Fondazione per la Scuola
Compagnia di San Paolo

Convegno Internazionale

"La sfida della valutazione"

Torino

24 - 25 Maggio 2011

www.fondazionescuola.it

NORBERTO BOTTANI

"Valutare in modo attendibile i sistemi scolastici"

Valutare cosa?

In questa relazione si parlerà della valutazione dei sistemi scolastici effettuata con prove strutturate che misurano i livelli di apprendimento o di competenza e si tenterà di fornire una risposta alle seguenti domande: perché valutarli? E' necessario valutarli? Non si affronterà invece la questione della valutazione delle scuole e degli insegnanti, delle riforme scolastiche o delle innovazioni. Il punto di riferimento delle considerazioni che seguiranno è un approccio particolare della valutazione di sistema, il che non vuol dire che non ce ne siano altri, che quanto presento sia l'unico valido , ma questi approccio è il più in voga oggigiorno nell'ambito internazionale.

Cosa è la valutazione dei sistemi scolastici?

Mi servo di una definizione fornita da Denis Meuret. La valutazione è:

"Un esame ragionato dei risultati e del funzionamento di un sistema scolastico (anche una scuola, un istituto, sono un sistema scolastico) per esprimere un giudizio sul suo conto (è buono, pessimo, discreto, accettabile, tollerabile, efficiente, efficace, ecc.) allo scopo di prendere decisioni sul suo sviluppo".

Questa definizione è però solo una categoria di valutazione, quella che è concepita e intrapresa per prendere una decisione. Per quanto rilevante possa essere per la politica la decisione, non è sempre necessario prefiggersi una decisione. Si può valutare anche per altri scopi e si può valutare anche alla cieca, senza scopi, per evitare di essere manipolati o influenzati dagli scopi. Nel mondo della valutazione ci sono molteplici approcci tra cui anche quelli determinati dalla preoccupazione della massima oggettività e neutralità possibili..Ma è bene precisare che nessuna valutazione è totalmente neutra o totalmente oggettiva.

Perché la valutazione si è diffusa a macchia d'olio nelle società contemporanee?

Le società contemporanee democratiche sono caratterizzate dalla libertà di scelta di determinate prestazioni sociali, di moltissimi servizi, di svariati prodotti. Per esempio, si possono scegliere per i viaggi o le vacanze le compagnie aeree o i servizi delle agenzie di viaggio che si preferiscono, si possono scegliere gli ospedali nei quali farsi curare, scegliere il tipo di automobile oppure la marca di televisione che aggradano, e via discorrendo. Queste scelte sono talora precedute da valutazioni esplicite, fatte dai consumatori o dai fornitori o da terzi. In ogni modo, la mancanza di scelta è sopportata male nelle società del mondo avanzato. La valutazione è proposta come un aiuto per fare scelte (House). Questa considerazione è però poco diffusa nel mondo scolastico dove la libertà di scelta è una rarità e non c'è accordo sul fatto che sia un bene o un male. Questo concorre probabilmente a spiegare perché la valutazione non attecchisce facilmente nei sistemi scolastici. Non ci sono scelte in una situazione di monopolio e quindi non servono aiuti alla scelta.

I sociologi francesi Duru-Bellat e Dubet fanno valere un altro argomento per giustificare la richiesta di valutazione: la perdita di fiducia nell'educazione stessa, nell'idea che è possibile cambiare gli uomini, educarli, che è stata accantonata progressivamente nel corso del ventesimo secolo ed è stata sostituita dalla richiesta di un'istituzione scolastica più equa e più utile. Non un sistema scolastico che educa ma un sistema scolastico equo e soprattutto utile.

La resistenza alla valutazione dei sistemi scolastici

Le scuole e una parte consistente di coloro che operano all'interno del servizio pubblico d'istruzione sono in generale centri di resistenza alla valutazione dei sistemi scolastici e alla loro valutazione. Non pochi insegnanti e pedagogisti si inventano argomenti a iosa per dimostrare che non si può valutare l'istruzione o quanto succede nelle scuole, quanto si insegna e si apprende. e per sostenere che la valutazione è perniciosa per l'istruzione. Per onestà occorre anche dire che le critiche alla valutazione non sono sempre infondate. La valutazione dei sistemi scolastici non è un assegno in bianco, ossia da prendere o lasciare in blocco.

La valutazione dei sistemi scolastici fatica ad attecchire : pochi governi e poche amministrazioni pubbliche sono attrezzate per valutare autonomamente i propri servizi scolastici. In genere ricorrono a organizzazioni internazionali o ad enti privati. Ogni valutazione è un'operazione delicata e lo è ancor più quando l'oggetto da valutare è il sistema scolastico.

Quando si valutano i sistemi scolastici?

Il clima politico ed economico contano molto nell'impostazione delle valutazioni dei sistemi scolastici su vasta scala, sia a livello internazionale che a livello nazionale. Le grandi valutazioni sistemiche non si sviluppano indipendentemente dal contesto socio-politico ed economico.

La valutazione dei sistemi scolastici in politica, può essere usata in modo strumentale sia quando tutto va bene, sia quando il malcontento dilaga nella società per una qualsiasi ragione, sia quando la società attraversa una profonda crisi economica. Queste tre circostanze offrono l'occasione per proporre e svolgere valutazioni dei sistemi scolastici e delle scuole. In un certo senso si può dire che le valutazioni dei sistemi scolastici sono strumenti polivalenti, utili in politica in circostanze tra loro diversissime, perché si prestano a finalità allettanti per chi detiene il potere. La conoscenza dei risultati dell'apprendimento non è che un pretesto.

Valutare è difficile

Quali criteri adottare per valutare? La diffusione delle pratiche valutative aumenta la domanda di valutatori e contribuisce alla proliferazione di agenzie di valutazione. Siccome la valutazione dei sistemi scolastici e delle scuole non può essere comparata alla valutazione di un prodotto in vendita sul mercato, occorre prestare grande attenzione alla professionalità dei valutatori. Valutatori incompetenti o disonesti fanno più danno che bene perché le valutazioni dei sistemi scolastici generano riforme a medio e lungo termine, gli effetti delle valutazioni sono a lunga scadenza, simisurano sull'arco di dieci o più anni. Non si modifica una scuola o un sistema scolastico in un battibaleno. Per questa ragione, la formazione dei valutatori dei sistemi scolastici deve essere rigorosa .

Non ci si imprrovisa specialisti competenti, onesti, con un alto senso di responsabilità, consapevoli, come dice House, del dovere di rispetto della verità, della giustizia e anche della bellezza della valutazione.

Quali gli effetti comprovati della valutazione dei sistemi scolastici?

E' assodato che la valutazione esterna ha un'incidenza sull'apprendimento, sul comportamento degli insegnanti, sulle reazioni delle scuole, sui curricoli. I risultati della valutazione quando sono chiari, spiegati bene, capiti (la comprensione non è automatica e la maggioranza degli insegnanti non è stata formata per capire il senso di molti dati), hanno un effetto, modificano la didattica,i curricoli, l'organizzazione e il funzionamento delle scuole, le decisioni delle autorità scolastiche e dei responsabili politici.

Occorre quindi porsi il problema dei costi delle valutazioni riguardo agli effetti. Una buona valutazione di sistema costa, richiede investimenti adeguati perché non si improvvisa. In genere occorrono un paio di anni per preparare una valutazione su larga scala e almeno un anno se non due per analizzare i risultati. Queste operazioni richiedono il coinvolgimento di specialisti che vanno pagati. Non è concepibile incaricare insegnanti di svolgere questi compiti. Non ci si improvvisa valutatori, anche se la conoscenza dei sistemi scolatici e delle scuole è una condizione necessaria per valutate un sistema scolastica. Una condizione necessaria ma non sufficiente.

Da ultimo un aspetto controverso anche per chi sostiene la necessità di una valutazione sistematica è l'opportunità di rendere pubblici i risultati. Poiché la pubblicazione dei risultati delle valutazioni scolastiche non lascia indifferenti nessuno, a partire dai genitori e dalle famiglie, la valutazione dei sistemi scolastici fa paura ed è osteggiata.

Questa ostilità è un pretesto per impedire qualsiasi valutazione esterna delle scuole o degli insegnanti. Molti ritengono che si inizia con la valutazione di sistema e si finirà inevitabilmente con la valutazione degli insegnanti.

Quindi anticipare queste conseguenze e impedire qualsiasi valutazione esterna.

Per queste ragioni i mandanti delle valutazioni scolastiche devono soppesare come usare i risultati delle valutazioni, avere idee chiare in materia, spegarle in anticipo per non cogliere di sorpresa i diretti interessati, "in primis" gli insegnanti o tutti coloro che forniscono le informazioni sulle scuole e sui sistemi scolastici oppure coloro che saranno valutati.

ERIC A. HANUSCHEK

Economic Impacts of Cognitive Skills

Abstract

While most analyses of growth and development emphasize the central role of human capital, measurement issues have plagued both research and policy development. Specifically, attention to school attainment and enrollment rates appears to be misdirected. In contrast, recent work has shown that the measures of cognitive skills that can be derived from international assessments greatly improve the ability to explain differences in economic growth rates across countries. Moreover, improved cognitive skills appear to have dramatic impacts on the future economic well-being of a country, suggesting that policy actions should focus directly on school quality and other means of improving cognitive skills. International assessments also permit investigation of educational institutions that promote higher achievement.

MARIA TERESA SINISCALCO

"Dieci anni di PISA: una foto aggiornata

In questo intervento si traccia un quadro dei principali risultati dell'Italia in PISA 2009. Questa rappresenta la quarta edizione dell'indagine e consente, per la prima volta, di operare un confronto tra due valutazioni incentrate sullo stesso ambito principale di competenza: la lettura.

Come nelle precedenti edizioni di PISA, nel 2009 il dato medio dell'Italia è più basso della media OCSE in tutti i tre ambiti di *literacy* valutati, anche se lo scarto viene ridimensionato o capovolto quando si tiene conto – nel confronto internazionale – dei fattori contestuali. Come già constatato nelle precedenti edizioni, inoltre, occorre andare oltre il dato medio, perché questo sintetizza gli esiti molto differenti delle diverse aree geografiche italiane. Proprio le disparità geografiche sono uno dei nodi problematici del nostro sistema scolastico, di non facile soluzione dal momento che ripetute analisi hanno evidenziato come le disparità geografiche siano in gran parte legate a disparità dei rispettivi contesti socio-economico-culturali delle scuole e del territorio.

Dal 2000 al 2009 i punteggi di lettura non sono cambiati (mentre sono aumentati, dal 2003 al 2009, quelli di matematica e, dal 2006 al 2009, quelli di scienze). Il dato di tendenza relativo alla lettura potrebbe però sottostimare il cambiamento, di segno positivo, dati i cambiamenti avvenuti nella popolazione scolastica.

L'Italia è, nel 2009, uno dei Paesi in cui la variabilità dei risultati è superiore alla media OCSE e, in particolare, le differenze dei risultati tra scuole sono particolarmente marcate e risultano ancora accresciute rispetto alle edizioni precedenti. Sulle elevate differenze tra scuole gioca il fatto che alle differenze tra tipi di istruzione, proprie dei sistemi in cui a 15 anni gli studenti sono canalizzati in indirizzi differenti, si sommano le differenze tra aree geografiche. Queste differenze non appaiono, invece, dai voti scolastici che prendono come punto di riferimento la distribuzione "locale" delle prestazioni.

PISA 2009 conferma l'impatto comparativamente ridotto delle origini familiari sui risultati dei singoli, che PISA utilizza come indicatore dell'equità dei sistemi. Il dato positivo, per il nostro sistema, è che la relazione tra background e risultati rimane debole anche con il crescere dei risultati complessivi,nelle aree del Nord. Un altro dato che conferma la relativa equità della scuola italiana è la percentuale di studenti resilienti, che è superiore alla media OCSE. Tuttavia gli studenti risultano distribuiti nei diversi indirizzi della scuola secondaria di II grado in buona parte in relazione al background e il background medio della scuola ha una relazione più forte con i risultati, che non il background dei singoli studenti all'interno di ciascuna scuola.

Le analisi multilivello condotte a livello nazionale e regionale sui dati delle diverse edizioni di PISA hanno evidenziato alcuni fattori, a livello di studenti e di scuole, associati con differenze significative nei risultati, anche se occorre ricordare la difficoltà di individuare fattori a livello di scuola, il cui effetto non sia "assorbito" dalla composizione – in termini di capacità e di origini sociali – del corpo studentesco, quando gli studenti sono iscritti a canali di istruzione differenti.

PIERO CIPOLLONE

"Gli esiti delle indagini e del politiche educative "

In attuazione del proprio mandato istituzionale sopra delineato l'INVALSI ha realizzato per l'a.s. 2009-10 la rilevazione degli apprendimenti nelle classi II e V della scuola primaria e nella classe I della scuola secondaria di primo grado, oltre che per la terza classe nell'ambito della Prova Nazionale. La rilevazione è stata censuaria e ha riguardato tutti gli studenti delle classi II e V della scuola primaria e della classe I (e III classe nell'ambito della prova nazionale) della secondaria di primo grado. Complessivamente sono state coinvolte circa 13.700 Istituzioni scolastiche, 115.120 classi (87.800 al netto della terza media) e circa 2,3 milioni di studenti (1, 7 senza la terza media) (cfr. tavola1).

Tavola 1: la rilevazione degli apprendimenti nell'anno 2009-10, scuole classi studenti.

Italia	Popolazione			Campione		
Livello	Numero scuole	Numero classi	Numero studenti	Numero scuole	Numero classi	Numero studenti
II primaria	7770	30175	555347	1385	2000	39299
V primaria	7770	30476	565064	1385	2000	39643
I secondaria di I grado	5895	27210	595732	1309	2095	46447
III secondaria di I grado	5896	27259	574652	1309	1309	28626
Totale	13666	115120	2290795	2694	7404	154015

Per fornire alle scuole in tempi brevi una informazione di ritorno sugli esiti complessivi delle rilevazioni, in modo da disporre di valori di riferimento (benchmark) con cui confrontare i propri risultati, è stato individuato un campione di scuole statisticamente rappresentativo, i cui risultati costituiscono la base dei risultati che l'INVALSI diffonde nei suoi rapporti. Il campionamento è stato effettuato su base regionale, coinvolgendo complessivamente, per i quattro livelli di classe, 2.694 scuole, 7404 classi e 154.015 studenti (cfr. Tavola 1). In ogni classe campionata la rilevazione si è

svolta con la presenza di un osservatore esterno (eccetto che nella prova nazionale dove la vigilanza e' affidata alla commissione d'esame).

Il dato centrale che dovrebbe guidare la definizione delle politiche scolastiche e' la grande dispersione dei risultati tra scuole anche all'interno della stessa area geografica. Questa variabilità dei risultati tra scuole in presenza di un sistema fortemente centralizzato di assegnazione delle risorse indica che le difficoltà della scuola italiana vanno affrontate a partire da quelle della singola unità scolastica. Il sistema di rilevazione degli apprendimenti con test standardizzati e' lo strumento indispensabile per costruire un sistema di valutazione che aiuti le scuole a raggiungere il proprio potenziale.

Come e' noto gli strumenti che l'INVALSI usa nella rilevazione degli apprendimenti degli studenti sono test standardizzati basati su una metodologia consolidata a livello internazione e trasparente verso le scuole e i ricercatori che vogliono utilizzare i dati. Sul sito INVALSI compaiono infatti non solo i rapporti con i risultati emersi, ma anche i quadri di riferimento delle singole prove, le griglie di correzione, una dettagliata descrizione dei singoli item con indicazione del riferimento ai quadri concettuali e tutti i dati elementari, cioè le risposte ai singoli quesiti fornite da ogni studente, naturalmente nel rispetto di tutte le normative sulla privacy a protezione degli alunni e delle scuole. La vera sfida è ora proprio che questo enorme patrimonio informativo venga utilizzato dalle scuole per il miglioramento della propria efficienza educativa e dai ricercatori che si occupano di istruzione per produrre conoscenza sul sistema scolastico del nostro paese. L'attenzione massima è naturalmente riservata alle scuole come dimostra il fatto che grande impegno e' dedicato alla restituzione dei risultati delle rilevazioni disaggregati per singola domanda per tutte le circa 115000 classi coinvolte. Per facilitare la lettura e la comparazione dei dati e' stato predisposto un set molto ampio di grafici (fino a 50 grafici per scuola) accompagnato da una analitica guida alla lettura per facilitarne l'interpretazione.

