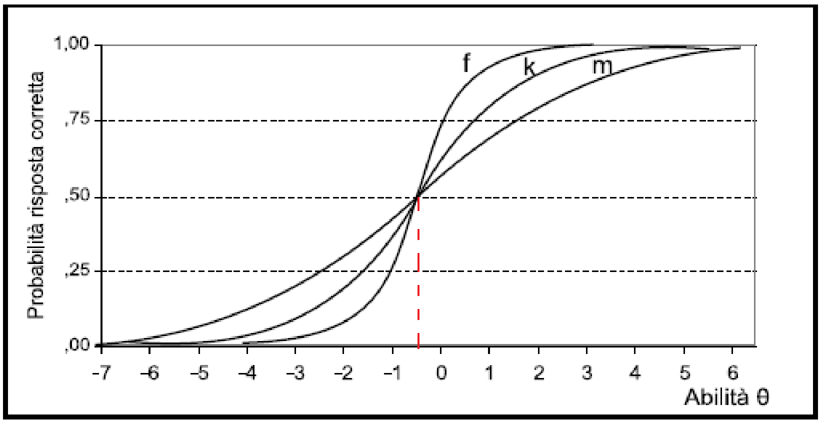


Figure 3.6: Curve caratteristiche di item con stesso livello di difficoltá ma diversa discriminativitá

In particolare, la curva caratteristica dell'item **f**, nella parte centrale, è la piú ripida, mentre la curva dell'item **m** è la meno ripida.
Prendiamo in considerazione i punti in cui le tre curve intercettano le rette orizzontali della probabilitá di rispondere correttamente uguale a .25 e .75 e individuiamo i valori corrispondenti di abilitá θ.

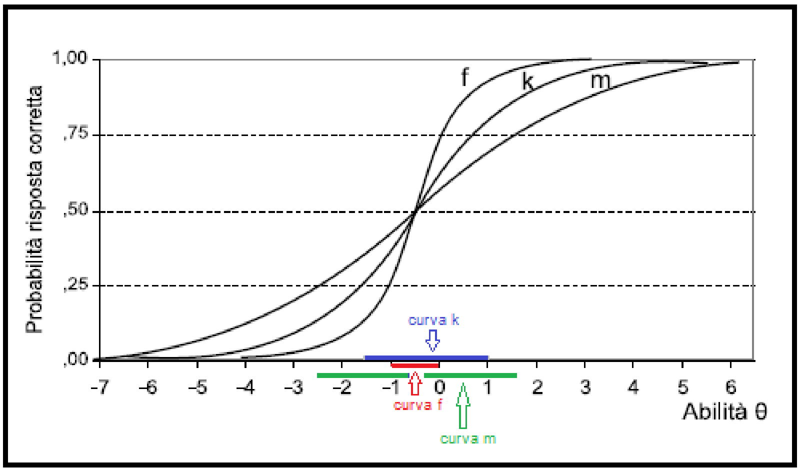


Figure 3.7: abilitá richiesta per l'item f: $\theta\epsilon[-1,0]$, abilitá richiesta per l'item k: $\theta\epsilon[-1.5,1]$, abilitá richiesta per l'item m: $\theta\epsilon[-2.5,1.5]$

Questo significa che nel caso dell'item **f** una variazione minima di abilitá (da -1 a 0) aumenta la probabilitá di rispondere correttamente dal 25% al 75% mentre per ottenere la stessa variazione negli altri item sono necessari aumenti maggiori nel livello di abilitá. In altri termini, l'item **f** è quello piú efficace nell'individuare differenze minime nel livello di abilitá.
Qual'è quindi il vantaggio del modello a 2-3 parametri, rispetto al modello ad un parametro solo? Poiché, secondo l'IRT, l'assunzione di base circa la probabilitá che un compito venga eseguito correttamente dipende:

- dal livello di sviluppo di una competenza specifica, ad esempio: la capacitá di ragionamento numerico
- dal grado di difficoltá del compito richiesto

lo studio di una curva caratteristica dell'item, fortemente discriminante, consente di scegliere quegli item che hanno un elevato potere informativo per una regione di θ specifica. Ricordiamo inoltre che l'errore di misurazione tende ad essere minimo intorno al valore di θ in cui l'informazione del test [5] è massima, pertanto il parametro della discriminativitá si rivela particolarmente vantaggioso al fine della costruzione di test ed item ad hoc per una specifica regione di abilitá, le cui funzioni informative vengono massimizzate per quella regione e solo per quella regione di θ l'errore di misurazione è minimizzato.
A questo punto ci chiediamo:

1. in che misura la differente modellizzazione assunta in PISA e TIMSS influisce sulla possibilitá di confrontare i risultati ottenuti dai due studi?
2. quanto sensibili sono i dati forniti dai due studi per modellare le supposizioni (circa il possesso di una specifica dimensione cognitiva)?

Brown, Micklewright, Schnepf e Waldmann (2005) hanno tentato di affrontare il problema confrontando i risultati di quattro indagini che includevano PISA (2003) e TIMSS. Tale confronto sottendeva un lavoro compiuto sulle edizioni di TIMSS 1995 e 1999 durante le quali, l'analisi statistica che ne è scaturita, ha fatto uso di entrambi i modelli: il modello di Rasch e il modello a 2-3 parametri, consentendo un confronto diretto (rispetto alla stessa indagine) dei risultati ottenuti. Si è riscontrato come la stima del tratto è indipendente dalla scelta del modello utilizzato (in altre parole la presenza di una competenza specifica), piuttosto sembra che le misure di dispersione siano suscettibili al tipo di modello IRT usato, con la possibile e conseguente inferenza sia sul livello di sviluppo della competenza osservata che sulla sua persistenza in termini campionari.

[5] ricordiamo che la funzione di informazione del test si ottiene dalla sommatoria delle singole curve degli item

Questo non risponde direttamente alle domande ma fornisce una ragione per sospettare che alcune differenze nei risultati possono essere ricondotte ad una diversa modellizzazione.

3.4 Stima dei Parametri

Note le risposte degli individui agli item del test, si devono stimare:

- i parametri degli item
- le abilitá degli individui (scoring)

Si distinguono a tal proposito le seguenti procedure:

- Stima congiunta per stimare tutti i parametri contemporaneamente (metodo della massima verosimiglianza o metodi Bayesiani)
- Stima dei parametri degli item, note le abilitá oppure stima delle abilitá, noti i parametri degli item

Noi ci soffermeremo, senza comunque andare in profonditá, sul metodo della massima verosimiglianza (congiunta).
Detta Y_i la variabile casuale dicotomica di risposta all'item i-esimo, $Y_i = 1$ se la risposta è corretta, $Y_i = 0$ se la risposta è errata; è definito un vettore casuale di dimensione p (p è il numero degli item osservati) associato alle risposte date all'insieme di tutti gli item:

$$\bar{Y} = (y_1, y_2, ..., y_p)$$

che puó assumere 2^p valori (tutte le possibili combinazioni binarie di 0,1). Poichè la singola Y_i ha una distribuzione bernoulliana, la sua media sará pari alla probabilitá di successo relativo al verificarsi dell'evento
$(\lambda = P(Y_i = 1|\theta))$. Allora il contributo del singolo individuo alla funzione di verosimiglianza è quindi λ nel caso in cui egli sperimenti l'evento $Y_i = 1$, $1 - \lambda$ nel caso contrario $Y_i = 0$. Pertanto la sua funzione di probabilitá è di conseguenza (data l'assunzione di **indipendenza locale**):

$$\mathcal{L}_i = P(\bar{Y}|\theta_j, a, b, c) = \prod_i P(Y_{ij} = y_{ij}|\theta_j) = \lambda^{y_i}(1 - \lambda_i)^{1-y_i}; \quad \text{i=1,..,p}$$

data l'assunzione di **indipendenza sperimentale**, la probabilitá associata al campione osservato:

$$P(\bar{Y}|\theta, a, b, c) = \prod_j (\prod_i P(Y_{ij} = y_{ij}|\theta))^6 \quad \text{j=1,...,n i soggetti nel campione}$$

ovvero

$$\mathcal{L} = \prod_i \mathcal{L}_i = \prod_j \prod_i \lambda^{y_i}(1 - \lambda_i)^{1-y_i}$$

[6] le lettere maiuscole rappresentano le variabili casuali, le lettere minuscole realizzazioni di esse

Sia i parametri degli item che le abilitá vengono trattati come parametri fissi ma incogniti, da stimare in modo simultaneo. Si ricercano (secondo metodi iterativi, come Newton-Raphson) i valori dei parametri che massimizzano congiuntamente la funzione di log-verosimiglianza:

$$ln\mathcal{L}(\bar{Y}|\theta, a, b, c) = \sum_{i=1}^{p}[Y_i ln\lambda_i + (1 - Y_i)ln(1 - \lambda_i)]$$

3.4.1 Analisi degli item e utilizzo del modello di Rasch nella fase pre-test

La valutazione delle caratteristiche di una prova standardizzata, relativamente alle caratteristiche psicosometriche degli item, implica un insieme di analisi di tipo statistico sulle risposte degli studenti, volte, in sostanza, a identificare eventuali item che non svolgono nel modo dovuto la loro funzione misuratoria. Le informazioni di base da cui si parte per l'analisi del funzionamento di un item sono sostanzialmente tre:

- quanti alunni hanno risposto correttamente
- quanti hanno scelto ciascuno dei possibili distrattori
- quanti hanno omesso la risposta

A partire da queste informazioni è possibile rispondere a tre domande:

1. quanto è difficile l'item?
2. l'item è in grado di discriminare fra gli alunni piú abili e quelli meno abili?
3. gli alunni hanno scelto tutte le opzioni, o ci sono alcune opzioni che non sono state scelte da nessuno?