Queste informazioni cosi dettagliate permettono alle scuole che le vogliano utilizzare di individuare con precisione i punti di debolezza nella preparazione dei propri alunni e di programmare una didattica coerente. Molte scuole si stanno orientando in questa direzione e in alcune aree del paese si stanno programmando interventi di sostegno alle scuole proprio sulla base delle risultanze dalla

misurazione degli apprendimenti. In molte realtà i risultati delle rilevazioni vengono utilizzati per quello che sono, uno tra i vari strumenti di diagnostica sullo stato degli apprendimenti degli allievi di una scuola. La specificità dei dati dell'INVALSI è la loro comparabilità tra scuole, caratteristica che manca invece alle rilevazioni adottate dentro le singole scuole

Questa dialettica, programmazione didattica-misurazione dei risultati-nuovo ciclo di programmazione, e' la spirale ascendente lungo la quale muoversi per accrescere il livello degli apprendimenti dei ragazzi. Questa visione laica e prammatica della misurazione degli apprendimenti si va progressivamente diffondendo nelle scuole e costituisce la base per uno sviluppo sano e non autoreferenziale dell'autovalutazione delle scuole. Le scuole hanno bisogno di conoscere quanto più possibile della propria situazione relativa rispetto a quella degli altri istituti e perché questa conoscenza e' un motore importante di cambiamento, grazie al suo utilizzo da parte dei dirigenti e degli insegnanti.

LUCIANO ABBURRA'

"Dal macro al micro, e ritorno. L'utilizzo delle indagini internazionali per comparazioni su base regionale: dall'esperienza del Piemonte suggerimenti e proposte"

L'esperienza del Piemonte può essere utile per esemplificare i modi e le ragioni per cui un'indagine internazionale sulle competenze fondamentali degli adolescenti può essere assunta come utile strumento di valutazione anche a scala territoriale regionale, senza che ciò diventi un semplice esercizio di trasposizione d'interesse localistico. Anzi, si può sostenere che, per paesi molto differenziati come l'Italia, solo da analisi articolate in chiave regionale è possibile ricavare una rappresentazione realistica della situazione nazionale. E una rappresentazione realistica è una condizione essenziale perché lo strumento che la produce risulti credibile come misuratore dello stato effettivo delle competenze e delle loro variazioni fra ed entro i territori, senza perdere la capacità di indicare vie e possibilità di miglioramento per tutte le parti del sistema, sia quelle più avanzate sia quelle più in difficoltà. Nella relazione si richiameranno processi e motivazioni che hanno portato il Piemonte a muoversi per primo per chiedere all'INVALSI un "regionalizzazione" della rilevazione PISA. Verranno poste in luce le ragioni e le conseguenze "locali" di una tale scelta, emersa come risposta ad un esplicito problema di "valutazione" dello stato e delle performance del sistema regionale alla luce dei convergenti processi di decentralizzazione e di autonomia delle istituzione educative.

Ma verranno ugualmente poste in luce le ricadute conoscitive sul livello nazionale che l'estendersi progressivo delle analisi regionali ha prodotto, con effetti benefici in primo luogo sulla capacità di comprensione dell'oggetto, ma anche sul livello di interesse e di attivazione sociale e politica che alle indagini hanno fatto seguito, per non dire delle competenze professionali nel campo delle analisi valutative che l'esperienza ha contribuito a produrre e a consolidare a scala ampia. Il problema è ora quello di come valorizzare e integrare - in un nuovo sistema di valutazione nazionale saldamente articolato su una organizzazione decentralizzata a scala regionale - energie, conoscenze, competenze ed attese alimentate da un programma come il PISA.

MELANIA RUDIN

"Risultati PISA e successo negli studi postsecondari: l'indagine longitudinale in Svizzera"

TRE

TRansition from **E**ducation to **E**mployment

Project design and results

Convegno internazionale "La sfida della valutazione"

Torino, 24 - 25 maggio 2011

Fondazione per la Scuola
Compagnia di San Paolo

TREE's Roots

"The opportunity should not be missed to equip Switzerland with a longitudinal survey of transitions at national level (...) Transition pathways to employment are becoming increasingly complex. To understand young people's decisions and options, and to take them into account in policy decisions, appropriate analytical instruments are needed."

OECD Thematic Review on Transitions from Initial Education to Work (TIEW)
Country Note on Switzerland, 1999, p. 53

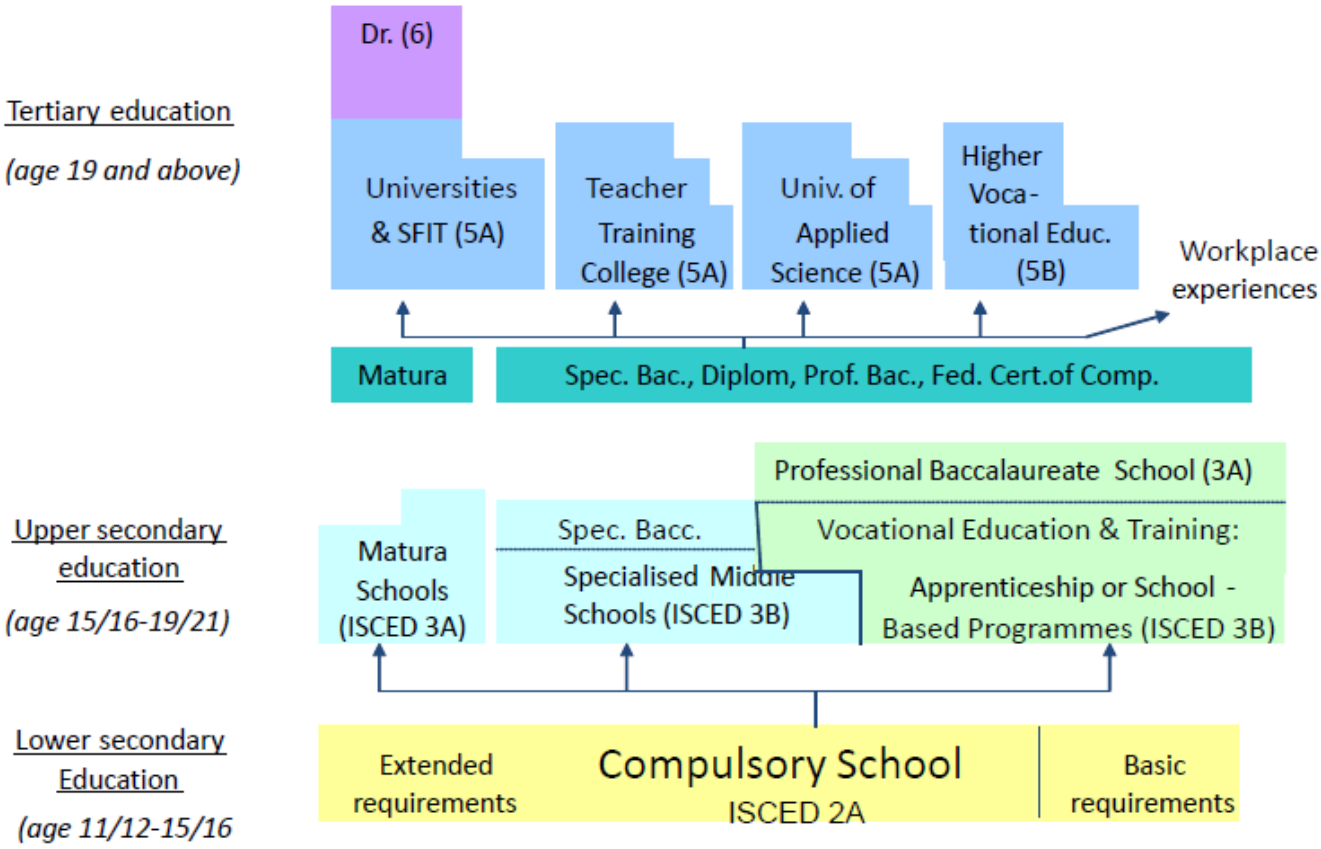
Fondazione per la Scuola
Compagnia di San Paolo

State of Research in Switzerland, before TREE start (late 1990s)

Existing research	TREE
cross sectional	prospective, longitudinal
cantonal / regional	national
particular / „risk group“	entire school leavers' cohort
disciplinary view	multidisciplinary view

Fondazione per la Scuola
Compagnia di San Paolo

The Swiss Educational System



Convegno internazionale "La sfida della valutazione"

Torino, 24 - 25 maggio 2011

Fondazione per la Scuola
Compagnia di San Paolo

The TREE Project: Context and Objectives

TREE is the first and so far only longitudinal study in Switzerland to comprehensively survey the postcompulsory educational pathways of young people and their entry into the labour market

Based on a sample of approximately 6,000 youths who participated in the PISA survey in 2000 and left compulsory school the same year.

Provides public use files

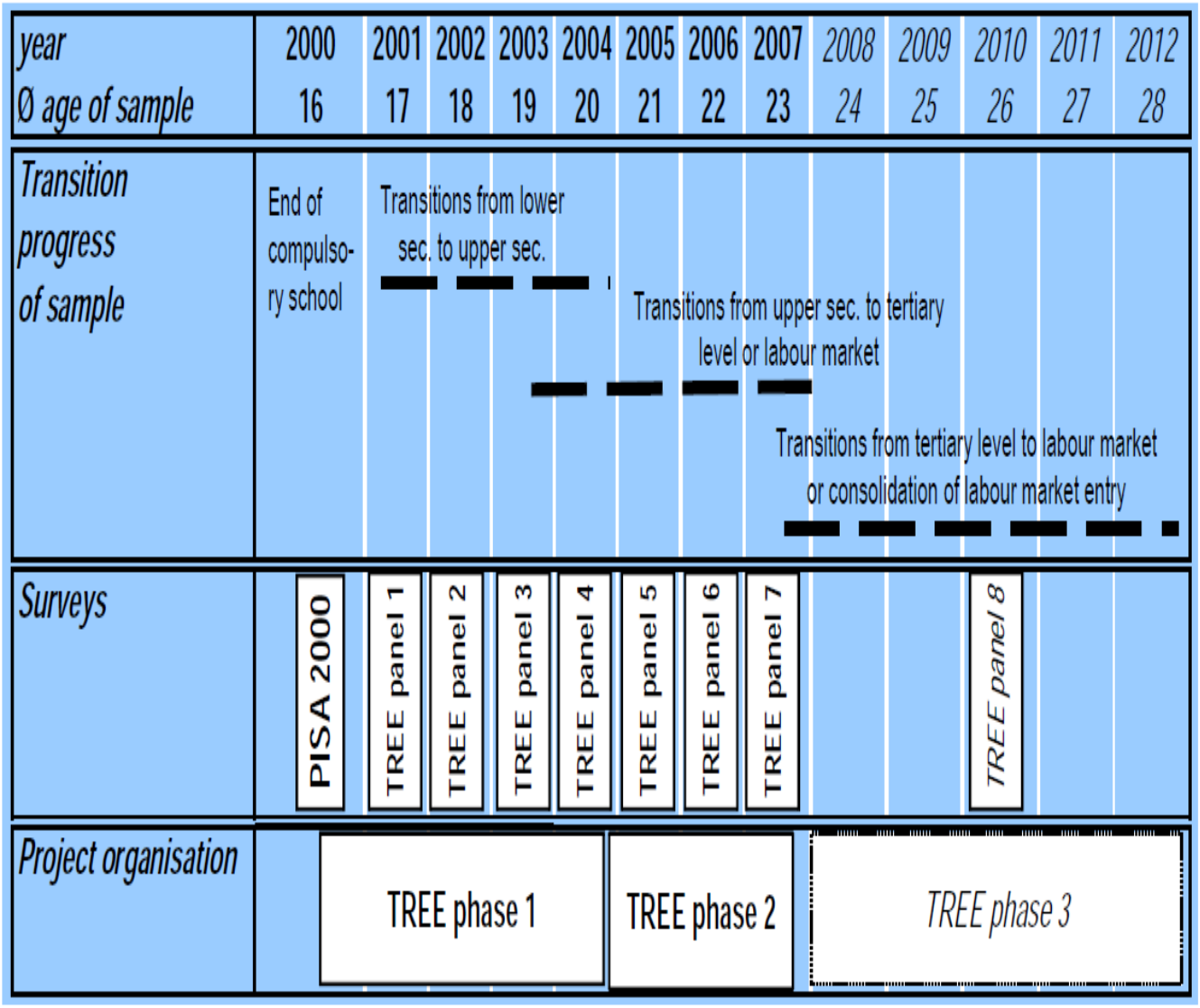
Serving both fundamental & applied research interests

Main focus on reasons for typical trajectories as well as consequences of irregular or critical educational careers

Particular focus on premature dropout

Fondazione per la Scuola
Compagnia di San Paolo

Survey design



Response rates

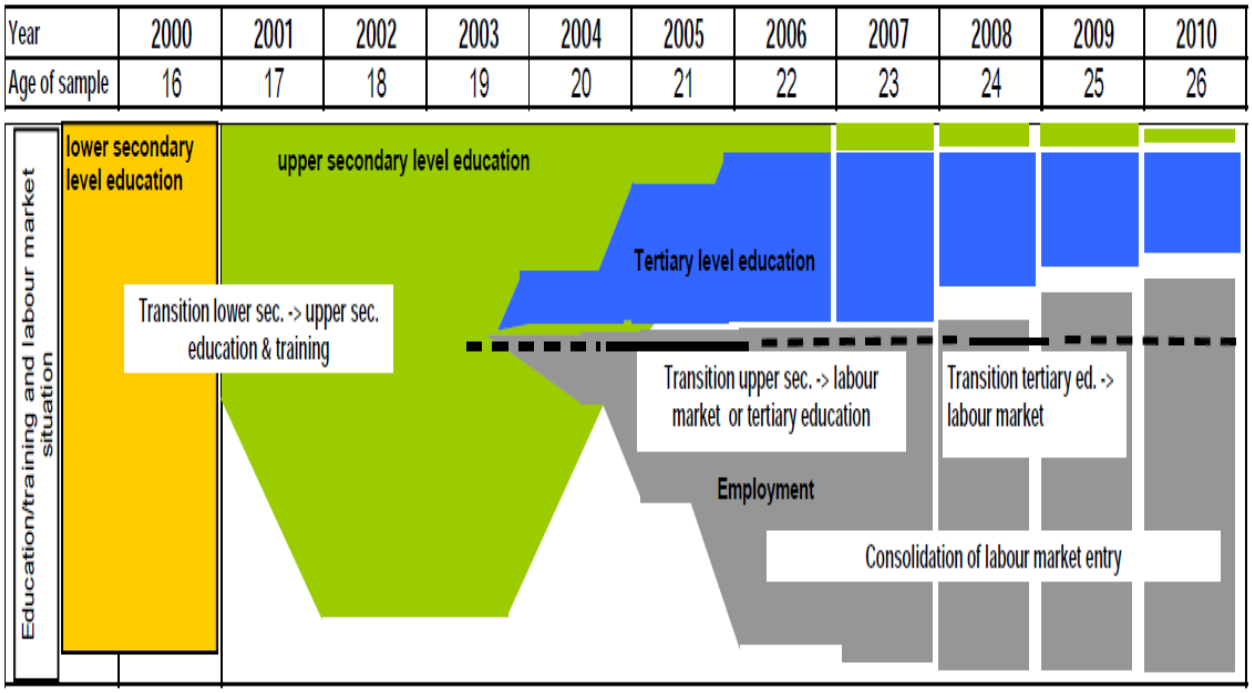
Survey year	*2000*	*2001*	*2002*	*2003*	*2004*	*2005*	*2006*	*2007*	
Gross sample	* 11,710	6,343	5,944	5,609	5,345	5,060	4,852	4,659	***20***
Responses (N)	** 6,343	5,528	5,206	4,877	4,679	4,506	4,133	3,979	***10*** 4,571
Return rate									3,418
% of wave	54%	87%	88%	87%	88%	89%	85%	85%	
% of 2001 survey			82%	77%	74%	71%	65%	63%	75%

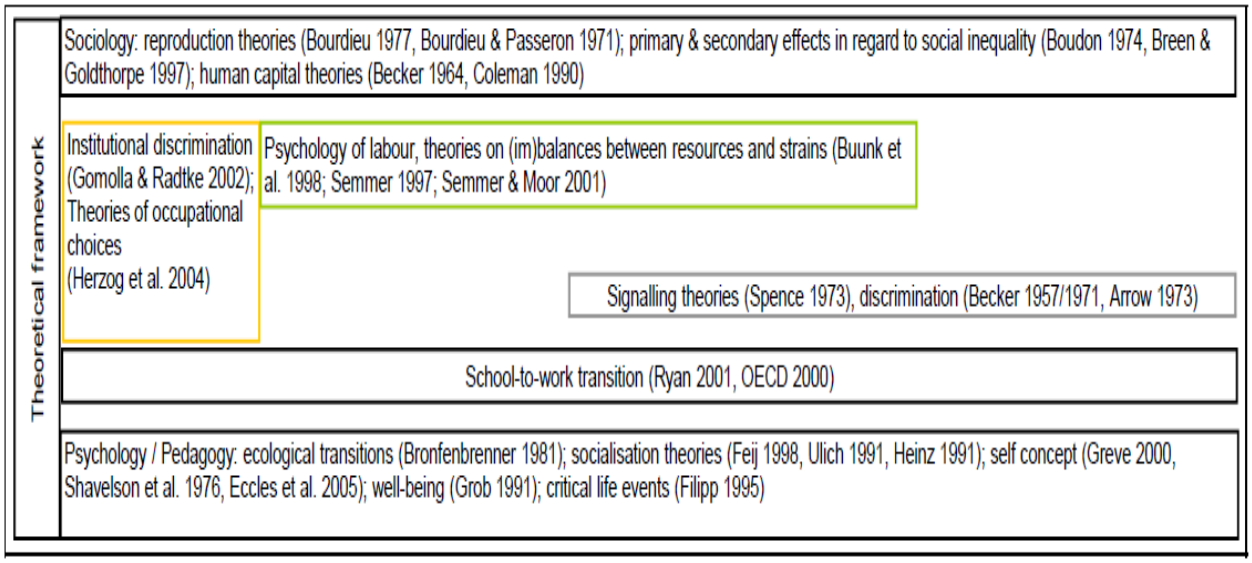
* Sample base ** Address survey: willingness to participate in TREE panel survey

41%

Various measures to prevent sample attrition (Alternative response channels, Hotlines, address research,
newsletter tailored to the target group informing about new results)
Panel weights to compensate for sample bias

Theoretical framework(s)





Fondazione per la Scuola
Compagnia di San Paolo

Used concepts and scales (1)

Detailed description of educational and labour market situation and pathways

Strains and resources in education and work

Social support in schools and VET training firms

Gender and ethnic (in)equality

Competences of trainers, teachers and supervisors

Job search and application strategies

Psychological/personality scales (coping, persistence)

Educational aspirations and plans

Fondazione per la Scuola
Compagnia di San Paolo

Used concepts and scales (2)

Fluctuation/dropout tendency

Achievement and academic success

Income, financial situation

Job requirements

Critical life events

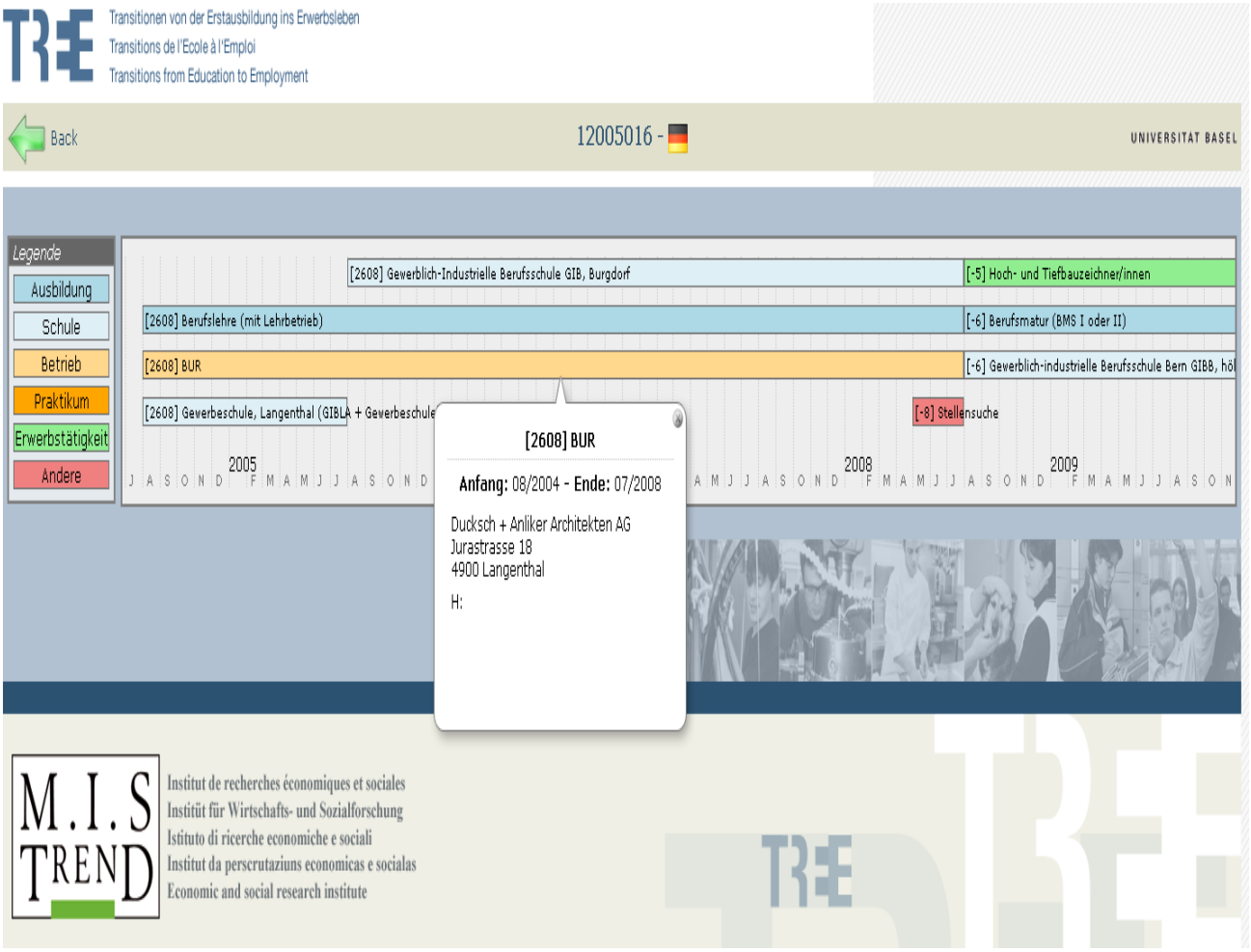
Health behaviour and wellbeing

Drug abuse

Family situation /status

Fondazione per la Scuola
Compagnia di San Paolo

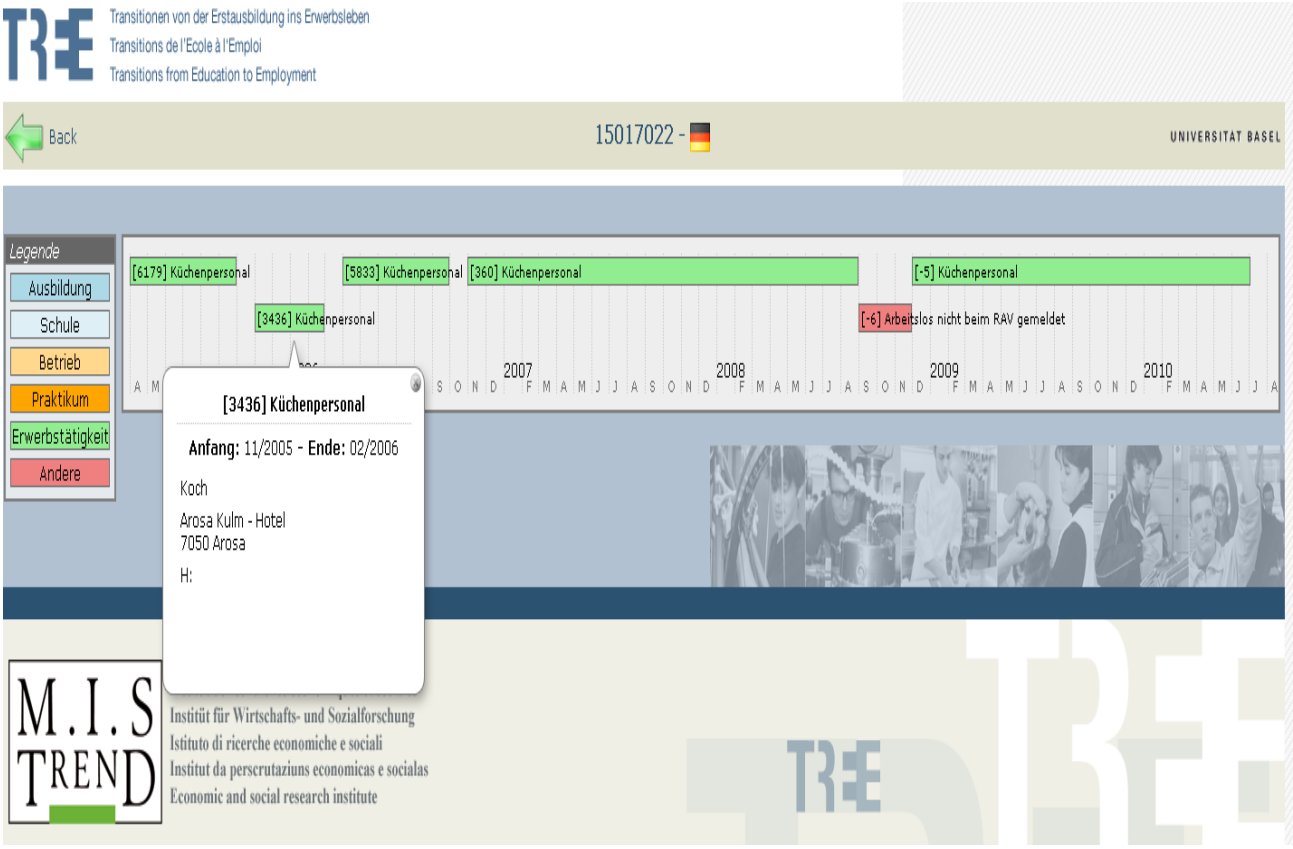
screenshot of visualisation tool for CATI 2010 (1)



Challenges
- many parallel or intersecting episodes
- Capacity of remembering former episodes

Fondazione per la Scuola
Compagnia di San Paolo

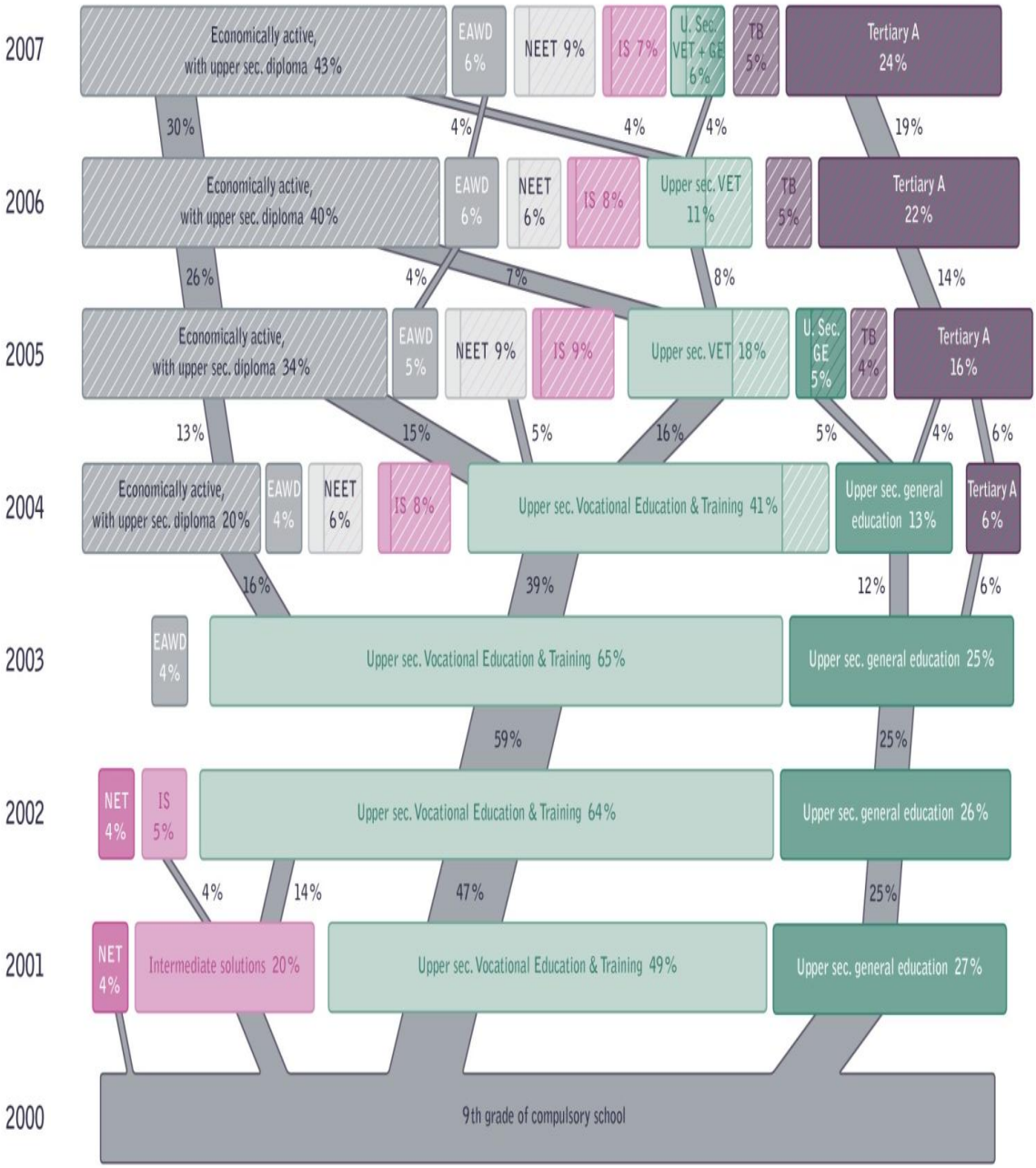
screenshot of visualisation tool for CATI 2010 (2)



Convegno internazionale "La sfida della valutazione"

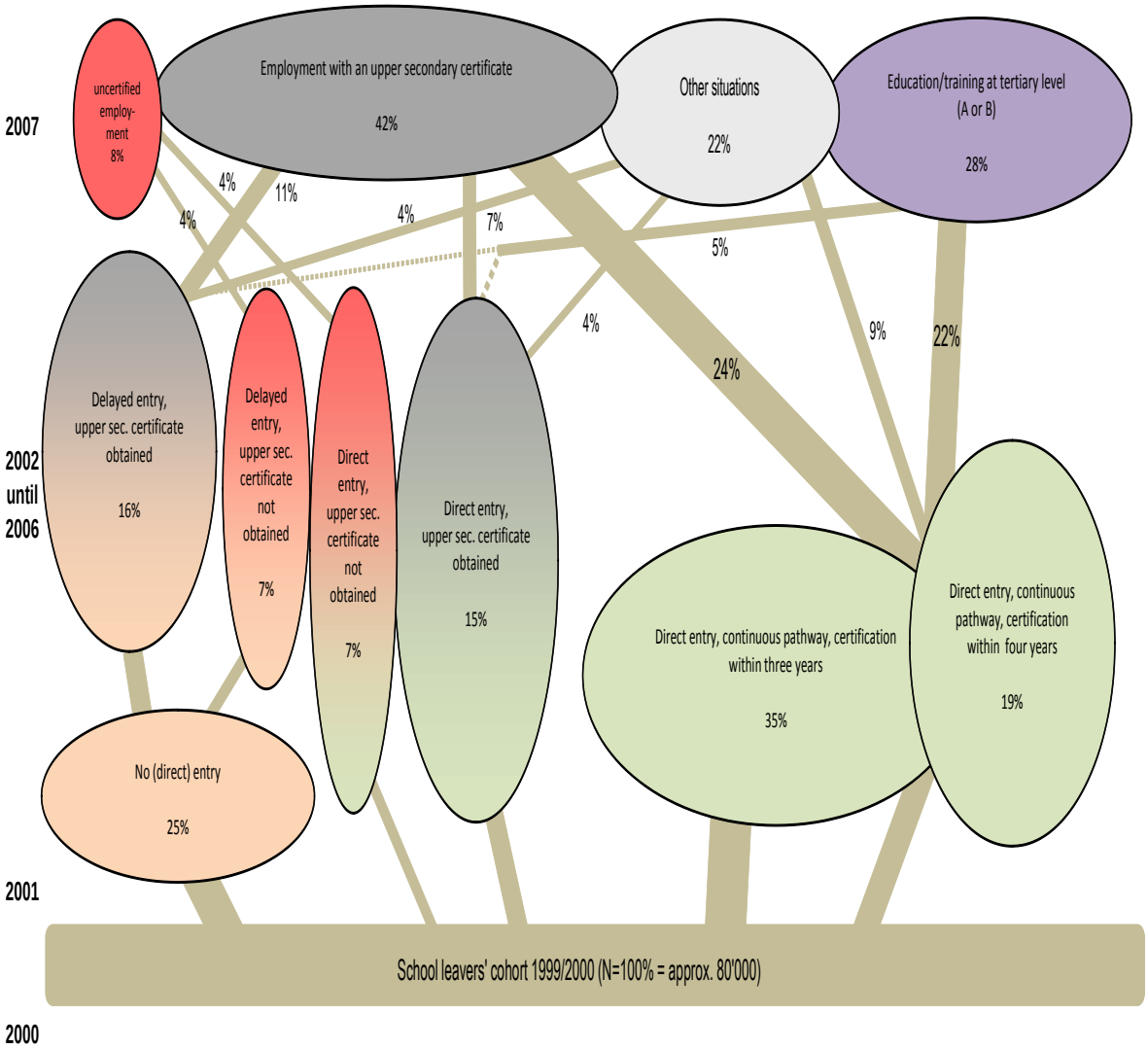
Torino, 24 - 25 maggio 2011

Results – Education and employment pathways



Keller, A.; Hupka-Brunner, S.; Meyer, T (2010). Post-compulsory education pathways in Switzerland: The first seven years