La valutazione della difficoltá di una prova è, *in primis*, effettuata basandosi sul giudizio di esperti e su premesse teoriche. Successivamente, dal punto di vista statistico e secondo la teoria classica dei test (calcolo dell'α di *Cronbach* per misurare la coerenza complessiva della prova in oggetto, e calcolo del coefficiente punto-biseriale [7]), viene considerata la percentuale di risposte corrette fornite da un numero adeguato di studenti, tale cioè da coprire una varietá di comportamenti sufficientemente ampia. L'elevato numero di risposte è, in altre parole, un elemento cruciale per ottenere informazioni tecnicamente fondate. Un elemento fondamentale per esprimere un

[7] questo coefficiente si basa sul confronto delle medie relative il totale delle risposte giuste nel test, tra i soggetti rispondenti correttamente e i soggetti con risposta sbagliata, rispetto ad una specifica domanda. Nel calcolo interviene anche la deviazione standard relativa il totale delle risposte giuste, calcolata su tutti gli N soggetti a cui il test è somministrato.

giudizio complessivo adeguatamente approfondito sulle prove e su ciascuna domanda che le compone è l'indice di discriminazione, ossia il coefficiente punto-biseriale (da non confondersi con il parametro di discriminativiá del modello 2 PL). E' necessario, infatti, valutare se le diverse domande contengano o meno delle ambiguitá che possono aver tratto in inganno i rispondenti. In una domanda formulata in modo adeguato e priva di ambiguitá, solo l'opzione corretta si deve associare positivamente con il risultato complessivo della prova, nel senso che coloro che rispondono correttamente al una data domanda devono, in media, ottenere un risultato complessivo nella prova migliore di coloro che rispondono in modo errato o non rispondono affatto. In termini piú tecnici, il quesito deve essere in grado di discriminare positivamente fra studenti piú abili e meno abili. Quanto piú una domanda è discriminante tanto piú essa è in grado di misurare la variazione di probabilitá di fornire la risposta corretta anche per piccole variazioni di abilitá del rispondente. Gli standard di letteratura indicano che l'indice di discriminazione deve raggiungere almeno il valore .20 e puó considerarsi buono quando supera il valore di .25.
Ora supponiamo che venga sottoposto un quesito (di un item) a risposta multipla, con quattro opzioni. Nella tabella sottostante vediamo un esempio di analisi statistica delle risposte date (ricordiamo che questa è la fase pre-test). La prima indicazione che troviamo nella tabella è il numero di rispondenti alla domanda (cases for this item), nel caso in questione l'item 18 (quesito A18) è stato somministrato a 330 studenti. L'informazione immediatamente successiva, che troviamo sulla stessa riga, è l'indice di discriminazione che, relativamente all'opzione corretta, è di .31 quindi, per quanto osservato precedentemente, la domanda in questione risulta avere una discriminazione decisamente buona.

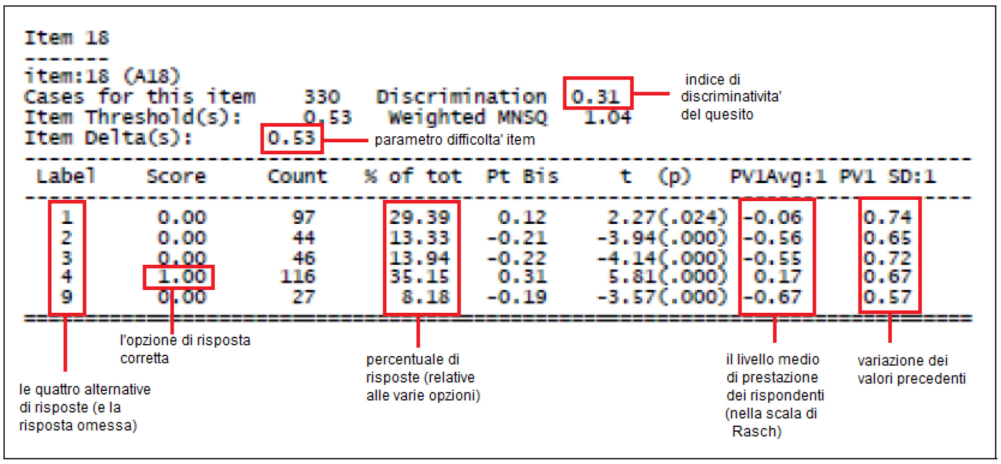
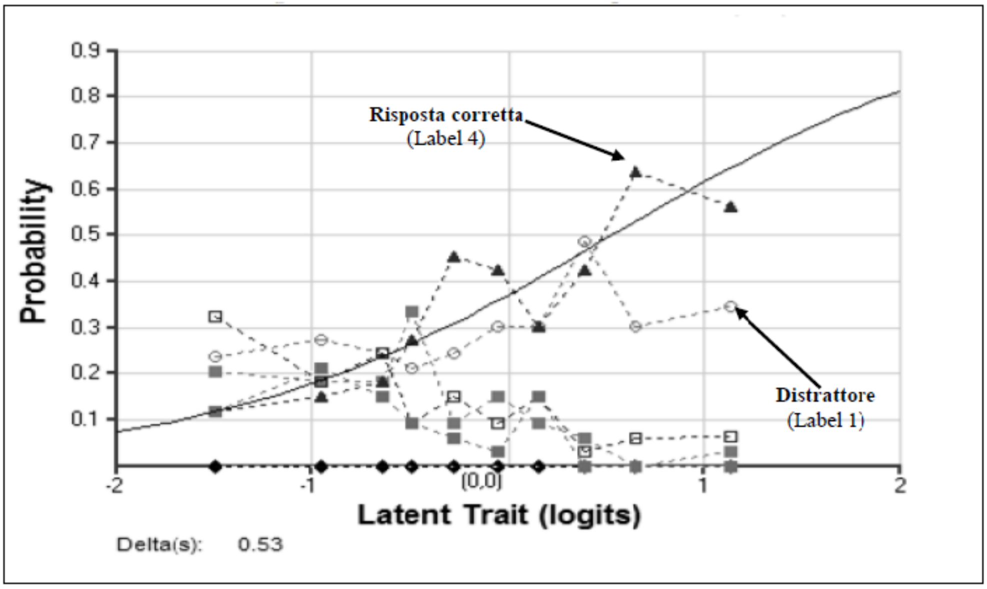

```
Item 18
-------
item:18 (A18)
Cases for this item      330   Discrimination    0.31
Item Threshold(s):      0.53    Weighted MNSQ    1.04
Item Delta(s):          0.53
------------------------------------------------------------------------------
  Label    Score     Count   % of tot   Pt Bis     t   (p)    PV1Avg:1  PV1 SD:1
------------------------------------------------------------------------------
   1       0.00       97     29.39      0.12     2.27(.024)   -0.06     0.74
   2       0.00       44     13.33     -0.21    -3.94(.000)   -0.56     0.65
   3       0.00       46     13.94     -0.22    -4.14(.000)   -0.55     0.72
   4       1.00      116     35.15      0.31     5.81(.000)    0.17     0.67
   9       0.00       27      8.18     -0.19    -3.57(.000)   -0.67     0.57
==============================================================================
```