Results – Discontinuous Pathways

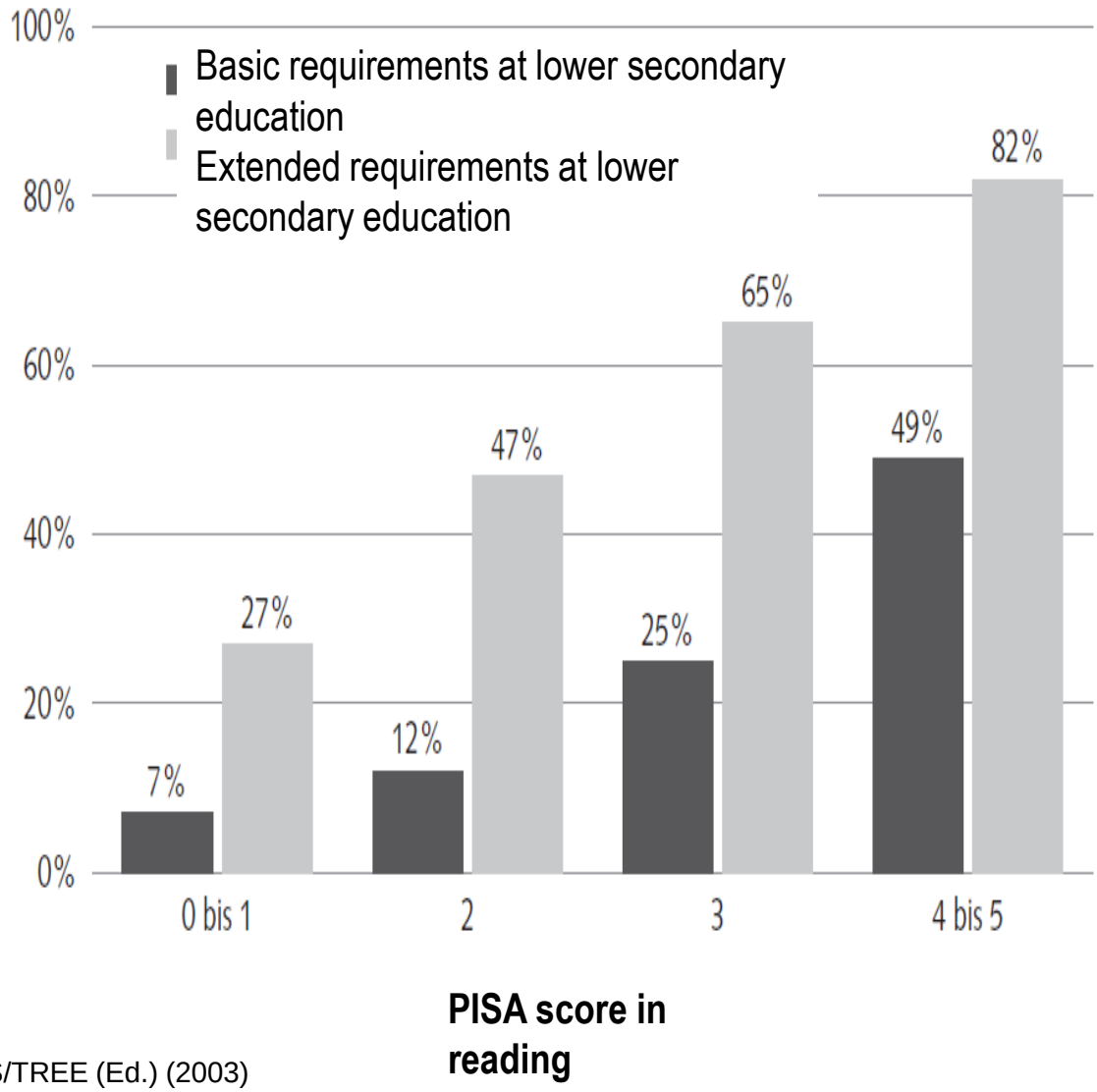


Keller, A.; Hupka-Brunner, S.; Meyer, T (2010). Post-compulsory education pathways in Switzerland: The first seven years

Fondazione per la Scuola
Compagnia di San Paolo

PISA and the transition into post-compulsory education

Percentage of young adults entering upper secondary education with high requirements



OFS/TREE (Ed.) (2003)

Fondazione per la Scuola
Compagnia di San Paolo

PISA and the transition into the labour market

	(1)	(2)	(3)	(4)	(5)
PISA score in reading/10	0.028 (0.011)**	0.025 (0.011)**	0.017 (0.013)	0.01 (0.013)	−0.001 (0.015)
Vocational training final score		0.286 (0.23)		0.43 (0.233)*	0.438 (0.248)*
Higher intellectual level			0.366 (0.197)*	0.45 (0.200)**	0.556 (0.199)***
Full-time vocational school					−0.42 (0.257)*
Upper-level school track * female					0.369 (0.174)**
Lower-level school track * female					−0.097 (0.291)
Lower-level school track * male					0.387 (0.336)
Older than 16					−0.070 (0.210)
Socio-economic index					0.129 (0.093)
Foreign language					−0.019 (0.401)
Immigrant					−0.067 (0.378)
Countryside					0.087 (0.17)
French speaking region					−0.357 (0.247)
Italian speaking region					−0.135 (0.324)
Unemployment Rate					0.088 (0.115)
Constant	−0.679 (0.58)	−1.895 (1.158)	−0.331 (0.614)	−2.08 (1.177)*	−1.629 (1.358)
$\bar{x}'\beta$	0.872	0.986	0.901	0.756	−0.16
Observations	642	642	642	642	642

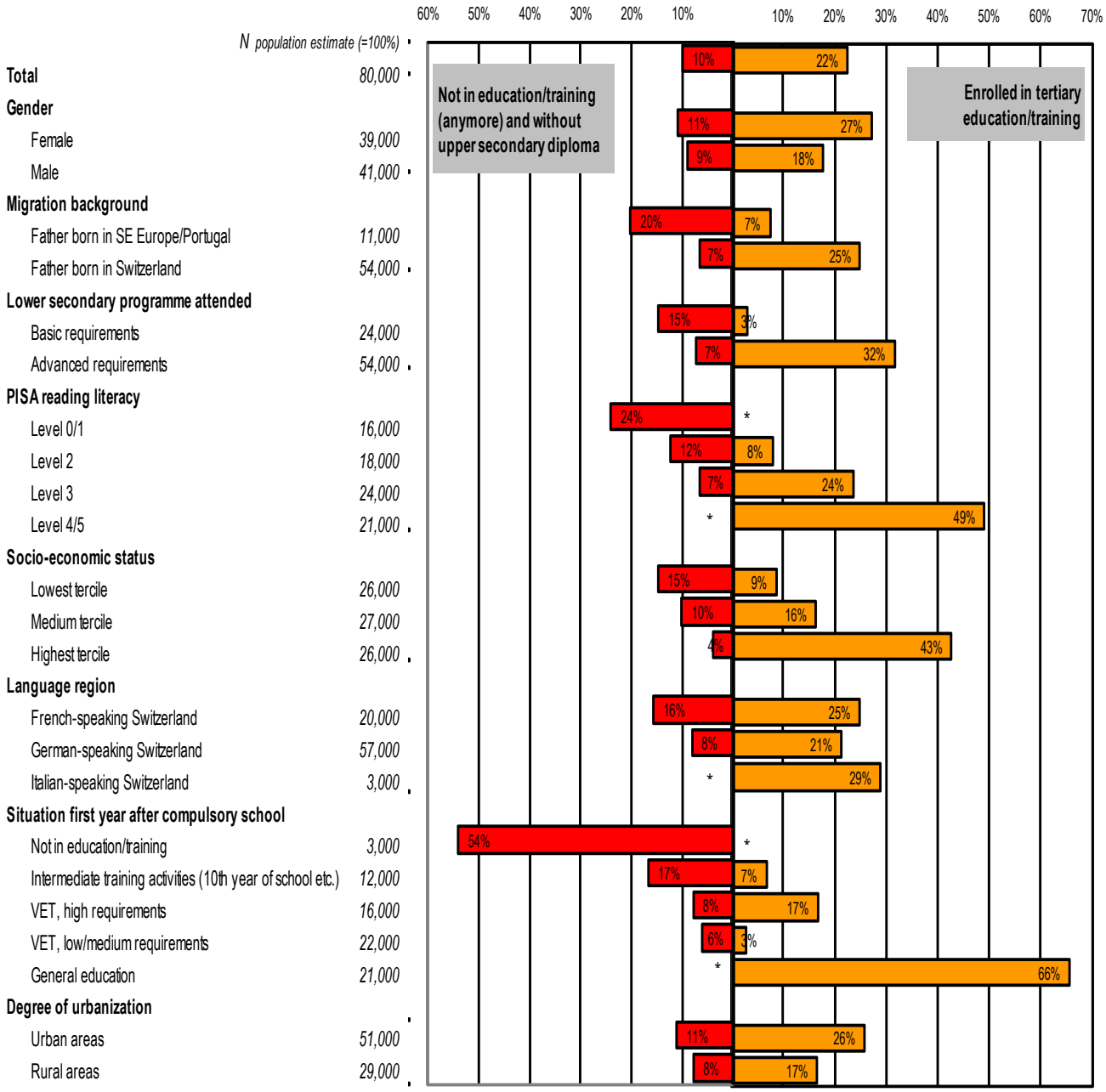
Probit Model of the probability of finding an adequate job

→ indirect effect of PISA results on labour market transitions through type of vocational education

Bertschy et al. (2009)

Fondazione per la Scuola
Compagnia di San Paolo

Non-completion/ enrolment in tertiary education, 2006

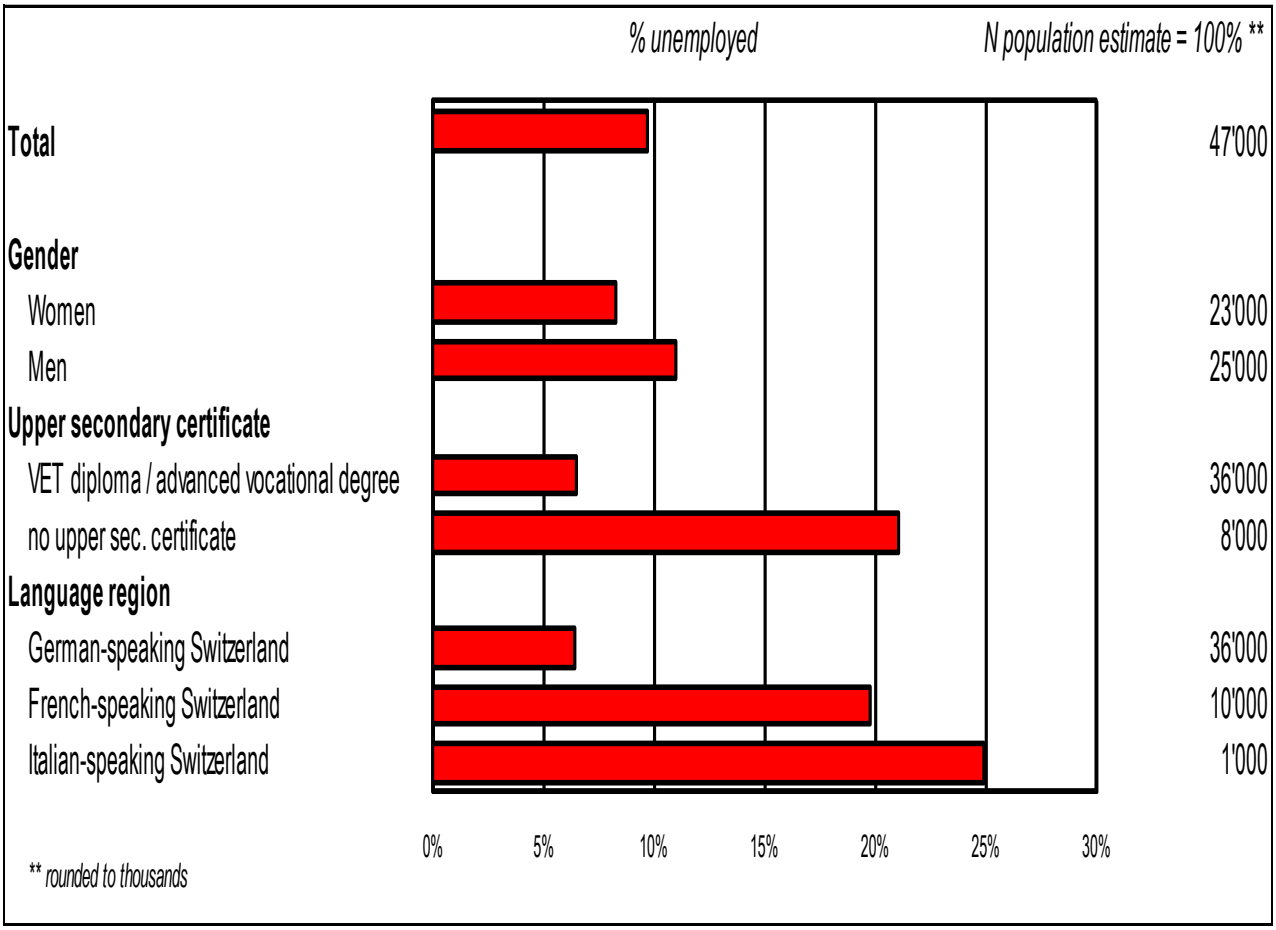


Bertschy, K.; Böni, E.; Meyer, T. (2007)

Convegno internazionale "La sfida della valutazione"

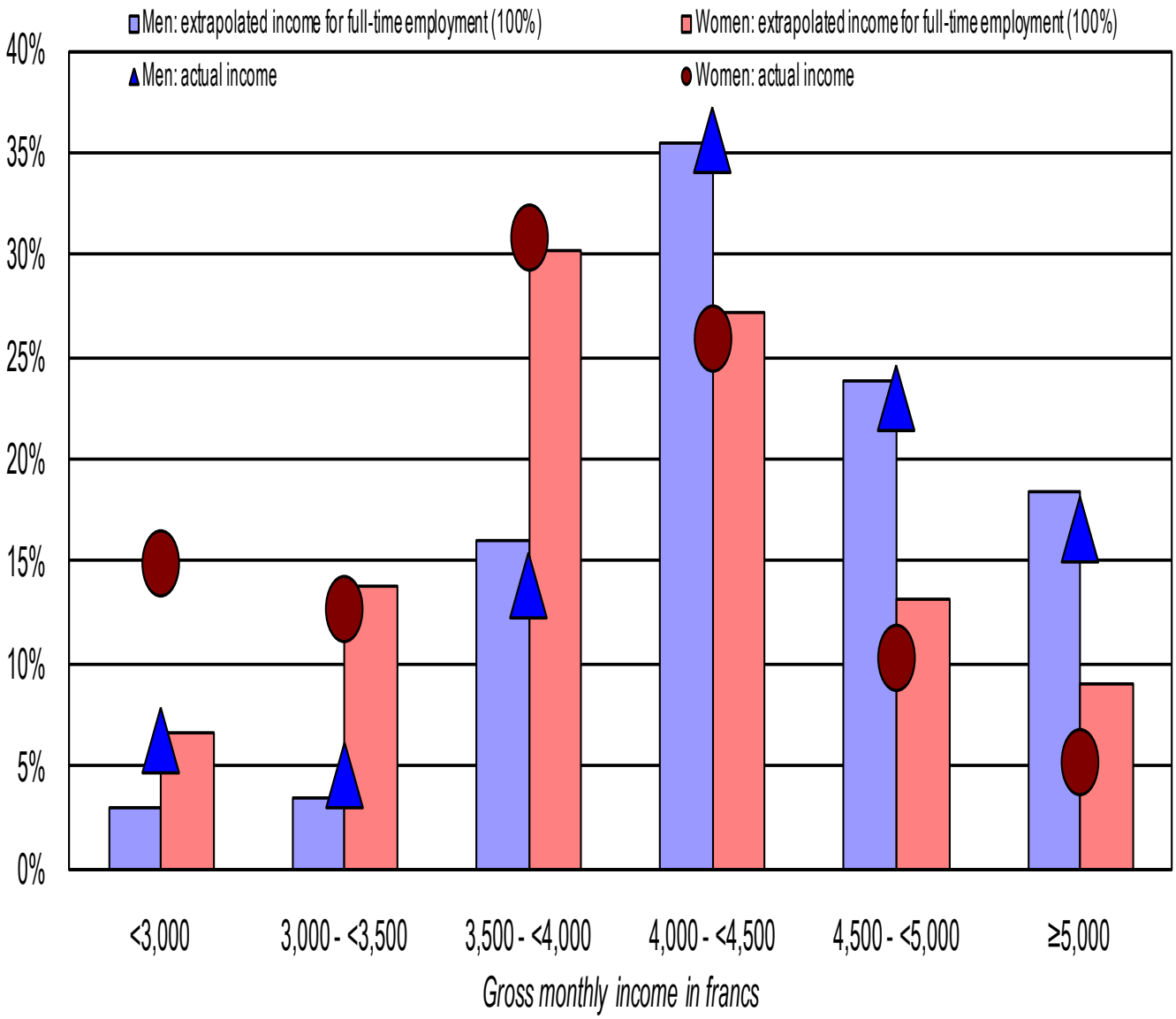
Torino, 24 - 25 maggio 2011

Unemployment by selected characteristics, 2006



Bertschy, K.; Böni, E.; Meyer, T. (2007)

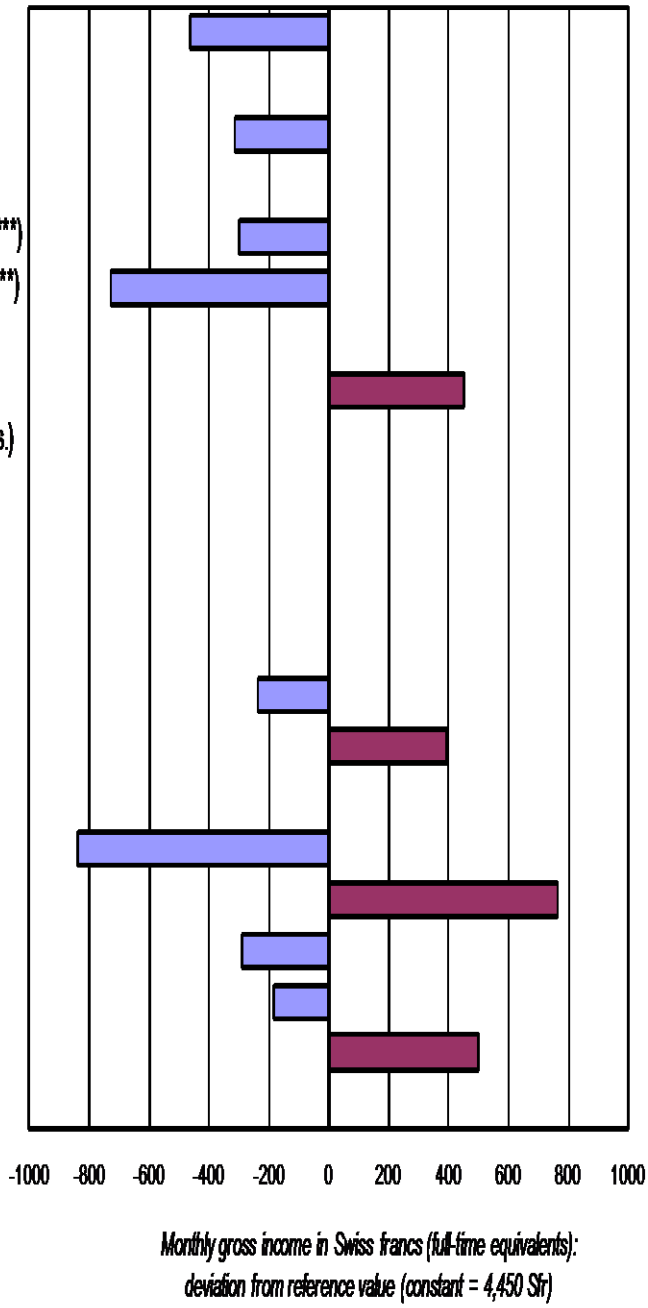
Income distribution by gender, 2006



Bertschy, K.; Böni, E.; Meyer, T. (2007)

Linear regression analysis on income

Characteristic	*Category of reference*	*Category of comparison*
Gender	Male	Female (****)
Upper sec. diploma	VET diploma	No upper sec. certificate (**)
Language region	German-speaking Switzerland	French-speaking Switzerland (***)
		Italian-speaking Switzerland (****)
Occupation	All other occupations	Clerk (****)
		Vocational business school (n.s.)
		Retail (n.s.)
		Electrician (n.s.)
		Pharmacy assistant (n.s.)
		Car mechanic (n.s.)
		Shop assistant (*)
		Medical assistant (***)
		Carpenter (n.s.)
		Hairdresser (****)
		IT specialist (***)
		Cook (**)
		General mechanic (*)
		Automation technician (***)
		Dental assistant (n.s.)



Example of how to read the graph: After controlling for all factors considered in the model, in terms of gross monthly income,
- a woman earns nearly 500 francs less than a man;
- a person from French-speaking Switzerland approximately 300 francs less than a person living in German-speaking Switzerland;
- an IT specialist over 700 francs more than the average income in other occupations.

Levels of statistical significance:
**** <0.001
*** .001 - <.01
** .01 - <.05
* .05 - <.1

Bertschy, K.; Böni, E.; Meyer, T. (2007)

Research foci & results

1st threshold“

Substantial dysfunctions at interface between lower & upper sec. in Switzerland (largely due to quantitative and qualitative rationing of educational „supply“)

Tremendous equity problem in Switzerland regarding access to and graduating from post-compulsory education & training (by gender, migration background and social origin mostly)

„2nd threshold“

Generally good labour market absorption and job-skills match of upper sec. VET graduates, even for low achievers

Strong „genderisation“ of education/transition pathways

Substantial gender-related wage discrimination from the very beginning of professional careers

Fondazione per la Scuola
Compagnia di San Paolo

Research areas to be developed

Transition dimensions beyond education and work (family founding, parenthood, etc.)

Analysis of episodes, particularly regarding dynamics and sustainability of labour market entry and careers

Risks/opportunities of (dis-)continuous pathways

Working with the TREE data:

Data of first 7 waves (2001-2007) available, 8th wave (2010) shall be available by end of 2011

>50 ongoing user contracts for public use files in 6 countries

Extended documentation in English, German and French

Individual support on demand

further information: www.tree.unibas.ch

Literature

Bergman, M., Hupka-Brunner, S., Keller, A., Meyer, T. & Stalder, B. E. (2011). Youth Transitions in Switzerland: Results from the TREE Panel Study. TREE. Zurich: Seismo.

Bertschy, K., Böni, E. & Meyer, T (2007). Young people in transition from education to labour market. Results of the Swiss youth panel survey TREE, update 2007. BernTREE.

Bertschy, K., Cattaneo, M. A., Wolter S. C. 2009: PISA and the transition into the labour market. Labour. Review of Labour Economics and Industrial Relations 23, Special Issue. 111-137.

CDPE (website). http://www.cdip.ch/dyn/16237.php . download: 13.4.2011. Bern: Conferenza svizzera dei direttori cantonali della pubblica educazione.

Keller, A.; Hupka-Brunner, S.; Meyer, T. (2010): Post-compulsory education pathways in Switzerland: The first seven years. Results of the Swiss youth panel survey TREE, update 2010. Basel: TREE.

OECD. (1999). Thematic Review of The Transition From Initial Education to Working Life. Switzerland. Country Note: OECD.

OFS/TREE (Ed.). (2003). Parcours vers les formations postobligatoires. Les deux premières années après l'école obligatoire. Résultats intermédiaires de l'étude longitudinale TREE. Série "Monitorage de l'éducation en Suisse". Neuchâtel: Office fédéral de la statistique.

TREE (2010). Project documentation TREE 2000-2007. Basel: TREE.

DANIELE CHECCHI

"Le gambe del sistema: misurazione, valutazione e supporto formativo.
Dove stiamo andando e chi paga il costo del sistema"

1. L'Italia sconta tra svantaggi rispetto ad altri paesi di area OECD. Innanzitutto non possiede la cultura del dato amministrativo come sistema di monitoraggio; la registrazione amministrativa avviene ancora su base cartacea in buona parte dei casi (basti pensare ai registri e/o alla verbalizzazione dei voti); parcellizzazione dei software (basta girare per le segreterie scolastiche e/o universitarie); le banche dati non comunicano e non sono interrogabili dagli amministratori che dovrebbero voler monitorare i processi che gestiscono. Secondo, siamo figli di una cultura idealistica di stampo crociano, che considera la dimensione quantitativa meno significante di quella qualitativa, e che rifugge quindi da operazioni intellettuali quali misurare o stimare, tanto più se applicate alla sfera della cultura. Per queste ragioni la cultura docimologica non ha mai attecchito in Italia, rimanendo confinata alla stregua di moda filo-statunitense. Da ultimo, la pubblica amministrazione è sempre stata tenuta al riparo di ogni tentativo di *accountability*, in nome di uno scambio perverso tra bassa retribuzione e bassa produttività (per definizione da non misurare).

2. Per queste ragioni l'Italia si presenta all'inizio del nuovo millennio totalmente impreparata alla sfida della valutazione dei sistemi scolastici. La maggior parte degli istituti di ricerca educativa (IRRE) fa *ricerca azione*, ovvero analisi descrittiva di casi studio, l'INVALSI sotto la presidenza Vertecchi produce ricerca sulle pratiche didattiche, l'indagine PISA 2000 passa sostanzialmente ignorata. Il ministro Berlinguer è costretto a dimettersi sull'ipotesi dei corsi-concorsi ad accesso limitato, ma nessuno propone di misurare l'efficacia didattica degli insegnanti. Simmetricamente, sociologi e pedagogisti sviluppano analisi qualitativi. Il cambio a livello ministeriale avviene con il Ministero dell'Istruzione Moratti e con la Commissione Muraro per il controllo della spesa pubblica.

Esce il Quaderno Bianco sulla scuola ad opera del Ministero dell'Economia. A sinistra si grida contro l'aziendalizzazione della scuola, a destra si gettano le basi per i tagli tremontiani.

3. Nessuno sembra condividere l'idea che l'informazione quantitativa sul funzionamento dei sistemi scolastici e universitari, fornita con il massimo livello di disaggregazione possibile, sia un bene pubblico. Come tutti i beni pubblici, non rivali e non escludibili, essi non vengono prodotti attraverso meccanismi di mercato perché nessuno è disposto a spendere per acquistare un bene di cui potrebbe fruire gratuitamente (qualora ci fosse qualcun altro che a sua volta lo acquistasse). Il progetto Almalaurea/Almadiploma è un buon esempio per illustrare questo paradosso. Conoscere la probabilità d'impiego associata ad un titolo di studio (diploma di maturità, laurea o laurea magistrale) è una informazione importante per un vasto spettro di attori: gli studenti che devono scegliere dove iscriversi; le amministrazioni scolastiche e universitarie che devono distribuire le risorse; le imprese che desiderano conoscere il grado di scarsità relativa dei diversi titoli di studio. Produrre questa informazione per via amministrativa è ancora oggi percepito come futuribile, ma tecnicamente facile da realizzare: data una anagrafe degli studenti che sia aggiornata sulle scelte scolastiche e che comunichi in tempo reale con le direzioni regionali per l'impiego (che ricevono le comunicazioni obbligatorie di avvio/cessazione degli impieghi), si potrebbe produrre il tasso d'impiego per titolo di studio in tempo reale, a qualunque scadenza temporale, con un semplice click. Appunto perché futuribile, lasciamo questa idea da parte e speriamo nelle scuole e nelle università: non sarebbe difficile avviare un monitoraggio periodico, telefonico o informatico, dei *past alumni*, le università estere lo fanno regolarmente, in Italia si invoca la mancanza di risorse finalizzate. Si lascia allora il campo ad altri: all'Istat, che encomiabilmente ogni 3 anni intervista diplomati e laureati di tre anni prima, con un campione rappresentativo su base regionale e/o di grande ateneo, ma che non permette di scendere ad un livello di dettaglio inferiore, perché la struttura è campionaria. Oppure al consorzio Almalaurea, che vende ai potenziali utilizzatori dell'informazione il dato stesso, facendosi nel contempo pagare la rilevazione. Le famiglie e gli studenti ne restano esclusi, perché non riescono a sapere se la facoltà che hanno in mente permetterà di trovare lavoro.

4. Negli ultimi cinque anni, Invalsi ha dato un importante contributo a sostegno dell'idea che la misurazione dei comportamenti scolastici è un bene collettivo da diffondere al massimo grado. Ha cercato di rendere universali le proprie rilevazioni, al costo di affiancarsi all'esame che conclude la scuola secondaria di primo grado.

Ha messo in luce l'importanza della rilevazione esterna per evitare i comportamenti opportunistici degli insegnanti (vedi correzioni ex-post applicate alle rilevazioni). Ha reso disponibili i propri dati alla comunità scientifica, come misura di validazione delle rilevazioni condotte. Rispetto al progetto complessivo di una rilevazione universale e obbligatoria in quattro punti della carriera scolastica (II e V primaria, III secondaria I grado, III secondaria di II grado) si è però dovuto piegare di fronte alla carenza dei finanziamenti ministeriali, ed è stato costretto a ripiegare sulla costruzione di un campione rappresentativo (chi non ricorda la distinzione introdotta da Fioroni tra valutare e monitorare ?).

5. Se la rilevazione non è universale, è impossibile individuare le aree di forza e di debolezza del sistema scolastico. Certo con una rilevazione campionaria possiamo sfruttare la variabilità temporale per monitorare i trend, ma certamente non possiamo tenere sotto controllo la variabilità territoriale, che è uno dei mali che affliggono il sistema scolastico italiano. In particolare non è possibile individuare quali siano le scuole in difficoltà su cui concentrare possibili interventi di sostegno.

6. In un ormai dimenticato contributo ("Un sistema di misurazione degli apprendimenti per la valutazione delle scuole: finalità e aspetti metodologici - Proposta preparata per l'INVALSI" curata dal sottoscritto insieme a Andrea Ichino (Università degli Studi di Bologna) e Giorgio Vittadini (Università degli Studi di Milano-Bicocca) – dicembre 2008) si era insistito sulla indispensabilità del rendere immediatamente operativa una anagrafe degli studenti, in cui riportare gli esiti individuali della prove.

Questo permette di misurare gli apprendimenti non in termini di livello ma di variazioni, eliminando in buona parte il problema delle differenze nelle condizioni di partenza.

Nello stesso contributo si forniva anche una stima dei costi di un sistema di rilevazione degli apprendimenti a regime. La stima dei costi si basava sulla considerazione di tre elementi:

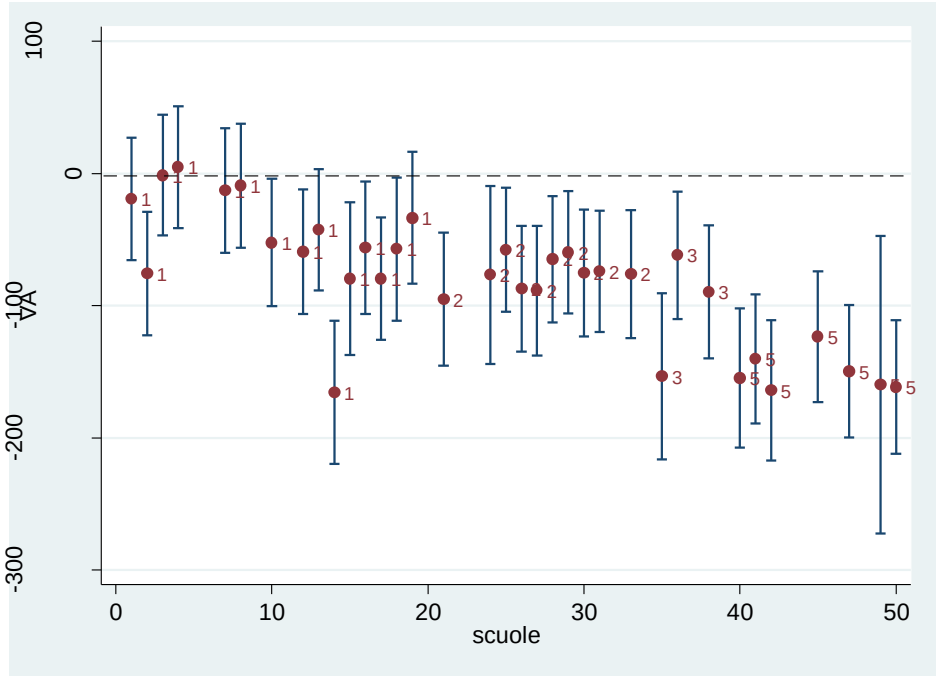
1) la preparazione dei materiali, la loro stampa, spedizione e raccolta;
2) la somministrazione;
3) la correzione delle prove.