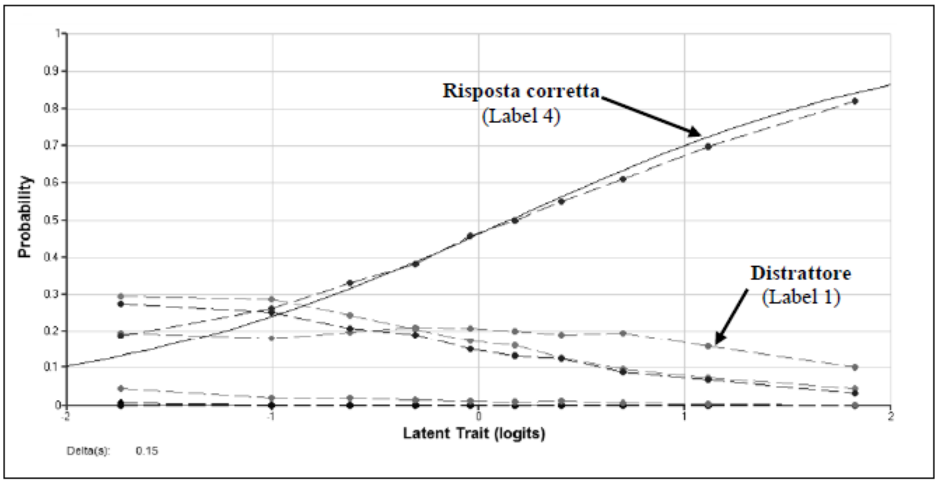

Figure 3.8: Analisi IRT delle risposte al quesito

Nelle due riche successive si trovano gli indici relativi alla difficoltá dell'item (Item Threshold(s), e Item Delta(s)), il valore di tali indici si riferisce alla soglia minima di abilitá che un rispondente deve avere per risolvere correttamente quell'item. I valori di riferimento della scala possono andare da -3 a +3 gli item con valore negativo sono gli item piú facili quelli con valore positivo sono piú difficili. In questo caso la difficoltá dell'item è pari a .53. Nella prima colonna (label) sono indicate le quattro alternative di risposta (1,2,3,4) e la risposta omessa (9). Nella seconda colonna (Score) viene indicata l'opzione di risposta corretta con il codice 1.00 (il codice 0.00 è attribuito alle opzioni errate o mancate). Le colonne tre (Count) e quattro (% of tot) riportano, rispettivamente, le frequenze assolute e percentuali delle risposte registrate. Questa informazione deve essere letta assieme ai dati della colonna cinque (Pt Bis) che riporta i valori della correlazione punto-seriale item-test. Tale correlazione deve essere negativa per le opzioni di risposta non corrette e positiva per quella esatta. Nell'esempio presentato, la correlazione punto-biseriale dell'opzione corretta è +.31. Per le altre opzioni di risposta, la 2 e la 3, hanno correlazioni punto-biseriali negative, invece la 1, che viene scelta da una percentuale di soggetti (circa 30%) vicina a quella di quanti scelgono la risposta esatta (circa 35%) ha una correlazione punto-biseriale positiva (+0.12). Tale risposta si rivela essere un distrattore piuttosto forte. Infine la colonna sette (PV1 Avg:1) permette di valutare su una scala di Rasch il livello medio di prestazione dei rispondenti che scelgono una determinata risposta, mentre la colonna otto fornisce una misura della variabilitá di ognuno dei valori riportati nella colonna sette. E' importante notare che, in media, il livello di abilitá dei rispondenti che scelgono l'opzione corretta di risposta è superiore alla media (che ricordiamo è posta a 0 nella scala di Rasch) e comunque piú elevato di quello di coloro che scelgono le altre opzioni. Ancora una volta peró gli alunni che scelgono l'opzione 1 (il distrattore) mostrano un'abilitá media con valore negativo ma vicina allo 0 (-.06).

Per valutare il funzionamento degli item se ne studia anche la Curva Caratteristica. In riferimento al quesito A18, preso come esempio, anche l'elaborazione della Curva Caratteristica conferma la necessitá di modificare la prima domanda (distrattore).



L'andamento della curva corrispondente alla risposta corretta (linea tratteggiata con triangoli), non perfettamente allineata con la Curva Caratteristica (linea continua) che il modello IRT richiede, risente del fatto che il distrattore (la domanda 1) identificato con la linea tratteggiata con cerchi, ha attirato anche studenti con un buon livello dell'abilitá misurata, i cui valori sono indicati in ascissa e conseguentemente l'andamento della curva relativa alla risposta corretta risulta male approssimata (troppi dati sono sottratti a motivo del distrattore).
Sia dunque i dati nella precedente tabella sia la Curva Caratteristica dell'item convergono nel segnalare l'esigenza di procedere alla sostituzione dell'opzione di risposta 1. Nella versione definitiva della prova il primo distrattore del quesito è stato completamente cambiato, come si puó visualizzare nella pagina seguente.



La Curva Caratteristica del quesito (nella figura in alto) mostra chiaramente che la modifica apportata ha funzionato: l'andamento della curva relativa al distrattore 1 (modificato) si allinea a quello degli altri (opzioni 2 e 3) e la curva dell'opzione di risposta corretta si allinea a quella prevista dal modello.

Diamo un cenno, infine, a titolo puramente descrittivo, dell'implementazione del modello di Rasch, nel quadro della teoria del tratto latente da cui IRT ha tratto ispirazione. Consideriamo la formula:

$$\lambda = \frac{exp(\theta - b)}{1 + exp(\theta - b)}$$

dove ricordiamo $\lambda = P(Y_j = 1|\theta)$ e $\theta_t = \zeta + \zeta_t$
I valori della variabile tratto latente ζ sono generati in modalitá random con distribuzione normale standard [8], quindi E[ζ]=0 e var(ζ)=1 (valore atteso e varianza). Similmente, i valori della variabile ζ_t tratto residuo, sono generati (modalitá random) con distribuzione normale e var(ζ_t)=.30. Ora sommando i valori di ζ_t e ζ per ogni persona si ottengono i valori della variabile θ. Questi parametri (incluso quello della difficoltá b che è preso invariante) consentono di calcolare la soluzione, in termini di probabilitá, dell'equazione data. Questi valori (le soluzioni) sono poi usati per generare, in modalitá random, le variabili manifeste (le risposte) 0,1, agli eventuali item; il valore 1 è generato con $P(Y_j = 1|\theta)$, il valore 0 è generato con $1 - P(Y_j = 1|\theta)$.
La procedura viene inizializzata assumendo che il soggetto abbia abilitá 0 (valore medio nella scala di Rasch). Se il primo item sottoposto, selezionato tra quegli item che massimizzano la funzione caratteristica del test, è risolto,

[8] ricordiamo l'assunzione secondo cui la capacitá intellettiva ha una distribuzione normale

l'abilitá è accresciuta di .25 e si procede con l'item successivo, selezionato analogamente al primo. Quando l'item non viene risolto la variabile relativa all'abilitá assume un valore ottenuto dalla funzione di verosimiglianza e si procede con gli item successivi. Se invece il primo item non è risolto l'abilitá è scalata di .25 e si procede come indicato.
I parametri essenziali da riprodurre in un'analisi statistica sono la varianza di ζ e ζ_t e la qualitá delle stime della varianza di ζ dipendono dalla dimensione del campione e dalla precisione del criterio di stop del programma.

3.5 Procedure di equating

L'equalizzazione è una procedura (che rientra tra quelle tecniche avanzate note come *scaling to achieve comparability*) il cui scopo fondamentale è di permettere la somministrazione di forme alternative di un test e al tempo stesso poter confrontare i punteggi di soggetti che ricevono versioni diverse. I punteggi conseguiti dagli studenti a cui sono somministrati test di profitto devono poter essere ricondotti su di un'unica scala per permettere confronti fra studenti di classi diverse o istituti diversi, o consentire un confronto del rendimento scolastico di uno stesso studente col passare del tempo. Infatti uno dei maggiori punti di forza di TIMSS è la misurazione dell'andamento nel rendimento in matematica (e scienze) nel corso del tempo. Quindi successivamente alla raccolta dati tutte le risposte degli studenti vengono riportate a una metrica comune per ciascun anno di scolaritá oggetto di indagine (4^o e 8^o) allo scopo di fornire un quadro generale dei risultati. Il motivo principale per cui si ricorre all'equalizzazione di versioni diverse dello stesso test è quello di annullare le differenze di difficoltá fra di esse, in modo da consentire un confronto equo fra i soggetti. Poiché la difficoltá degli item è uno dei parametri centrali, insieme all'abilitá dei soggetti, nella teoria della rispsta all'item sono state sviluppate strategie di equalizzazione anche in questa cornice di riferimento teorica. L'equalizzazione è un processo a tre fasi:

1. la progettazione di un metodo di raccolta dei dati che consenta successivamente l'equalizzazione

2. la verifica che le stime dei parametri provenienti dalle calibrazioni delle due forme di un test possano essere considerate sulla stessa scala

3. l'impiego della relazione fra le abilitá e i punteggi veri nelle due forme, per stabilire una relazione fra il punteggio vero e quello riscalato per la forma del test da equalizzare

Compatibilmente alla fase della progettazione, si pone il problema della conversione del punteggio, ovvero la procedura che consente di mettere in corrispondenza la scala dei punteggi ottenuti dagli item nelle due forme di test somministrati. La conversione del punteggio puó essere ottenuta con diverse procedure quali:

- *equalizzazione delle medie*, in cui le medie delle due forme vengono poste uguali (cioé se nella forma A del test si è ottenuto una media di 55 e nella forma B una media di 60, occorre che il punteggio 55 e il punteggio 60 corrispondano allo stesso punteggio sulla scala comune, in modo che rappresentino lo stesso livello nel costrutto)