La stima complessiva del costo del sistema nazionale di valutazione come prospettato in quel documento è riportata nella tavola seguente. I costi variano da 31 milioni di euro nell'ipotesi di costo dei somministratori di 250 euro e nessuna domanda a risposta aperta a 81 milioni nel caso di somministratori a 300 euro tutte le domande a risposta aperta e correttori pagati 2500 euro per mese-persona. Il Ministero ha messo meno di un decimo dell'ipotesi più ottimistica. Ma cosa sono questi costi rispetto ai risparmi conseguiti con il taglio al corpo insegnanti ?

	II primaria		V primaria		III secondaria I°		II secondaria II°		V secondaria II°		totale		
	costo per studente	num. studenti	costo per studente	num. studenti	costo per studente	num. studenti	costo per studente	num. studenti	costo per studente	num. studenti	costo per studente	num. studenti	spesa totale
1) materiali stampa e spedizione	2.0	580.000	2.0	580.000	2.0	570.000	2.0	570.000	2.0	470.000	2.0	2.770.000	5.540.000
2) somministrazione													
Ipotesi a: 250 euro a somministratore	10.1	580.000	10.1	580.000	8.8	570.000	8.8	570.000	9.1	470.000	9.4	2.770.000	26.050.000
Ipotesi b: 300 euro a somministratore	12.2	580.000	12.2	580.000	10.5	570.000	10.5	570.000	11.0	470.000	11.3	2.770.000	31.260.000
3) correzione													
nessuna domanda a risposta aperta	0.0	580.000	0.0	580.000	0.0	570.000	0.0	570.000	0.0	470.000	0.0	2.770.000	0
16 domande a risposta aperta	6.4	580.000	6.4	580.000	6.4	570.000	6.4	570.000	6.4	470.000	6.4	2.770.000	17.728.000
40 domande a risposta aperta	16.0	580.000	16.0	580.000	16.0	570.000	16.0	570.000	16.0	470.000	16.0	2.770.000	44.320.000

7. Questo è il costo di una informazione di buona qualità. Difficile stimare l'equivalente economico dei benefici che ne potrebbero derivare. Quanto vale un miglior *match* tra qualità della scuola scelta e capacità del singolo studente ? Quanto vale la riduzione dell'*over-education* ? Più facili (e potenzialmente più misurabili) i benefici legati al miglioramento della prestazione lavorativa degli insegnanti. Nel momento in cui almeno una dimensione del risultato diventa conoscibile in modo oggettivo, vengono a cadere molte delle scusanti, vere o fittizie che siano. La concorrenza implicita che può scatenarsi tra le scuole, nel momento in cui i risultati della misurazione degli apprendimenti vengono resi pubblici (seppur con le cautela del misurarli in termini di variazione degli apprendimenti) può essere un potente incentivo al miglioramento.

8. Ma c'è davvero da innovare ? La figura seguente mostra la misurazione del valore aggiunto nelle scuole secondarie di Trento che hanno ripetuto il test PISA nel 2010 ad un anno di distanza dalla rilevazione ufficiale. Se si tiene conto dell'intervallo di confidenza sono davvero poche le scuole che sono "oggettivamente" meglio o peggio di altre. Questo non è un caso isolato. Ogni esercizio valutativo che produca una graduatoria tra scuole e/o operatori deve essere cosciente che si tratta di confronti tra medie, che esiste una dispersione intorno alle medie e che esiste un intervallo di confidenza per le misure fornite. Per questo diventa altrettanto difficile poter associare premi e punizioni a queste graduatorie.



Nota. La figura riporta le stime del VA di scuola in lettura nel 2010 (pallino centrale) e i relativi intervalli di confidenza ottenuti col LCVA (*longitudinal contextual value added*) model. 1=liceo, 2=ist. tecnico, 3=ist. professionale, 5=formazione professionale. Il modello LCVA include come controlli: sesso, struttura familiare, status di immigrato, livello più elevato di istruzione dei genitori (HISCED) e indicatore di prestigio occupazionale più elevato tra i genitori (HISEI).

9. Ma il buon funzionamento di una scuola non si riduce alla rilevazione di un buon livello di apprendimento attraverso test uniformi sul territorio nazionale. Buon funzionamento può voler dire promozione della convivenza, collaborazione con gli enti locali, recupero del disagio giovanile, oculata gestione dei finanziamenti pubblici, partecipazione a progetti nazionali o internazionali, organizzazione di reti tra scuole, innovazione nei metodi didattici, formazione del corpo insegnante. Tutto questo è poco standardizzabile, e ancor meno quantificabile in modo oggettivo. Le rilevazione degli apprendimenti è la punta di un iceberg, cui correlano molte delle dimensioni precedentemente elencate. E i nessi di causalità non sono così evidenti: quando una scuola va male, di chi è la colpa ? del basso livello sociale di provenienza degli alunni ? della trascurata gestione di un dirigente scolastico, magari appena giunto nella sede ? della bassa motivazione degli insegnanti ? del disinteresse di genitori e amministratori locali ?

10. I numeri senza modelli interpretativi non servono a nulla. Invalsi ha (di nuovo meritoriamente) promosso lo studio di un modello interpretativo (Valsis), ma la sua applicazione richiede buona capacità di discernimento delle cause. Si è ipotizzato che una rivalorizzazione del ruolo degli ispettori possa supportare questo approccio. Ma il singolo ispettore, lasciato senza linee guida di indirizzo, può rivelarsi poco efficace nell'individuare cause e nel proporre rimedi. Il Ministero (o al massimo le Direzioni Regionali) non può esimersi dall'esprimere linee di indirizzo a cui deve corrispondere l'esercizio valutativo. Non si possono lasciare alle singole scuole la scelta di un proprio *benchmarking* (come accade invece implicitamente oggi quando si parla di autovalutazione). Il Ministero deve fornire obiettivi di riferimento, in modo del tutto analogo a quanto viene scritto negli obiettivi di Lisbona 2020 (soglia massima nell'analfabetismo funzionale, soglia minima di conseguimento di titoli di studio secondari). Solo così la valutazione non si rivela sterile esercizio di concorrenza.

11. Il Ministero ripone molta fiducia alla formazione sugli insegnanti come leva di miglioramento per le situazioni deficitarie. Personalmente credo che ci sia più da ottenere intervenendo sull'organizzazione interna. Agli insegnanti deve essere lasciata fiducia di innovare la pratica educativa, specialmente quando essi vengono resi responsabili delle conseguenze della propria progettualità.

Una parte del corpo docente non è pronta a questo saldo, e la macchina organizzativa statale lo è ancor meno. In questo la formazione può cambiare la cultura del fare scuola. Ma non mi sembra che manchino le idee. Manca piuttosto l'auto-riconoscimento e l'auto-valorizzazione, dopo decenni di mancato riconoscimento del contributo individuale e di forzato appiattimento retributivo.

OLAF KÖLLER

"Educational Standards in Germany: Conceptualization and Assessment"

Abstract

The disappointing findings of the Third International Mathematics and Science Study (TIMSS) published in the Nineteen Nineties (e.g., Beaton, Martin, Mullis et al., 1996) led to an intense scientific and public discussion about educational goals in the German school system and about reforms necessary to improve the academic achievement of our students in mathematics and science. As a consequence, the 16 federal states with the German Federal Ministry of Education and Research established a program to improve teaching quality in science and mathematics (so-called SINUS-Program, cf. Fischer, Rieck, & Prenzel, 2010). Furthermore, they decided to participate in OECD's PISA. The findings of PISA 2000, however, again revealed relatively low competencies of German students and led to a second wave of reforms in Germany. One substantial change was the introduction of educational standards for different subjects at the end of primary school (grade 4) and at the end of lower secondary school in 2003 and 2004. These standards define core learning outcomes for seven subjects (biology, chemistry, English, French, German, mathematics and physics). In order to test whether German primary and secondary school students reach these standards and to establish a German system of national assessment, the Institute for Educational Progress (IQB) was founded in 2004. Since that time large item pools have been generated providing hundreds of calibrated items for different subjects. In addition, important grant programs by the German Science Foundation (DFG) and the Federal Ministry of Education and Research (BMBF) have strongly supported research programs on competencies. Research questions in these initiatives refer not only to the assessment of competencies but also address competence structures, their development, their domain-specificity, and their relationship to other constructs like intelligence, social background etc. This paper provides an overview of educational standards in Germany, their assessment, and current national research activities on competencies of primary and secondary students in school.

1. Outcome Orientation, Standards and Competencies in the German School System

As a consequence of the international large-scale assessments like PIRLS, PISA, and TIMSS, many countries across the world have introduced new concepts of educational governance based on the idea that a focus on students' cognitive outcomes in core subjects (mathematics, mother tongue, science) is more effective than input orientation in improving the quality of national educational systems. Consequently, several national and international assessment programs have been initiated, the number of participating countries in international studies on students' academic achievement has steadily increased, and educational standards have been established that define students' learning outcomes in school subjects as central goals of primary and secondary schooling. Note that, recently of 50 North American states 48 have agreed on so-called *Common Core State Standards* for mathematics, English language arts and literacy in history/social studies, science, and technical subjects (Common Core State Standards Initiative, 2010). It is also noteworthy that PISA started in 2000 with 31 participating countries and by 2009 this number has increased to 71 (see Klieme, Artelt, Hartig et al., 2010).

Until 2006, the cognitive framework in PISA has proposed literacy in reading, mathematics and science as core learning outcomes of schools in industrial countries (OECD, 2006).National educational goals often rely on the idea of content and performance standards and several countries have recently linked these standards to theoretical frameworks on competencies. Interestingly, OECD has also moved to the concept of competencies (OECD, 2009) in the PISA-2009 framework. Such competencies are often conceptualized in the sense of the German psychologist Weinert (2001, p.47), who described specialized cognitive competencies as "clusters of cognitive prerequisites that must be available for an individual to perform well in a particular content area". In this definition a competence is a combination of skills, knowledge, characteristics, and traits that contribute to performance in a particular domain. Similar to the development of expertise, building up competencies requires long-term learning and learning opportunities in which students gain deep understanding of the domain.

1.1 Standards in Germany

In 2003 and 2004, based on an expertise by Klieme, Avenarius, Blum et al. (2003) and recommendations by experts in teacher education, psychology, and educational research, the Standing Conference of Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany (KMK) agreed on common standards for primary and secondary schools. These standards were partly based on Weinert's (2001) conceptual work of competencies, and referred to subject-specific educational traditions regarding goals of teaching. They describe, in a broader sense, domain-specific cognitive student outcomes at a certain grade level for a certain school-leaving qualification. Although German standards are based on theoretical frameworks of international studies like PIRLS, TIMSS, and PISA, and take national traditions of the subjects into account, it has been criticized that much more research is necessary to justify the underlying competency models. However, it has been argued elsewhere (cf. Köller, 2008a) that linking such standards to theory, educational models, and empirical research must be part of future research, which can help to modify such models and educational standards. Furthermore Köller (2008a) argues that at least standards in mathematics, foreign languages, and to some extent science, are well grounded in theory. The standards in foreign languages are based on the Common European Framework of References for Languages (CEFR, e.g., Council of Europe, 2001). Mathematics standards are rooted in the work of the National Council of Teachers of Mathematics (NCTM, 2000), the PISA framework (e.g., OECD, 2003), Freudenthal's "Realistic Mathematics Education" (Freudenthal, 1983), and a publication by Winter (1995). Finally, science standards are strongly connected to the large body of literature on scientific literacy (e.g., OECD, 2006, Bybee, 2008; Waddington, Nentwig & Schanze, 2008) and the German tradition of science education in secondary schools (see, for example the expertise on core curricula for upper secondary school, edited by Tenorth, 2004).

Table 1 provides an overview of subjects and grade levels for which standards have been developed in Germany. Missing bullets in the cells indicate that no standards have been published yet for the corresponding subject at a specific grade level. For the end of upper secondary school, standards are currently (March 2010) being developed and will probably be available in 2012.

Contrary to curricula that substantially differ among German states, standards define the same goals for all 16 federal states. In primary school (grade 4) however, they are limited to German and mathematics. Primary school standards for foreign languages were excluded for several reasons:

- Curricula in English and French are very different in the 16 federal states.
- Some states start teaching a foreign language in grade 1 while others start in grade 3.
- There is no consensus among states about common textbooks that are suitable for primary education in foreign languages.
- There is no existing tradition of primary teacher education in foreign languages.

Furthermore the KMK did not publish any science standards for primary education. Reasons here might had been that no common curricula across our federal states exists, that lesson tables in primary school differ among states and that science instruction is usually embedded in a subject called general studies (in German: "Sachkunde") which is a mixture of civics education, geography, physics, and biology.

Concerning lower secondary school, standards exist for the end of grade 9 and 10. Grade 9 standards define outcomes for students, who strive for the lowest certificate of secondary education (LCSE; in German: "Hauptschulabschluss"), whereas standards at the end of grade 10 describe goals for the more prestigious middle school certificate (MSC; in German: "Mittlerer Schulabschluss"). Science standards, however, only exist for the MSC and differentiate outcomes for biology, chemistry, and physics. The paper by Schecker and Parchmann (2007) provides a good insight into the German science standards. However, it is noteworthy that both science and mathematics standards in Germany are a mixture of so-called content standards and performance standards, although the idea of linking standards to competencies clearly favors the concept of performance standards.

Table 1. Subjects and grade levels for which educational standards have been developed in Germany*

	Grade level			
	Primary school	End of lower secondary school		End of upper secondary school[1]
Subject	Grade 4	Grade 9	Grade 10	Grades 12/13
German (mother tongue)	●	●	●	●
Mathematics	●	●	●	●
Foreign Language				
English		●	●	●
French			●	●
Science				
Biology			●	●
Chemistry			●	●
Physics			●	●

Note: *Missing bullets indicate that no standards are available for the corresponding subject; [1]Standards for the end of upper secondary school will be published in 2012.

2. Research on Competencies

Many research initiatives have been started following the publication of the PISA findings and the German educational standards. Both the German Science Foundation (DFG) and the Federal Ministry of Education and Research (BMBF) have initiated large grant programs to support empirical studies on

- structures of students competencies,
- assessment and measurement models of competencies,
- development of competencies, and
- definition of proficiency levels

In 2007 the DFG started a priority program on competence structure models and assessment of competencies. Many findings of the program have recently been published (Klieme, Leutner, & Kenk, 2010). Furthermore, this research has been extended to teachers' professional competencies (e.g. Krauss, Brunner, Kunter et al., 2008) and higher education competencies, particularly for teacher education. The following paragraphs report on these different research activities.

2.1 Research on Competence Structures

A first step in assessing standards requires the development and testing of scientifically grounded structure models describing the different facets of a competence. As an example for such models, figure 1 displays different facets of mathematics competence in primary and secondary school (cf. Köller, 2010). Contents are combined with processes supporting the argument above that mathematics standards represent a mixture of content and performance standards. As already mentioned, the model is based in the literature. It remains, however, an open question whether this holds true empirically. Therefore, research is needed to construct measurement models and then test the structure on the basis of large student and item samples (for an overview, see Hartig, Klieme, & Leutner, 2008). Here multidimensional IRT models (e.g., Adams, Wilson, & Wang, 1997) represent adequate procedures to test such structural models, given that they are not too complex.

Figure 1. Theoretical structure of mathematics competence in the German primary and secondary school standards (KMK, 2003, 2004)

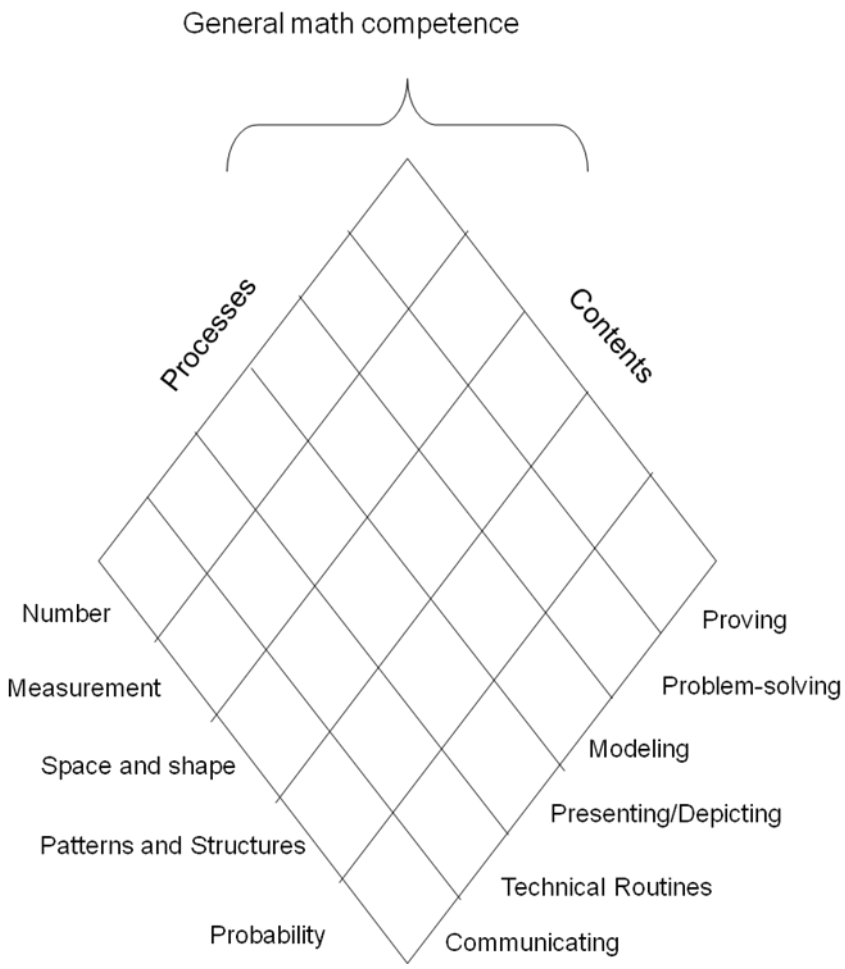


Figure 1 suggests 11 distinct dimensions of mathematics competence, which clearly indicate a high level of complexity. Winkelmann and Robitzsch (2009) have thus argued in a recent publication, that testing the model is only possible in two steps, first testing whether the different contents can be distinguished, and second whether the processes are separable in a multidimensional IRT-analysis. In a large sample of primary school students from grades 3 and 4, the authors found convincing

evidence for the validity of the model, i.e., the content dimensions as well as the process dimensions were strongly correlated but all correlation coefficients were substantially below 1, indicating multidimensionality.

Furthermore, the intercorrelations suggested that collapsing all sub-scores to a total score in order to get a measure of general math competency is appropriate and particularly sufficient for large-scale assessments. Similar analyses have not yet been conducted for the science standards but findings from the German extension of the PISA 2003 study (Rost, Carstensen, Walter et al. 2003) provide evidence that similar results can be expected. Rost, Carstensen, Walter et al. analyzed a national science test which consisted of items measuring seven different facets of cognitive processes (decision-making, divergent reasoning, using graphical representations, convergent reasoning, mental models, verbalizing, and data interpretation). Statistical analyses revealed substantial correlations among the facets ranging from .58 to .88, but all coefficients were below 1. The findings both for mathematics and science thus provide evidence that *a priori* structures hold true in empirical tests but there is also empirical support for a general competence factor.

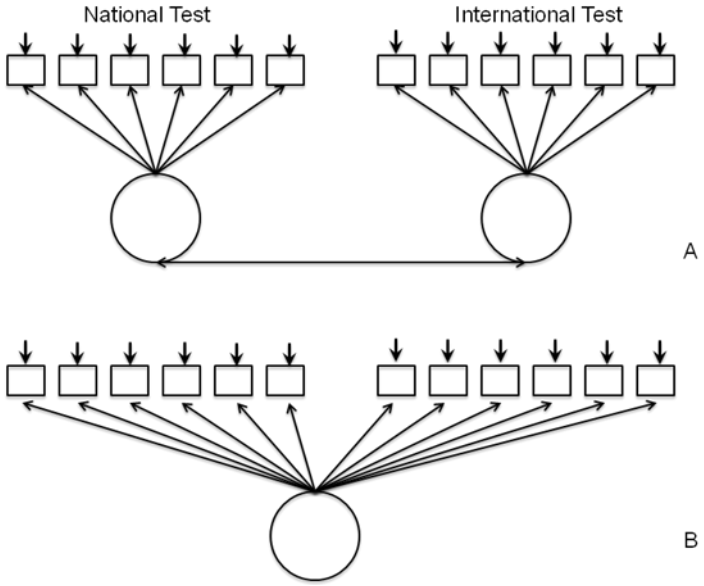
This is also true for other domains, e.g., foreign languages and mother tongue.

Recently, Leucht, Retelsdorf, Möller, and Köller (2010) analyzed tests on reading and listening comprehension in English and found correlations between the two tests above .80.

Also of particular importance are studies in which data from national and international large-scale assessments are used to test whether the instruments measure different constructs. Figure 2 shows two competing models. Rectangles indicate items or subtests; circles symbolize constructs (latent variables) like mathematics or reading competence.

In model A the national and international test measure different but correlated constructs, in model B the different tests measure the same construct.

Figure 2. Competing competence structure and measurement models of national and international tests



Two German studies have recently been published on this issue. Pietsch, Böhme, Robitzsch, and Stubbe (2009) analyzed data from *n* = 4,728 4th graders who worked on items from the Progress in International Reading Literacy Study (PIRLS) and on items from the national test to assess standards in reading in 2006. A two-dimensional IRT analysis revealed a correlation of *r* = .84 indicating that despite a strong overlap between the two constructs, they were not the same (in model fit comparisons the two-dimensional model outperformed the one-dimensional solution).

Of even more interest is a study conducted at the IPN (Frey, 2010). In this study data from approximately 10,000 9th graders who worked on international PISA mathematics items and national mathematics items for the assessment of educational standards were analyzed by means of confirmatory factor analysis. The analyses revealed a correlation of *r* = .95 between the national and the international mathematics test, providing evidence that both instruments more or less measure the same construct. Overall, the two studies show that international and national instruments have a strong overlap in what they measure. It might thus be possible to use national instruments to calibrate students on the international scale and vice versa.

2.2 Research on the Development of Competencies

From an educational perspective it is not only interesting to learn about competence structures but also how competencies and their structures change over time. Therefore, three different types of developmental research have been established, the first one investigating domain-specific achievement or competence gains over time, the second one studying how competence structures change over time, and the last one addressing the robustness of measurement models over time. Figure 3 provides graphical representations of these three research lines.

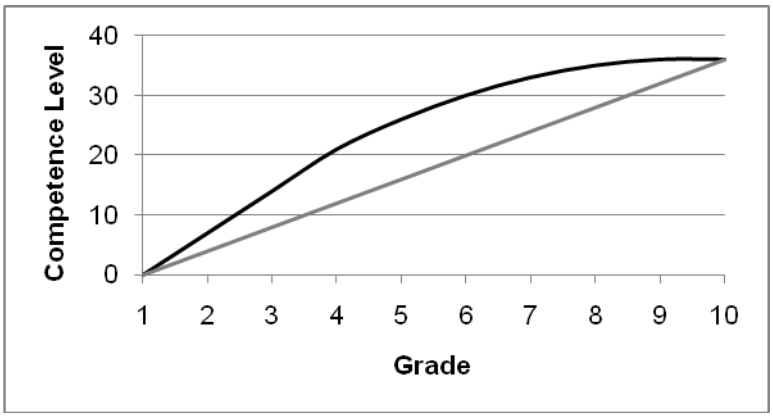
The first graph (3A) shows different growth trajectories (linear vs. non-linear change) in correlation to years of schooling. Studies investigating change of competencies usually use global measures and ask how much students learn in a domain per year. Recent U.S. studies by Lee (2010) and Hill, Bloom, Black, and Lipsey (2008) suggest, that at least for mathematics and reading, non-linear growth over time (similar to the black line in figure 3A), i.e., learning rates are much higher in earlier years (about one standard deviation in primary school) than in later school years (about .10 standard deviations in grade 10). German studies (for an overview see Köller & Baumert, in press) revealed similar results. Findings of the Institute for Educational Progress (IQB) show substantial growth rates in primary school (reading and math scores in grade 4 were about .70 standard deviations higher than in grade 3) becoming much smaller at the end of lower secondary school (.10 standard deviations in reading and about .40 standard deviations in mathematics). Studies on longitudinal change in science competencies are rare but suggest that learning rates are smaller in general (on average .25 standards deviations per year, e.g., Köller & Baumert, in press). Viering, Fischer, and Neumann (2010) report findings on the development of content knowledge in physics.

Based on the educational standards in physics, the authors were interested how students' basic concept of energy changes over time. Students from grades 6 to 10 worked on standardized tests. Cross-sectional as well as longitudinal analyses revealed that students' knowledge increased from grade to grade. Growth rates, however, were relatively small (e.g., .24 standard deviations from grade 6 to grade 7).

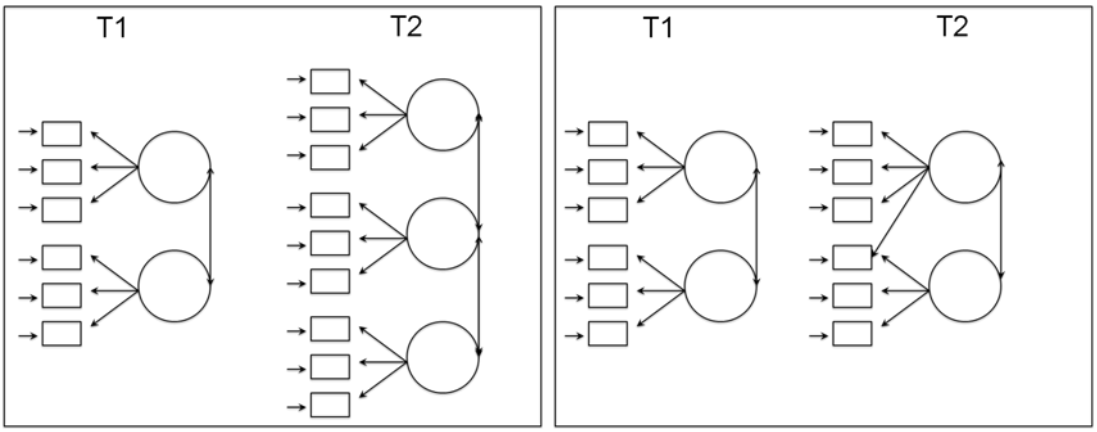
The second graph in figure 3 (B) displays how competence structures can change over time, i.e., become more complex as a result of differentiation of cognitive abilities (cf. Carlstedt, 2001). As an example, the figure shows that a competence structure with two correlated factors in earlier years (T1) becomes more complex (three correlated factors) in later years (T2). Alternatively, differentiation of competencies can result in decreasing correlations among factors over time.

Figure 3: Three research lines on the development of competencies; (A) Growth of competencies; (B) Change of competence structures in terms of differentiation; (C) Change of measurement models

A



B **C**



One explanation of this very popular differentiation hypothesis is that competence acquisition in earlier school years is strongly influenced by general cognitive abilities, whereas later opportunities to learn dominate competence acquisition thus resulting in more complex structures (e.g., Undheim, 1978). Empirical findings of research in academic achievement, however, do not provide much evidence for such differentiation processes. More convincing evidence has been found for a de-differentiation, i.e., correlations among factors increase over time (e.g., Brunner, 2006). The analyses of Winkelmann and Robitzsch (2009) for mathematics in primary school revealed lower correlations among different factors than was the case in Brunner's (2006) analyses of the PISA 2000 mathematics data at the end of lower secondary school. Future studies on the change of competence structures over time are necessary however, because not much research has been conducted in this area.

Invariance of measurement models is often an important issue in studying developmental processes. The last graph (C) in figure 3 demonstrates one example where invariance does not hold true over time. At T1 all items only have one factor loading (so-called *between-item dimensionality*), whereas at T2 at least one item has loadings on both factors (so-called *within-item dimensionality*; for the distinction of between- and within-item dimensionality see Zhang, 2004, 2005). Such changes of measurement models over time typically represent a serious problem because they question the validity of the measures. If measures change their characteristics over time, it is often impossible to interpret mean level changes in students (for a more detailed discussion of measurement invariance see Marsh, Lüdtke, Muthen, et al., 2010). The already mentioned study by Viering et al. (2010) provides some evidence at least for content knowledge, in terms of the basic concept energy, that measurement models can be robust over time. However, the discussion about linking errors in large-scale assessments (e.g., Mazzeo & von Davier, 2008; Xu & von Davier, 2010) clearly shows how changes in competencies can be biased when measurement models change over time.

2.3 Research on Standard-setting

Educational standards and competencies inevitably result in a system of national assessment in order to test whether students reach the expectations formulated in the standards. The German

standards from 2003 and 2004 were labeled by the KMK as so-called *regular standards* (in German: "Regelstandards"). This was strongly criticized and people argued to change the definition to *minimum standards*. A closer look into the corresponding documents, however, clearly suggested that these standards were neither basic nor regular standards. Instead they represented more or less broad ideas of what students should achieve by the end of primary and secondary school without any substantial evidence of proportions of students reaching the proposed achievement levels. Consequently, researchers argued (e.g., Klieme et al., 2003; Köller, 2008b; Schecker & Parchmann, 2007) that more theory building and empirical studies are necessary to come to a deeper understanding of standards and how they can be measured. Furthermore it was argued that data from empirical studies should form the basis for a more concrete description of regular and minimum standards (e.g., Köller, 2000b).

Starting in 2005, the Institute for Educational Progress (IQB), in collaboration with many universities and research institutes, has created large item banks consisting of calibrated items to assess competencies (see Köller, 2008a and 2008b for an overview) and to form a basis for national assessments of educational outcomes. Items have been calibrated on the basis of nationally representative samples of 3rd and 4th graders in primary school and 8th, 9th, and 10th graders in lower secondary school (Granzer, Köller, Bremerich-Vos et al. 2009; Köller, Knigge, & Tesch, 2010). The calibration studies (for an overview see www.iqb.hu-berlin.de) resulted in national competence scales with means of m = 500 and standard deviations of sd = 100.[1] Standard-setting procedures (Cizek, 2001; 2006; Pant, Rupp, Tiffin-Richards, & Köller, 2009) were applied to define proficiency levels (competence levels) and to link these proficiency levels to educational standards.

The term standard-setting subsumes various consensual approaches that consist of having panels of experts set discrete cut-scores on continuous proficiency scales. These cut-scores mark the boundaries of successive achievement levels (e.g., below minimum, minimum, proficient, and advanced) and are used to classify individuals. These cut-scores are often set as a means to facilitate reports about students' achievements (see e.g., the PISA-reports).

[1] Scale means of 500 and standard deviations of 100 were defined for grade 4 in primary school and grade 9 in secondary school.

It is believed that criterion-referenced interpretations of performance based on discrete proficiency levels can be communicated more easily to a wide variety of stakeholders than norm-referenced interpretations based on continuous proficiency scales.

In terms of educational standards and their assessment, students can thus be categorized and assigned to proficiency levels; the simplest being dichotomous: *proficient* or *not*. In the German debate about standards, political expectations were that standard-setting procedures should provide cut-scores for five different competence levels, i.e.:

Level 1: below minimum standard

Level 2: minimum standard

Level 3: regular standard (proficient)

Level 4: regular standard plus (proficient plus)

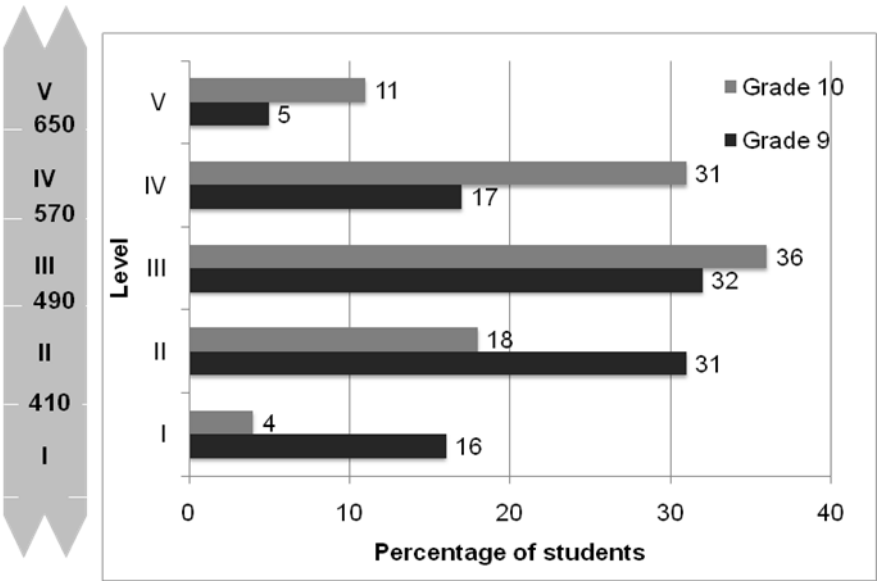
Level 5: ideal standard

Minimum standards describe a competence level, which in the optimal scenario, should be reached by one hundred percent of the students. Proficient (level 3) corresponds to the expectations that were published by the KMK; proficient plus represents a competence level that most German students should reach in the long run and ideal standards describe expectations for high achievers under optimal instructional conditions. Documents have been published (see www.iqb.hu-berlin.de) in which all five proficiency levels are described providing ideas about the skills students express from level 1 to level 5 for different subjects and grade levels. Figure 4 shows the distribution of German 9th and 10th graders for the MSC standards in mathematics, indicating that large proportions of students reach the competence expectations of our national standards (level 3 or higher; 78% in grade 10 and 58% in grade 9). These cut-scores are the result of a standard-setting procedure based on the so-called bookmark method (e.g., Cizek & Bunch, 2007). With this method experts set their cut-scores by placing a marker between groups of items arranged by ascending difficulty in an *ordered item booklet* (Mitzel, Lewis, Patz, & Green, 2001). The item immediately following the bookmark is considered to be the first item of a proficiency level and the last item in the ordered item booklet (OIB) a borderline examinee of that level is likely to answer correctly.