- *equalizzazione lineare*, in cui viene imposta l'uguaglianza di medie e deviazioni standard

- *equalizzazione equipercentile*, in cui si impone uguaglianza di medie, deviazioni standard e di skewness e curtosi

Commentiamo l'*equalizzazione lineare*. Supponiamo di aver somministrato ad un sottocampione casuale di soggetti la forma A e ad un altro sottocampione casuale la forma B di un test. Diversamente da quanto avviene nell'equalizzazione delle medie, la differenza di difficoltá relativa alle due forme (A e B) è considerata una variabile, per cui, ad esempio, la forma A potrebbe essere relativamente piú difficile per chi ha un livello basso nel costrutto (o tratto latente) rispetto a chi ha un livello alto. Pertanto, per calcolare il punteggio (la scala comune) si utilizza una conversione di tipo lineare secondo la formula:

$$\frac{X_A - M_A}{S_A} = \frac{X_B - M_B}{S_B}$$

dove M_A, S_A, M_B, S_B sono media e deviazione standard delle forme A e B rispettivamente. Ora risolvendo per X_A abbiamo che:

$$X_A = \frac{s_A}{s_B} X_B + \left(M_A - \frac{s_A}{s_B} M_B \right)$$

che ci permette di convertire i punteggi della forma A (X_A) su una scala comune, quella della distribuzione della forma B (X_B), tenendo conto del fatto che la difficoltá delle due forme varia in funzione del livello del soggetto nel costrutto.

Relativamente alla fase della progettazione possono essere adottate diverse tecniche quali: la *single group design*, la *random group design* o la *common item nonequivalent groups design*, sulla quale ci soffermeremo. In questo metodo, le versioni A e B del test hanno alcuni item in comune, che possono o meno essere considerati nel calcolo del punteggio totale; se sono considerati (gli item comuni) si parla di *internal design* e gli item sono somministrati in ordine casuale insieme agli altri, altrimenti si parla di *external design* in cui gli item comuni vengono somministrati separatamente e in questo caso possono costituire un vero e proprio terzo test a sè stante. Se per esempio abbiamo somministrato due versioni di un test composto da 50 item a due gruppi casuali di soggetti e la media di riposte corrette nella versione A è 26 e nella versione B è 33, a cosa possiamo attribuire questa differenza? Diversa preparazione/abilitá dei soggetti? Item piú difficili nella versione A? O una combinazione dei due fattori? In base alle informazioni che abbiamo a nostra disposizione non possiamo saperlo. Supponiamo peró che 20 item fossero stati uguali nelle due versioni e che nel gruppo che ha ricevuto la versione B la media delle risposte corrette, relativamente ai 10 item uguali, sia stata 16 e nell'altro gruppo 15. Abbiamo quindi che i soggetti del gruppo A hanno risposto correttamente in media, al 52% delle domande totali e al 75% di quelle comuni, mentre quelli del gruppo B hanno risposto correttamente al 66% delle domande totali e al 80% di quelle comuni. Questo significa che la differenza di prstazione al sottogruppo di item comuni non è sostanziale, mentre quella nelle altre domande sí: in qualche modo la prestazione agli item comuni funge da "controllo" per quella nel test totale, per cui possiamo concludere che dato che la prestazione negli item comuni è simile, la differenza a livello di punteggio totale puó essere attribuita al fatto che la versione A era piú difficile.
A questo punto gli item comuni vengono utilizzati per creare la scala comune di riferimento fra le due forme del test, in particolare nella fase di stima dei paramentri. Per quanto riguarda l'ampiezza campionaria necessaria, l'aspetto fondamentale è che il numero di soggetti consenta di ottenere stime dei parametri stabili in base al modello IRT utilizzato: questo significa che piú complesso sará il modello di riferimento maggiore sará la numerositá campionaria necessaria. Cook e Eignor (1991) riferiscono che i modelli 3PL necssitano di non meno di 1800 soggetti.

Per quanto riguarda il posizionamento delle stime dei parametri su una stessa scala si consideri il seguente esempio, tratto da Cook e Eignor (1991): supponiamo di aver somministrato la stessa forma del test a due gruppi indipendenti di soggetti per cui ogni stima dei parametri avviene due volte, una per gruppo, e di aver utilizzato un modello 2PL che permette la stima non solo dell'abilitá θ dei soggetti e della difficoltá **b** dell'item, ma anche della discriminativitá **a** di ogni item: in base alla proprietá dell'invarianza (secondo la quale la conversione del punteggio grezzo equalizzato deve essere indipendente dai dati utilizzati per derivare la conversione e deve essere applicabile in tutte le situazioni simili), le stime dei parametri nei due gruppi dovrebbero essere uguali al netto dell'errore di campionamento. Stocking e Lord (1983), peró, hanno mostrato come l'espressione della probabilitá di fornire una risposta corretta sia sí funzione di $a(\theta - b)$ ma anche che l'origine e l'unitá di misura dell'abilitá del soggetto e della difficoltá dell'item siano indeterminate. Se il modello per la probabilitá di risposta corretta all'item **i** nella forma 1 del test è data da:

$$P_{1i} = \frac{e^{Da_{1i}(\theta_{1i} - b_{1i}})}{1 + e^{Da_{1i}(\theta_{1i} - b_{1i}})}$$

una trasformazione lineare, per ottenere una stima su una scala comune, di θ_i e b_i sará:

$$\tilde{\theta}_{1i} = A\theta_{1i} + B \qquad \tilde{b}_{1i} = Ab_{1i} + B$$

dove A e B sono la pendenza e l'intercetta che definiscono la trasformazione lineare. Il valore trasformato di a_{1i} invece puó essere ottenuto moltiplicando il reciproco della pendenza A per a_{1i}:

$$\tilde{a}_{1i} = \frac{1}{A} \cdot a_{1i}$$

Nel momento in cui andiamo a considerare come la nuova funzione di risposta $P(\tilde{\theta}_{1i})$ è in relazione con quella iniziale $P(\theta_{1i})$ otteniamo cosí:

$$\tilde{a}_{1i}(\tilde{\theta}_{1i} - \tilde{b}_{1i}) = \frac{1}{A}a_{1i}[A\theta_{1i} + B - (Ab_{1i} + B)] = a_{1i}(\theta_{1i} - b_{1i})$$

Come si puó notare, l'uguaglianza vale per qualunque trasformazione lineare dei parametri, da cui l'indeterminatezza della metrica per l'abilitá (Harris, 1989). La stessa indeterminatezza puó essere mostrata anche per il modello 3PL. I software permettono di risolvere questo problema stabilendo un'origine e un'unitá di misura per la metrica dei parametri in base a quelli del gruppo utilizzato per calibrare gli item.

Poiché le stime dei parametri di due diverse procedure di calibrazione differiscono solo relativamente all'origine e all'unitá di misura (ammesso che i due gruppi non abbiano la stessa abilitá in quel caso le due procedure non differiranno affatto) la trasformazione richiesta per ricondurre ad una metrica comune le due stime è una classica trasformazione lineare.
La procedura piú semplice di equalizzazione è quella che cerca di determinare i valori di A e B in modo che la media e la deviazione standard trasformate della distribuzione delle difficoltá stimate degli item della seconda calibrazione siano uguali alla media e alla deviazione standard delle difficoltá stimate degli item della prima calibrazione.
Supponiamo di aver eseguito la calibrazione di due diverse forme di test con un modello 3PL e che queste forme abbiano 30 item comuni. La media e la deviazione standard della stima della difficoltá degli item per i 30 item nella prima calibrazione sono rispettivamente: .80 e 1.00; mentre nella seconda sono: .50 e .90. I parametri della trasformazione lineare possono essere calcolati come:

$$\frac{b_1 - .80}{1.00} = \frac{b_2 - .50}{.90} \quad \Rightarrow b_1 = \frac{1.00}{.90}(b_2 - .50) + .80$$
$$\Rightarrow b_1 = 1.11b_2 + 2.4$$

e per la relazione:

$$\tilde{\theta}_{1i} = A\theta_{1i} + B \qquad \tilde{b}_{1i} = Ab_{1i} + B$$

vale:

$$A = 1.11 \quad e \quad B = .24$$

L'uso delle stime di difficoltá degli item come parametri di equalizzazione è giustificato dal fatto che sono le piú stabili. Una volta determinati A e B, questi valori vengono applicati a tutte le stime dei parametri che necessitano la trasformazione. Per cui se $\tilde{b}_i$ è la stima della difficoltá degli item ottenuta dalla calibrazione dell'item i e $\tilde{b}_i^*$ è la trasformata della stima della difficoltá dell'item, allora $\tilde{b}_i^*$ puó essere calcolato come:

$$\tilde{b}_i^* = A\tilde{b}_i + B$$

Allo stesso modo, il parametro di discriminazione **a** e l'abilitá θ possono essere stimati come:

$$\tilde{a}_i^* = \frac{\tilde{a}_i}{A} \quad e \quad \tilde{\theta}_a^* = A \cdot \tilde{\theta}_a + B$$