The difficulty (on an IRT-500, 100 scale) of this item marks the cut-score as presented in Figure 4.

There is a lot of research questioning the validity of such standard-setting procedures (cf. Pant et al., 2009). Authors have also argued that different methods of standard-setting lead to different cut-scores and different proficiency level distributions of students. In a recent paper, Tiffin-Richard, Pant, and Köller (2011) compared results of the *bookmark* and the *Angoff method*. This last technique requires experts to read through a test and for each item judge whether a barely proficient or *borderline examinee* could answer the item correctly. The *modified Angoff Method* suggests experts judge the *probability* with which a borderline examinee could answer each item correctly. The borderline examinee is conceptualized as having an ability which places him or her at the lowest end of a proficiency level and thus places him or her just over the threshold between two adjacent levels.

Figure 4. Percentages of German 9th and 10th graders reaching different proficiency levels in mathematics (basis: standards for the middle school certificate MSC)



The results of Tiffin-Richards et al. clearly suggest that Angoff panelists placed higher cut-scores than the bookmark panelists, resulting in larger proportions of students in lower proficiency levels.

This underlines the notion of the arbitrariness of standard-setting methods.

Researchers as well as policy makers are thus requested to publish how such cut-scores were defined.

From a scientific perspective, however, the findings call for more research on how setting cut-scores is influenced by different methods and different compositions of expert panels.

4. Conclusion

As a consequence of the international large-scale assessments like PISA, PIRL, and TIMSS, a worldwide paradigm shift from input to output orientation has taken place in educational governance. Despite a lot of criticism regarding these tests, achievement and/or competence tests measuring students' educational outcomes have become increasingly popular. This paradigm change has strongly stimulated international and national research on competence structures, their assessment, and their development in primary and secondary school. Hence, in the last few years Germany has initiated systems of national assessment as well as large grant programs.

A strong focus has been on psychometric models (e.g., the priority program of the DFG) to better understand what competencies are and how they can be measured. However, from a more educational point of view, it seems to be important to extend this research program to projects on learning and instruction. The idea of competencies requires a type of teaching that is strongly student-oriented and expects students to be cognitively activated by teachers so that they can self-regulate work on demanding tasks and problems.

Not much research, however, has been carried out to better understand how we can implement such a cognitively activating type of instruction in German classes and what this means for programs in teacher professional development. Consequently, future research in the field of competencies should more strongly focus on relating instruction to growth in competencies.

References

Adams, R. J., Wilson, M. R. & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit. *Applied Psychological Measurement, 21*, 1-24.

Beaton, A., Martin, M. O., Mullis, I. V. S., Gonzales, E. J., Smith, T. A., & Kelly, D. L. (1996). *Science achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Blömeke, S., Kaiser, G. & Lehmann, R. (2011). Messung professioneller Kompetenz angehender Lehrkräfte. „Mathematics Teaching in the 21st Century" und die IEA-Studie TEDS-M. In H. Bayrhuber, U. Harms, B. Muszynski, B. Ralle, M. Rothgangel, L. H. Schön, H. J. Vollmer, & H. G. Weigand, (Eds.), *Empirische Fundierung in den Fachdidaktiken* (= Fachdidaktische Forschungen, Band 1, pp. 9-26). Münster: Waxmann.

Blömeke, S., Kaiser, G., Lehmann, R., König, J., Döhrmann, M., Buchholtz, C. & Hacke, S. (2009). TEDS-M: Messung von Lehrerkompetenzen im internationalen Vergleich [TEDS-M: Assessment of teacher competencies in a internatioal comparison]. In R. Mulder, O. Zlatkin-Troitschanskaia, K. Beck, N. Reinhold & D. Sembill (Eds.), *Professionalität von Lehrenden – Zum Stand der Forschung* [Professional knowledge of teachers – State of research] (pp. 181-210). Weinheim: Beck.

Bybee, R. (2008). Scientific literacy, environmental issues, and PISA 2006: The 2008 Paul F-Brandwein lecture. *Journal of Science Education and Technology, 17*, 566-585.

Carlstedt, B. (2001). Differentiation of cognitive abilities as a function of level of general intelligence: A latent variable approach. *Multivariate Behavioral Research*, 36, 589-609.

Cizek, G. (2001). Conjectures on the rise and call of standard-setting: An introduction to context and practice. In G. J. Cizek (Ed.), Setting performance standards: Concepts, methods, and perspectives (pp. 3–17). Mahwah, NJ: Erlbaum.

Cizek, G.J. (2006). Standard setting. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 225–258). Mahwah, NJ: Erlbaum.

Cizek, G. J., & Bunch, M. B. (2007). *Standard-setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.

Council of Europe (2001). *Common European Framework of Reference for Languages*. Cambridge: Cambridge University Press.

Fischer, C., Rieck, K., & Prenzel, M. (2010) (Eds.). *Naturwissenschaften in der Grundschule: Neue Zugänge entdecken* [Science in primary school: Exploring new approaches]. Seelze: Klett, Kallmeyer.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht: Reidel.

Frey, A. (2010). Konstruktvalidität des Tests zur Messung der Bldungsstandards in Mathematik für den Mittleren Schulabschluss [Construct validity of a test assessing the educational standards in mathematics at the end of lower secondary school]. In IPN (Ed.), *Research Report 2009/2010* (pp. 132-134). Kiel: Leibniz Institute for Science and Mathematics Education.

Granzer, D. Köller, O., Bremerich-Vos, A. et al. (Eds.), *Bildungsstandards Deutsch und Mathematik*. *Leistungsmessung in der Grundschule* [Educational standards for German and mathematics: Assessment in primary school]. Weinheim: Beltz.

Hartig, J., Klieme, E., & Leutner, D. (Eds.) (2008). *Assessment of competencies in educational contexts*. Göttingen: Hogrefe.

Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement.* London: Routledge.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*, 172-177.

Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W., & Stanat, P. (Hrsg.) (2010). PISA 2009. *Bilanz nach einem Jahrzehnt* [PISA 2009: Conclusion after a decade]. Münster: Waxmann.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E., & Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards: Eine*

Expertise. [On the development of national educational standards: An expertise]. Berlin: Federal Ministry of Education and Research.

Köller, O. (2008a). Bildungsstandards in Deutschland: Implikationen für die Qualitätssicherung und Unterrichtsqualität [Educational standards in Germany: Implications for assessment and teaching quality]. *Zeitschrift für Erziehungswissenschaft, 10, Sonderheft 9/2008*, 47-59.

Köller, O. (2008b). Bildungsstandards – Verfahren und Kriterien bei der Entwicklung von Messinstrumenten [Educational Standards – Procedures and criteria for test development]. *Zeitschrift für Pädagogik, 54*, 163-173.

Köller, O. (2010). Bildungsstandards [Educational Standards]. In D. H. Rost (Hrsg.), Handwörterbuch *Pädagogische Psychologie* [Handbook of Educational Psychology] (4th Ed., pp. 77-83). Weinheim: Beltz/PVU.

Köller, O., & Baumert, J. (in press). Schulische Leistungen und ihre Messung [Academic achievement and assessment]. In R. Oerter, & L. Montada (Eds.), *Entwicklungspsychologie* [Developmental psychology] (7th ed.). Weiheim: Beltz/ PVU.

Köller, O., Knigge, M. & Tesch, B. (Hrsg.) (2010). Sprachliche Kompetenzen im Ländervergleich [Verbal competencies in a national comparison]. Münster: Waxmann.

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary Mathematics Teachers' Pedagogical Content Knowledge and Content Knowledge: Validation of the COACTIV Constructs. *The International Journal on Mathematics Education (ZDM), 40*, 873-892.

Krauss, S., Brunner, M., Kunter, M., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical Content Knowledge and Content Knowledge of Secondary Mathematics Teachers. *Journal of Educational Psychology, 100*, 716-725.

Krauss, S., Neubrand, M., Blum, W., Baumert, J., Brunner, M., Kunter, M., & Jordan, A. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und –Lehrer im Rahmen der COACTIV-Studie [An investigation on professional knowledge of mathematics teacher. The COACTIV project]. *Journal für Mathematik-Didaktik (JMD), 29*, 223-258.

Lee, J. (2010). Tripartite growth trajectories of reading and math achievement: Tracking national academic progress at primary, middle, and high school levels. *American Educational Research Journal, 47*, 800-832.

Leucht, M., Retelsdorf, J., Möller, J., & Köller, O. (2010). Zur Dimensionalität rezeptiver englischsprachiger Kompetenzen [On the dimensionality of receptive English competencies]. *Zeitschrift für Pädagogische Psychologie, 24*, 123-138.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the Big Five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22*, 471-491.

Mazzeo, J., & von Davier, M. (2008). *Review of the Programme for International Student Assessment (PISA) test design: Recommendations for fostering stability in assessment results*. EDU/PISA/GB(2008)28.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The Bookmark procedure: Psychological perspectives. In G. Cizek (Ed.), *Setting performance standards: Concepts, methods and perspectives* (pp. 249-281). Mahwah, NJ: Erlbaum.

NCTM (2000). *Pricinples and standards for school mathematics*. Reston, VA: NCTM.

OECD (2003). *The PISA 2003 framework – Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.

OECD (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: OECD.

OECD (2009). *PISA 2009 assessment framework – Key competencies in reading, mathematics and science*. Paris: OECD.

Pant, H. A., Rupp, A. A., Tiffin-Richards, S., & Köller, O. (2009). Validity issues in standard-setting studies. *Studies in Educational Evaluation*, 15, 95-101.

Pietsch, M., Böhme, K, Robitzsch, A., & Stubbe, T. (2009). Das Stufenmodell zur Lesekompetenz der länderübergreifenden Bildungsstandards im Vergleich zu IGLU 2006 [Comparison of the proficiency models of reading comprehension in national educational standards and PIRLS 2006].

In. D. Granzer, O. Köller, A. Bremerich-Vos et al. (Eds.), *Bildungsstandards Deutsch und Mathematik*. *Leistungsmessung in der Grundschule* [Educational standards for German and mathematics: Assessment in primary school] (pp. 401-425). Weinheim: Beltz.

Rost, J., Walter, O., Carstensen, C. H., Senkbeil, M., & Prenzel, M. (2004). Naturwissenschaftliche Kompetenz [Science competency]. In PISA-Konsortium Deutschland (Ed.), PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs [Academic achievement of German adolescents – Results of the second international comparison] (pp. 111-146). Münster: Waxmann.

Schecker, H., & Parchmann, I. (2007). Standards and competence models: The German situation. In D. Waddington, P. Nentwig, & S. Schanze (Eds.), *Making it comparable: Standards in science education* (pp. 147-164). Münster: Waxmann.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15(2)*, 4–14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–22.

Tenorth, H. E. (Ed.) (2004). Kerncurriculum für die Oberstufe II: Biologie, Chemie, Physik, Geschichte, Politik. Expertisen - im Auftrag der KMK [Core curriculum for upper secondary level: Biology, chemistry, physics, history, civics. Expertises fort he KMK]. Weinheim: Beltz.

Tiffin-Richards, S. P., Pant, H. A., & Köller, O. (2011): *Setting standards for English foreign language assessment: Methodology, validation and a degree of arbitrariness*. Paper submitted for publication.

Undheim, J. O. (1978). Broad ability factors in 12- to 13-year-old childree, the theory of fluid and crystallized intelligence, and the differentiation hypothesis. *Journal of Educational Psychology, 70*, 433 – 443.

Viering, T., Fischer, H. E., & Neumann, K. (2010). Die Entwicklung physikalischer Kompetenz in der Sekundarstufe I [The development of physics competence in lower secondary school]. *Zeitschrift für Pädagogik, Beiheft 56*, 92-103.

Waddington, D., Nentwig, P. & Schanze, S. (Eds.) (2008). *Making it comparable. Standards in science education*. Münster: Waxmann.

Weinert, F. E. (2001). Concept of competence – A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45-65). Göttingen: Hogrefe.

Winkelmann, H., & Robitzsch, A. (2009). Modelle mathematischer Kompetenzen: Empirische Befunde zur Dimensionalität [Models of mathematical competencies: Empirical findings on their dimensionality]. In D. Granzer, O. Köller, A. Bremerich-Vos et al. (Eds.), *Bildungsstandards Deutsch und Mathematik*. *Leistungsmessung in der Grundschule* [Educational standards for German and mathematics: Assessment in primary school] (pp. 172-201). Weinheim: Beltz.

Winter, H. (1995). Mathematikunterricht und Allgemeinbildung [Mathematics instruction and general education]. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 61, 37-46.

Xu, X., & von Davier, M. (2010). Linking errors in trend estimation in large-scale surveys: A case study (ETS Research Report RR10-10). Princeton: ETS.

Zhang, J. (2005): *Estimating multidimensional item response models with mixed structure* (ETS Research Report RR-05-04). Princeton, NJ: ETS.

Zhang, J. (2004): *Comparison of unidimensional and multidimensional approaches to IRT parameter estimation* (ETS Research Report RR-04-44). Princeton, NJ: ETS

LUCREZIA STELLACCI

Esperienze di miglioramento sul territorio: Puglia

Abstract

Analisi dei risultati OCSE-PISA 2009 in Puglia (n. 30 slides circa)

Si registra un deciso miglioramento.

Manca la possibilità di un confronto oggettivo con le edizioni precedenti (2000 – 2003 – 2006)

Migliora la media della macroarea Sud che contribuisce in maniera determinante ad elevare la media italiana in tutti e tre i campi d'indagine, atteso che la media delle altre macroaree resta più o meno costante.

Fra le Regioni del Sud che migliorano, la Puglia migliora di più e si colloca al 1° posto delle Regioni del Sud e del Centro.

Come si spiegano questi risultati?

1. Con le politiche europee di investimento nella Scuola delle Regioni ricomprese nell'obiettivo convergenza, che negli ultimi anni si sono concentrate sulle misure idonee ad innalzare le competenze di base degli studenti .

Il Rapporto sulla Scuola in Italia 2010 della Fondazione Agnelli ha misurato in 68 punti lo svantaggio con cui parte uno studente del Sud rispetto ad uno studente del Nord,indipendentemente dalle caratteristiche individuali e della scuola frequentata.

Il costo dell'istruzione per alunno al Sud è di circa mille euro in meno che al Nord (ricerca Banca d'Italia)

2. Valorizzando l'autonomia progettuale e gestionale delle istituzioni scolastiche, senza rinunciare al coordinamento, composto da una seria valutazione dei progetti e da una puntuale rendicontazione dei risultati.

Funzione di governance dell'Amministrazione scolastica, che affida le risorse e gli obiettivi, precisa i compiti, valuta i risultati, curando le patologie.

La Scuola ha un grande potenziale di sviluppo e di innovazione che va incanalato nel senso giusto, guidato, sostenuto, valutato (dare valore) per separare il grano dall'oglio.

Ogni risultato è il gradino di una lunga scala di avanzamento verso un traguardo che deve mantenere sempre alta la tensione, senza cedimenti.

Differenze di performance fra i diversi indirizzi di studio hanno prodotto la programmazione di piani di intervento regionali diversi nei tempi e nei contenuti.

Se ad impegnarsi per il miglioramento delle competenze degli studenti non è solo la Scuola, ma tutto il sistema sociale che le sta intorno, ed in questo l'Amministrazione scolastica può svolgere un ruolo strategico, solo allora la speranza acquista la consistenza di una promessa.