Nel caso particolare del modello 3PL il parametro di guessing **c** non necessita trasformazione poiché viene determinato dalla metrica di probabilitá, che è necessariamente uguale per tutti.
Nel caso particolare dei modelli di Rasch (o 1 PL) l'unico elemento di indeterminatezza è l'origine della metrica dell'abilitá. I software permettono di risolvere l'indeterminatezza stabilendo l'origine della metrica dell'abilitá in base all'abilitá del gruppo di soggetti utilizzato per calibrare gli item, ossia fissando le stime di difficoltá a zero, oppure fissando la scala in qualche altro modo, ma la relazione fra le scale derivate dall'applicazione di due diversi modelli di Rasch differirá sempre solo di una costante. Se eseguiamo un modello di Rasch per due gruppi separati di soggetti che hanno ricevuto due forme del test con 40 item comuni, e la stima delle difficoltá degli item nel primo gruppo è .75 e nel secondo gruppo è .45 avremo che:

$$b_1 - .75 = b_2 - .45 \quad \Rightarrow b_1 = b_2 + .30$$

la stessa costante .30 dovrá essere applicata per equalizzare le stime di abilitá. Una volta determinate le stime dei parametri per le due forme del test e averle riportate sulla stessa scala, la stima di abilitá ottenuta per ogni soggetto sará la stessa, al netto di misurazione, indipendentemente dalla forma del test somministrata, per cui le stime di abilitá possono essere riportate come stime del livello dei soggetti nel costrutto.
Questo lavoro viene fortunatamente realizzato dai software, ma è possibile trasformare qualunque valore di θ nel punteggio vero stimato $\hat{T}$ nelle due forme e utilizzare tali punteggi veri stimati come punteggi equalizzati. Se le forme J e K, di un test, misurano la stessa abilitá θ, allora la stima dei punteggi veri per le due forme sará in relazione con θ in base alle funzioni caratteristiche del test:

$$\hat{T}_J = \sum_{m=1}^{n_J} \hat{P}_m(\theta) \quad \text{e} \quad \hat{T}_K = \sum_{q=1}^{n_K} \hat{P}_q(\theta)$$

dove $\hat{P}_m(\theta)$ e $\hat{P}_q(\theta)$ sono rispettivamente le funzioni di risposta all'item stimate per gli item m, nel test J, e gli item q, nel test K.
Tali equazioni possono essere utilizzate per trasformare qualunque θ (non solo quelli stimati effettivamente dalla somministrazione delle due forme) in un punteggio vero, stimato nelle due forme. Poiché è vero questo e poiché le stime dei parametri nei due gruppi sono indipendenti, la conversione (o relazione fra i punteggi veri nelle due forme) puó essere considerata indipendente dai gruppi utilizzati nello studio. Si noti inoltre come la stima delle abilitá per gli individui ottenute separatamente per le due forme del test dovrebbe essere la stessa, una volta trasformata, ma se i due test differiscono in difficoltá, le relazione fra i punteggi veri stimati sará non lineare, come illustrato in figura.

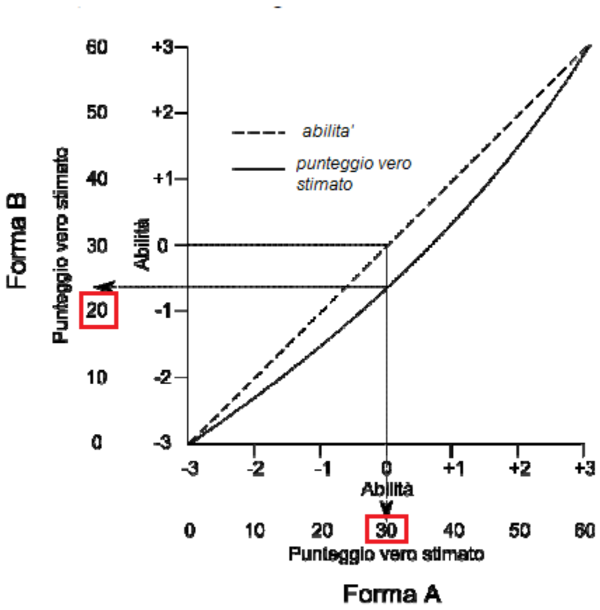


Figure 3.9: Relazione fra le stime di abilitá e i punteggi veri stimati per due forme (A e B) di un test

La figura mostra come lo stesso livello di abilitá consenta di ottenere un punteggio piú alto nella forma A, rispetto alla forma B, che quindi risulta piú difficile. La stima dell'abilitá, peró, è invariante rispetto alla forma somministrata. Cook e Eignor (1991) sottolineano come i punteggi veri stimati non vadano confusi coi punteggi osservati, la cui equalizzazione in base a questa procedura ad ogni modo, non pare portare a risultati particolarmente distorti, nonostante sia metodologicamente scorretto (Lord e Wingersky 1983). Se c'è necessitá di riportare i punteggi su una scala comune come la College Board Scores (CBS) le abilitá vengono trasformate in punteggi veri stimati nelle due forme del test, con le formule che abbiamo visto. Una volta che è stata stabilita la relazione fra i punteggi veri stimati fra le due forme, il passo successivo è individuare la relazione fra i punteggi veri stimati di una forma e i punteggi CBS. Questa fase implica una trasformazione punteggio osservato-punteggio CBS per entrambe le forme, e la relazione fra i punteggi stimati delle due forme viene utilizzata per derivare una stima della trasformazione punteggio vero-punteggio CBS.

Chapter 4

I modelli dei concetti matematici

4.0.1 Il ruolo delle concezioni operative e strutturali nella formazione dei concetti matematici

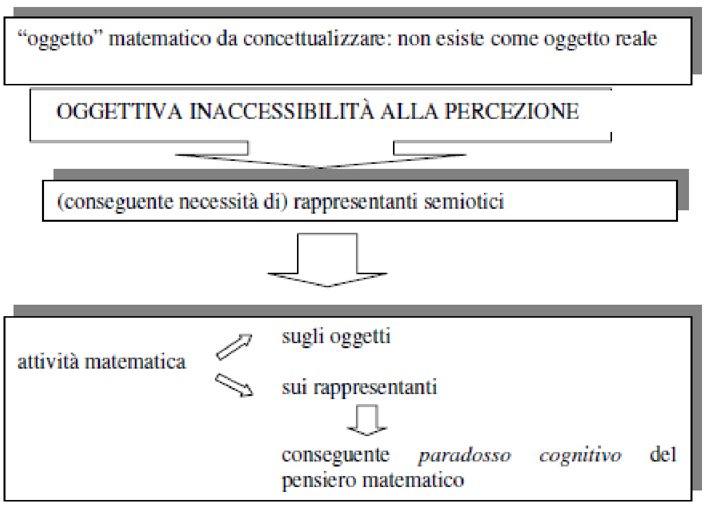
A questo punto della trattazione risulterebbe interessante individuare quali possono essere le caratteristiche della matematica che la rendono cosí difficile da essere appresa o quali livelli risultano facilmente assimilabili e riproducibili in un contesto di verifica/valutazione. Esaminare la natura delle concezioni matematiche potrá aiutarci a rispondere alla domanda: in che cosa l'astrazione matematica richiesta da PISA differisce, nella sua natura, nelle sue funzioni e applicazioni, dall'astrazione richiesta da TIMSS?
Di conseguenza, appare ovvia la duplice riflessione: a che cosa si deve educare quando, a scuola, si fa scienza (intendendo con la parola scienza in realtá tutte le discipline ad essa ascrivibili):

- al metodo scientifico, inteso come padronanza della metodologia
- all'acquisizione e padronanza dei concetti essenziali della scienza

Lungi dal voler fare una disamina sulla natura ontologica degli "oggetti" della conoscenza matematica, non si puó non essere d'accordo con il fatto che diversamente dagli oggetti materiali i costrutti della matematica sono totalmente inaccessibili ai nostri sensi, essi possono essere visti solo dagli occhi della mente. Ogni concetto matematico ha rinvii a "non-oggetti", dunque la concettualizzazione non è e non puó essere basata su significati che poggiano sulla realtá concreta. D'altra parte la capacitá di vedere, in qualche modo, tali "oggetti invisibili" sembra essere una componente essenziale del sapersi muovere nella matematica. La mancanza di questa capacitá puó essere una delle principali ragioni per cui la matematica sembra essere praticamente impermeabile a molti soggetti.

Un'altra ragione sembra essere riconducibile ad una difficoltá di allineamento tra aspettative *semantiche* che insorgono nella mente di chi apprende, rispetto alle diverse argomentazioni e definizioni che la Matematica produce, e la natura *sintattica* della matematica. Ad esempio: non capire nè la definizione di operazione binaria nè il relativo esempio sui numeri razionali, vuol dire aspettarsi una definizione in termini di significato, una definizione che spieghi che cos'è un'operazione binaria su un piano, per cosí dire, filosofico. Mentre la definizione di "operazione" vista all'interno della Matematica vuol subito mettere in chiaro il suo aspetto sintattico, cioè il richiamare una certa "applicazione" che soddisfi a determinate funzionalitá e conseguentemente, nel corso della trattazione teorica, che siano soddisfatte certe proprietá formali dell'operazione, sempre ignorando volutamente il loro eventuale significato.
Perchè questo fa la Matematica: si occupa dell'esistenza e delle proprietá di certe entitá astratte senza preoccuparsi troppo dell'aspetto semantico.
"Vedere" un'entitá matematica come un oggetto significa essere in grado di riferirsi ad essa come se fosse una cosa reale, una struttura statica esistente. Significa essere in grado di riconoscere l'idea nel suo insieme ed essere in grado di manipolarla (approccio strutturale). In contrasto interpretare una nozione come processo implica interpretarla come un entitá "potenziale" (piuttosto che un'entitá vera) la quale esiste nel momento in cui viene utilizzata in una sequenza di azioni (approccio operativo). Vale la pena quindi, chiarita la differenza tra concetto e concezione, tentare di capire la natura dei (o come si differenziano i) concettti matematici e la loro relazione con la prestazione. Nella prossima sezione accenneremo ad un modello esplicativo di un particolare tipo di concetto, che nonostante abbia come privilegiato campo d'applicazione quello delle scienze naturali (fisiche), trova riscontro autorevole nell'idea secondo cui, il modo di procedere della conoscenza (scientifica, ma non solo) sia quello della rete concettuale.
Quando parliamo di concetto matematico ci riferiamo ad un "blocco" di conoscenza matematica inserito in un preciso contesto teorico matematico. In altre parole, un "blocco" di conoscenza vista all'interno della Matematica. Una concezione è un insieme di rappresentazioni interne evocate da un concetto. Questa puntualizzazione è tutt'altro che superflua, perchè spesso la prestazione matematica è il risultato di *come* l'entitá è stata percepita dall'individuo piuttosto che da una suo riconoscimento all'interno di un impianto teorico, con quegli strumenti o processi intrinsechi che la Matematica offre. Prendiamo come esempio la definizione di literacy matematica, la base teorica su cui è costruita l'indagine PISA. Il linguaggio utilizzato per la definizione di literacy matematica è incentrato su un coinvolgimento attivo con la matematica ed è inteso a comprendere la formulazione di un ragionamento matematico e l'utilizzo di concetti, fatti e procedure per descrivere, spiegare e prevedere determinati fenomeni. Ora, se noi analizziamo nello specifico il processo cognitivo della formulazione (rispetto al quale sono

emerse lacune in sede di indagine), questo include la "*capacitá di prendere una situazione cosí come si presenta e trasformarla in una forma gestibile in modo matematico, creando strutture e rappresentazioni matematiche, identificando le variabili e formulando ipotesi semplificative che aiutino a risolvere il problema o la sfida*". Ora, per quanto osservato, l'insieme delle rappresentazioni interne sono evocate da un concetto matematico, il quale per definizione è costretto a servirsi di rappresentazioni, dato che non vi sono "oggetti" da esibire in loro vece o a loro evocazione. Dunque la concettualizzazione (che la formulazione necessariamente implica) deve passare attraverso registri rappresentativi che, per vari motivi, soprattutto se sono a carattere linguistico, non possono essere univoci. Sembra pertanto un paradosso, richiedere/valutare un processo la cui attivazione ha richiesto il processo stesso, o nella migliore delle ipotesi un'istanza piuttosto dispendiosa e farraginosa. Riassumiamo parte di quanto giá detto nel seguente schema (Bruno D'Amore, interpretando Duval, 1993):



dove per paradosso cognitivo del pensiero matematico si intende: "da una parte, l'apprendimento degli oggetti matematici non puó che essere un'apprendimento concettuale e d'altra parte, è solo per mezzo di rappresentazioni semiotiche che è possibile un'attivitá su degli oggetti matematici. L'impossibilitá di un accesso diretto agli oggetti matematici, al di fuori di ogni rappresentazione semiotica, rende la confusione quasi inevitabile. E, al contrario come possono, gli studenti, acquisire la padronanza dei trattamenti matematici, necessariamente legati alle rappresentazioni semiotiche, se non hanno giá un apprendimento concettuale degli oggetti rappresentati?"(Duval, 1993). Il dibattito potrebbe continuare all'infinito, ma un primo elemento da rilevare, collegato alla concettualizzazione e quindi con una grande ricaduta su tutti i processi dell'apprendimento, puó essere individuato nel ruolo della rappresentazione semiotica. Anche se la corretta gestione dell'apprendimento mediante le rappresentazioni semiotiche puó essere causa di difficoltá (settorializzazione delle conoscenze), potrebbe essere preferibile una continua e sistematica integrazione delle varie forme di rappresentazione (o in altre parole un coinvolgimento di piú registri rappresentativi) per favorire l'acquisizione di quelle proprietá concettuali che consentono non solo di distinguere la rappresentazione semiotica dall'oggetto stesso ma di riconoscerne ed estenderne il campo di applicabilitá (*formulazione*).

È possibile osservare che, due entitá matematiche, sulle quali ad esempio si basino due esercizi, possono apparire formalmente analoghe se esaminate dal punto di vista della loro presentazione simbolica, ma una loro considerazione piú profonda, condotta nell'ambito di diversi registri rappresentativi puó rivelare differenze decisive. Illustriamo la situazione con un esempio semplicissimo: agli studenti è ben noto che la disequazione

$$x^2 + 1 > 0$$

è verificata da tutti i valori reali della variabile x. Analoga "universalitá" potrebbe essere (sbrigativamente) attribuita per analogia formale alle soluzioni di

$$[f(x)]^2 + 1 > 0$$

senza considerare le caratteristiche della funzione f, ad esempio il suo dominio il quale potrebbe non essere piú esteso a tutto il campo reale.

In pratica, la sola considerazione della situazione algebrica (l'esame del segno di una somma di quadrati) risulterebbe primaria ed esclusiva, nella mente dello studente, conducendolo cosí a perdere di vista il significato degli stessi simboli o della relazione con essi espressa.

Molto spesso i concetti matematici possono essere definiti o presentati sia in maniera operativa che strutturale e i due modi di interpretare la matematica sembrano essere entrambi necessari (indichiamo ad esempio: A.Sfard, La doppia natura delle concezioni matematiche, 1991). La parola strutturale indica un approccio in cui le nozioni sono trattate come se esse corrispondessero in qualche modo ad oggetti astratti. Come abbiamo osservato, questo non è l'unico approccio possibile, esiste anche quello operativo. Ad esempio una funzione puó essere considerata come un *processo di calcolo*, la simmetria puó essere interpretata come *trasformazione*; in entrambi questi due casi si puó parlare di concezione operativa di una nozione matematica. Vediamo nella tabella che segue altri esempi (A.Sfard 1991):

concetto	**definizione strutturale**	**definizione operativa**
funzione	insieme di coppie ordinate	processo computazionale o metodo ben definito per passare da un sistema ad un altro
simmetria	proprietá di una generica forma	trasformazione di una forma geometrica
numero naturale	proprietá di un insieme o la classe di tutti gli insiemi finiti di uguale cardinalitá	0 o ogni umero ottenuto da un altro naturale aggiungendo uno
numero razionale	coppia di interi (considerata un'opportuna relazione di equivalenza)	risultato della divisione tra interi
cerchio	il luogo dei punti equidistanti da un punto dato	una curva ottenuta ruotando un compasso intorno ad un punto dato

Ora la doppia natura dei concetti matematici si puó notare non solo nelle descrizioni verbali ma anche analizzando vari modi di rappresentare simbolicamente i concetti in questione. Anche se una proprietá, come la strutturalitá di una concezione è interna al soggetto piuttosto che nei simboli, è anche vero che alcuni tipi di rappresentazioni sono piú suscettibili ad una interpretazione strutturale di altri. Ad esempio differenti modi di interpretare una funzione possono essere identificati nei tre modi di rappresentare la funzione $y = 3x^4$ riportata in figura

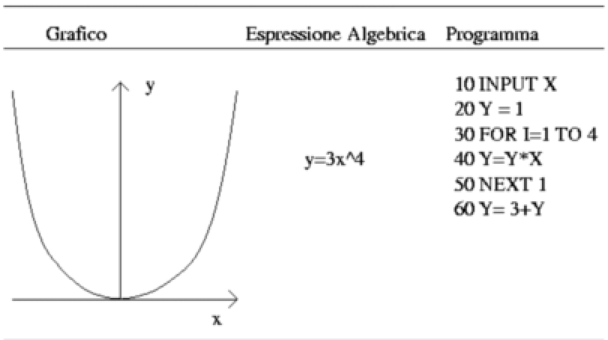


Figure 4.1: Differenti rappresentazioni di una funzione

Il programma, con cui è stata implementata l'espressione analitica della funzione, sembra corrispondere alla concezione operativa di funzione, in quanto la presenta come un processo computazionale, non come un'entitá. Dall'altro lato abbiamo la rappresentazione grafica che (essa stessa potrebbe essere considerata una concezione operativa) combina le infinite coppie di punti della funzione in un'unica linea continua, in questo modo queste infinite componenti della funzione possono essere "prese" come un tutt'uno; il grafico quindi incoraggia l'approccio strutturale.
L'espressione algebrica puó essere facilmente interpretata in entrambi i modi: puó essere spiegata operativamente come una descrizione sintetica di certi calcoli, oppure strutturalmente come una relazione statica tra due grandezze.
La tipologia di una concezione si manifesta anche attraverso le speciali rappresentazioni interne al soggetto. Secondo quanto ormai noto, i concetti matematici a volte sono visualizzati tramite "immagini mentali" mentre altre volte gli stessi concetti sono trattati soprattutto tramite rappresentazioni verbali. Le immagini mentali, essendo compatte ed unificanti, sembrano supportare concezioni strutturali. La visualizzazione, quindi, rende le idee astratte piú tangibili ed incoraggia a trattarle quasi come se fossero entitá materiali. In effetti le immagini mentali possono essere manipolate quasi come gli oggetti reali, inoltre, dalla rappresentazione visiva possono essere estratti vari aspetti del concetto matematico tramite "accesso casuale", senza un ordine prestabilito.

Di contro una rappresentazione verbale non puó essere presa nel suo insieme come tutt'uno ma deve essere processata sequenzialmente, quindi sembra essere piú appropriata per rappresentare procedure di calcolo (o procedure in generale). Per cui le rappresentazioni non visive sembrano essere piú pertinenti al modo operativo di pensare.
Con questo non si vuole dire che ci sia necessariamente una corrispondenza biunivoca tra i tipi di visualizzazione e tipi di concezione, piuttosto si vuole osservare che alcuni tipi di rappresentazioni interne sembrano piú vicine all'approccio strutturale ed altre a quello operativo. In questa prospettiva, la capacitá, richiesta dagli studenti, di utilizzare i concetti, i fatti, i procedimenti e ragionamenti matematici, che si traduce nel saper utilizzare la comprensione del contesto per orientare o organizzare il processo matematico di risoluzione potrá significare, per gli studenti, dare un senso al procedimento risolutivo, mettendo in relazione una molteplicitá di rappresentazioni nell'interazione con il problema (come è sottolineato nel prospetto Relazione tra processi matematici e capacitá fondamentali, PISA).
In altre parole tali capacitá fondamentali potranno essere usate se di fatto sono state opportunamente attivate.
Poiché dei due tipi di definizioni matematiche, quelle strutturali sembrano essere piú astratte, sembrerebbe quindi che l'approccio strutturale possa essere considerato come lo stadio piú avanzato dello sviluppo del concetto. In altre parole ci sono buoni motivi per aspettarsi che: nel processo di formazione dei concetti le concezioni operative precedano quelle strutturali. Questa potrebbe essere una ragione per cui, rispetto all'indagine TIMSS, la prestazione matematica (italiana) sembra godere di un trend significativamente superiore. Ricordiamo che il framework di TIMSS è focalizzato sui domini cognitivi: Conoscenza, Applicazione, Ragionamento. Il dominio Applicazione prevede l'applicazione degli strumenti matematici ad una varietá di contesti; i fatti, i concetti e i procedimenti potrebbero risultare familiari per gli studenti, essendo presentati come problemi di routine. Anche se i problemi si differenziano per livello di difficoltá si prevede che ciascuno di questi, simili a quelli che si incontrano nei libri di testo, siano sufficientemente familiari. Dunque, l'indagine TIMSS, è svolta su un piano di prevedibilitá e rappresenta un'opportunitá per misurare capacitá e abilitá in un contesto cognitivo piuttosto rassicurante, in linea a quanto, gli studenti, si sarebbero aspettati di "dover imparare" o di "saper fare". All'interno del dominio Ragionamento invece si richiede di riflettere e risolvere problemi nuovi o complessi che includono il trasferimento di conoscenze e abilitá a nuove situazioni, ma ricordiamo che, il dominio Ragionamento rappresenta solo il 20% della ripartizione totale, quindi l'esposizione alla prestazione piú impegnativa è piuttosto contenuta. Pertanto, nel disegno di ricerca e impostazione metodologica di TIMSS troviamo che, la capacitá procedurale è quanto ci si aspetta dall'apprendimento della matematica, è il suo naturale epilogo e coinvolgendo forse solo una parte del processo di apprendimento,

risulta cosí piú facilmente performante. Mentre in PISA, se pur con delle forti contraddizioni interne, risulta chiaro che la profonda "conoscenza" dei processi sottostanti i concetti matematici e forse anche l'abilitá nello svolgere tali processi è visto come una base per la comprensione di tali concetti, piuttosto che come obiettivo finale. In altre parole: capire e saper eseguire procedure come base per capire concetti, invece che saper eseguire procedure come scopo finale.

4.0.2 Il Modello della classi di coordinazione

Ora introduciamo una teoria di un particolare tipo di concetto, chiamato: *classe di coordinazione*, con l'esplicita intesa che quanto diremo non ha la pretesa nè tanto meno l'obiettivo di fornire un resoconto di funzionalitá della concettualizzazione o della metacognizione, in una visione unitaria di struttura mentale. Piuttosto quello che verrá descritto, (tratto da "Che cosa cambia nel cambiamento concettuale" di Andrea A. di Sessa e Bruce L. Sherin 1998) assomiglierá di piú ad un sistema di conoscenza (knowledge-system), il cui focus non sará tanto sul concetto, quanto sulle parti costituenti il sistema, coinvolte nel processo della conoscenza e loro performance.
Il concetto, almeno dal punto di vista dell'apprendimento cognitivo, è interpretato oggi come qualcosa di sempre piú vasto, per citare Cornu e Vergnioux:" l'apprendimento di un concetto isolato è impossibile, dato che ogni concetto è correlato ad altri. Si deve parlare allora di trame concettuali".
La seguente esposizione ha pertanto il desiderio di mostrare che ha senso parlare di concetto correlato ad un'acquisizione di conoscenza qualsiasi, purchè questa sia definita e incorporata in un contesto o in un sistema.
Quando ci riferiamo al modello teorico di un concetto, in realtá ci stiamo riferendo ad una classe di concetti, importanti nell'apprendimento scientifico. Una classe di coordinazione è un set di strategie sistemiche volte a ricavare un certo tipo di informazioni da un preciso contesto di riferimento; i concetti matematici e fisici ne rappresentano delle possibili applicazioni.
Perchè parlare del *modello* teorico di un concetto? E come si pone tale modello in relazione agli oggetti della conoscenza scientifica, che sono giá di per sè delle modellizzazioni della realtá? Innanzitutto, avere un articolato modello teorico sulla caratterizzazione del concetto scientifico puó consentire ad una prima istanza, l'organizzazione e la differenziazione del processo cognitivo coinvolto, fondamentalmente rispetto a due situazioni particolari, ovvero:

- l'applicazione di un concetto
- il cambiamento concettuale

e in seconda analisi di localizzare le difficoltá che gli studenti hanno nella concettualizzazione, identificando, almeno teoricamente, le tipologie o componenti stesse del cambiamento.

4.0.3 Le funzionalitá delle classi di coordinazione

Come abbiamo precedentemente osservato, una profonda conoscenza che permette la creazione matematica puó essere difficilmente raggiunta senza l'abilitá di "vedere" gli oggetti astratti e d'altra parte, ricavare o ritenere informazioni utili al fine di costruire una concezione strutturale è altrettanto difficile. La sfida quindi potrebbe essere non tanto "vedere" gli oggetti astratti quanto vedere "in differenti situazioni", penetrare la diversitá e la ricchezza delle diverse situazioni che compiono un affidabile "*readout*" di una particolare classe di informazione. Qualcuno potrebbe obiettare che si è provveduto semplicemente a spostare il focus dell'attenzione dal vedere un *concetto* astratto ad una *capacitá* di lettura della "*realtá*" finalizzata all'astrazione. D'altra parte, in molti casi, il "vedere" è un sostanziale e progressivo compimento dell'apprendimento e solo in parte dipende dalla dotazione di particolari capacitá percettive. Il nostro interesse è comprendere come "vedere in differenti situazioni" puó costituire la funzione centrale della classe di coordinazione e uno spostamento di ció che si "vede" corrispondere al cambiamento concettuale. A tal proposito useremo il verbo *coordinare* o *determinare* informazioni al posto di "vedere", per enfatizzare una prospettiva in termini di classe di coordinazione.
Quando un soggetto immerso in un contesto di apprendimento è sollecitato su piú fronti: operativo e strutturale (relativamente al contesto scolastico) e dalla ricchezza fenomenologica che, richiamata dal curriculum, entra in relazione con l'oggetto della conoscenza e si trova nella condizione di combinare, caratteristiche diverse a seconda della situazione e osservazioni correlate, attraverso una forma (rete) di ragionamento che consenta di inferire le informazioni necessarie, il problema si riconduce sostanzialmente a saper ricavare le relazioni necessarie: *coordinare*, per comprendere: *determinare* il significato delle idee matematiche. Ad esempio molte persone non riescono a comprendere chiaramente che i numeri usati per definire delle unitá di misura, come i grammi, i centimetri, le ore..., hanno un significato diverso rispetto a quelli usati per compiere delle operazioni come l'addizione, la sottrazione ecc. Mentre 2+2 puó risultare 4 in un'operazione, 2 ore piú 2 minuti corrisponde a 122 minuti e non a 4 ore o a 4 minuti. In altre parole non è cosí ovvio che esistano due significati distinti di "numero" uno per rappresentare delle unitá di misura e l'altro usato in operazioni come quelle citate. I problemi inoltre si moltiplicano quanto piú aumenta la complessitá dei concetti matematici o quando si considerano le applicazioni della matematica alle scienze o ad altri campi.

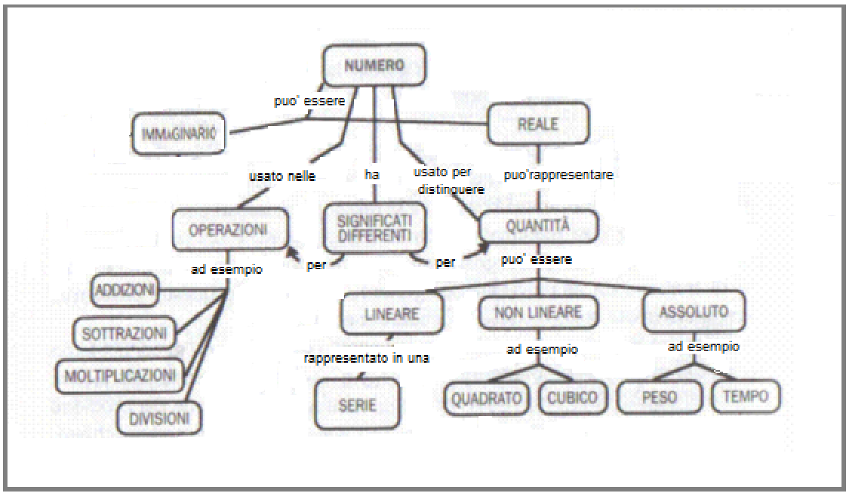


Figure 4.2: Concetti chiave coinvolti nella comprensione del concetto matematico di numero

La coordinazione (il termine scelto al posto di "vedere") ha perció un duplice significato: da una parte si riferisce al fatto che, all'interno di una data situazione molteplici osservazioni o aspetti potrebbero aver bisogno di essere coordinati (come nel grafo concettuale della figura precedente) per poter determinare la necessaria informazione. Questa versione di coordinazione potrebbe essere descritta piú precisamente come **integrazione**. D'altra parte, attraverso i molteplici esempi e le diverse situazioni, la conoscenza che compie un "readout" di informazioni deve, verosimilmente, determinare la stessa informazione (in termini di coerenza). Questo secondo significato di coordinazione puó anche essere definitio **invarianza**.

4.0.4 Le strutture delle classi di coordinazione

Nella nostra introduzione euristica alle classi di coordinazione abbiamo utilizzato il termine "readout" in riferimento a quelle situazioni di ricchezza e diversitá di informazioni e aspetti caratterizzanti le idee e i concetti.
Le strategie di readout costituiscono le prima delle due principali strutture delle classi di coordinazione. Le strategie di readout si caratterizzano a loro volta nelle due performance, precedentemente introdotte, **integrazione** e **invarianza**. Infatti in una specifica situazione di apprendimento sono richiamate azioni mentali di collezione, selezione o combinazione di diversi aspetti per riuscire a rappresentare e formalizzare l'oggetto della conoscenza. Quindi la "capacitá" di vedere in differenti situazioni ha piú a che vedere con una performance di integrazione. D'altra parte la modalitá organizzativa attraverso la quale le osservazioni conducono ad una 'presa' stabile e consolidata di un certo concetto è strutturalmente resa dall'invarianza. La generale

classe di conoscenza e strategie di ragionamento (o pensiero formale) che determina quando e come alcune osservazioni sono in relazione all'oggetto in questione è chiamata: rete casuale. Questo termine si rivela particolarmente adatto alle quantitá fisiche in quanto esse si prestano particolarmente per relazioni di regolaritá che connettono certe precondizioni a risultati attesi (pensiamo ad esempio alle leggi di Newton). La rete casuale rappresenta la seconda delle due strutture delle classi di coordinazione. Readout e rete casuale dovrebbero co-evolvere in un processo di apprendimento, ovvero si dovrebbero manifestare episodi, alla lettera, di *"bootstrapping"* (concetti "tiranti" che fanno da *tramite*) in cui le assunzioni casuali guidano, orientano l'apprendimento, in cui ogni concezione provvisoria aggiunge ed integra la precedente oppure aggiunge solo qualcosa (superposizione e accumulazione, B. D'Amore). Si potrebbero fare molti esempi in Matematica di concetti che si sono evoluti attraverso fenomeni di "bootstrapping" (potremmo ricordare l'introduzione della pseudo-oggetto retta, nella sua presentazione ingenua di segmento lungo e sottile, che cede il passo, con il progredire della formazione, alla retta euclidea, idealizzazione della precedente concettualizzazione, a sua volta arricchitá da nozioni di densitá...B.D'Amore). Ecco che cambiamenti in readout e rete casuale, che costituiscono i parametri del cambiamento concettuale, potrebbero anche essere visti come accessori della conoscenza, la cui "gestione" consapevole/assisitita potrebbe facilitare il "monitoraggio" del processo: *cambiamento concettuale*.

Figure 4.3: Il processo della conoscenza secondo il modello delle classi di coordinazione

La teoria delle classi di coordinazione è coerente con una prospettiva per cui la conoscenza é un sistema complesso e ció che serve per capire la Matematica in modo significativo quindi non sono definizioni chiare (o perlomeno questa è solo la condizione necessaria ma non sufficiente) ma un'esperienza

operativa concreta e autonoma, da far evolvere progressivamente e sintotticamente in un pensiero formale.
Le principali difficoltá nella costruzione di una classe di coordinazione, che rimangono un problema ancora aperto, sono:

- il problema dello span (fornire prospettive diversi, registri diversi della stessa nozione)
- il problema dell'alignement (rispettare la coerenza interna della teoria pur coniugando informazioni diverse in situazioni diverse)

Conclusioni
In questa esposizione abbiamo discusso il ruolo e l'impatto delle Indagini Comparative in Matematica sui sistemi scolastici.
Se da una parte costituiscono un importante benchmarking globale, diventando l'evidenza del successo di alcune riforme educative, dall'altra dovrebbero consentire una continua riflessione sulla qualitá della pratica educativa e fornire le basi teoriche (nonchè pratiche e strutturali) per il progresso delle metodologie e delle strategie di insegnamento.
Le rilevazioni in sostanza misurano il grado di efficacia con la quale i diversi paesi preparano gli studenti all'utilizzo della matematica, fare buon uso dei dati forniti da tali rilevazioni dovrebbe quindi tradursi in una "scuola" realmente in grado di farlo.

Bibliography

[1] [PISA 2012 Quadro di Riferimento analitico]

[2] [Indagini IEA 2011 PIRLS e TIMSS]

[3] [Rapporto Nazionale OCSE-PISA]

[4] [Rapporto tecnico Rilevazioni Nazionali]

[5] [PISA in Finland: An Education Miracle or an Obstacle to Change, Pasi Sahlberg]

[6] [Comparison Between PISA and TIMSS Are We the Man with two Watches?, by Dougal Hutchison, Ian Schagen]

[7] [Education policies for raising student learning: the Finnish approach, Pasi Sahlberg]

[8] [Come si costruisce un test, Guido Amoretti-Giorgio Bolondi]

[9] [Teoria e Tecnica Psicometrica, Carlo Chiorri]

[10] [Latent State-Trait Theory in Computerized Adaptive Testing, Rolf Steyer and Ivailo Partchev]

[11] [La doppia natura delle concezioni matematiche, A. Sfard]

[12] [What changes in conceptual change, Andrea di Sessa, Bruce L.Sherin]