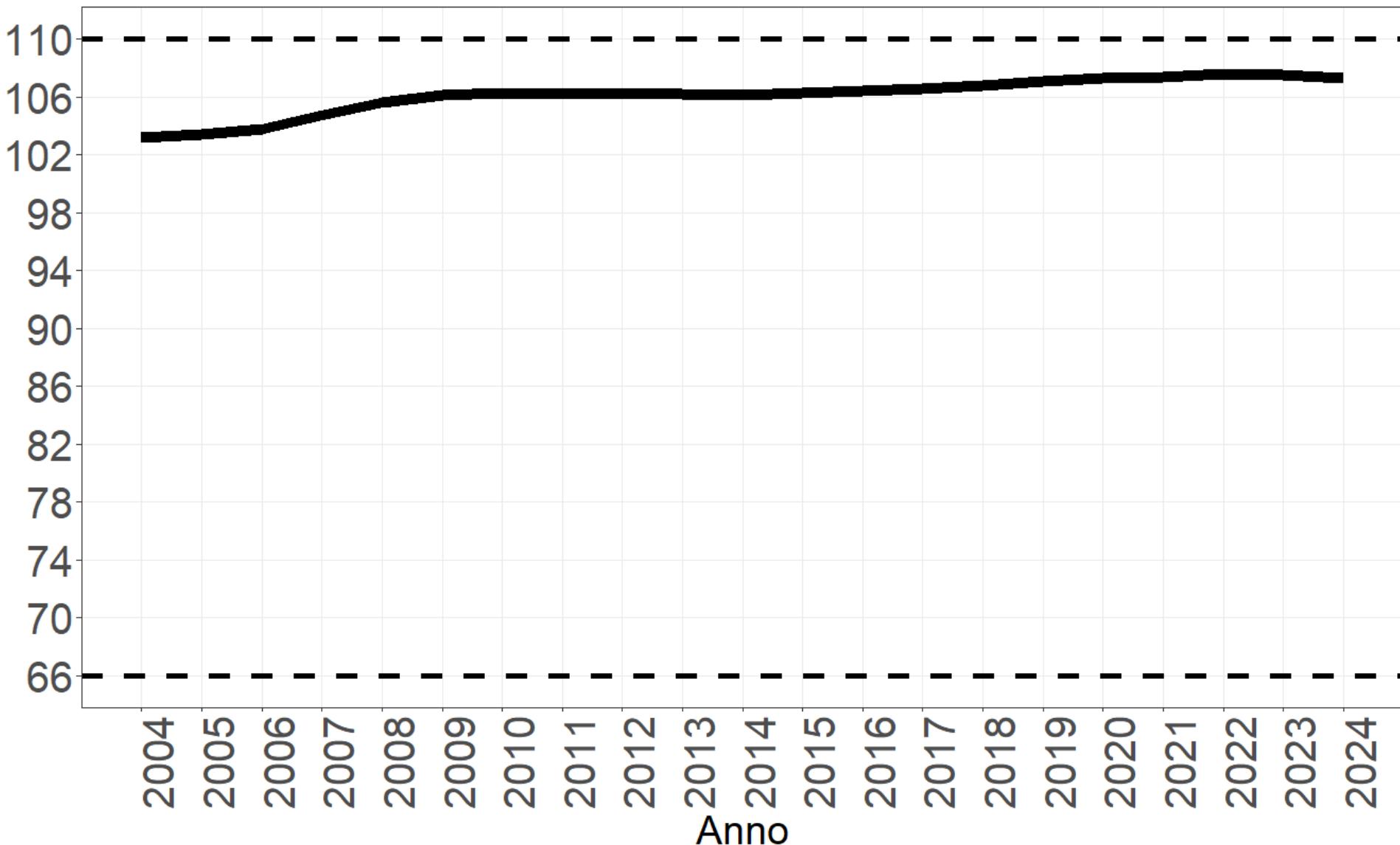


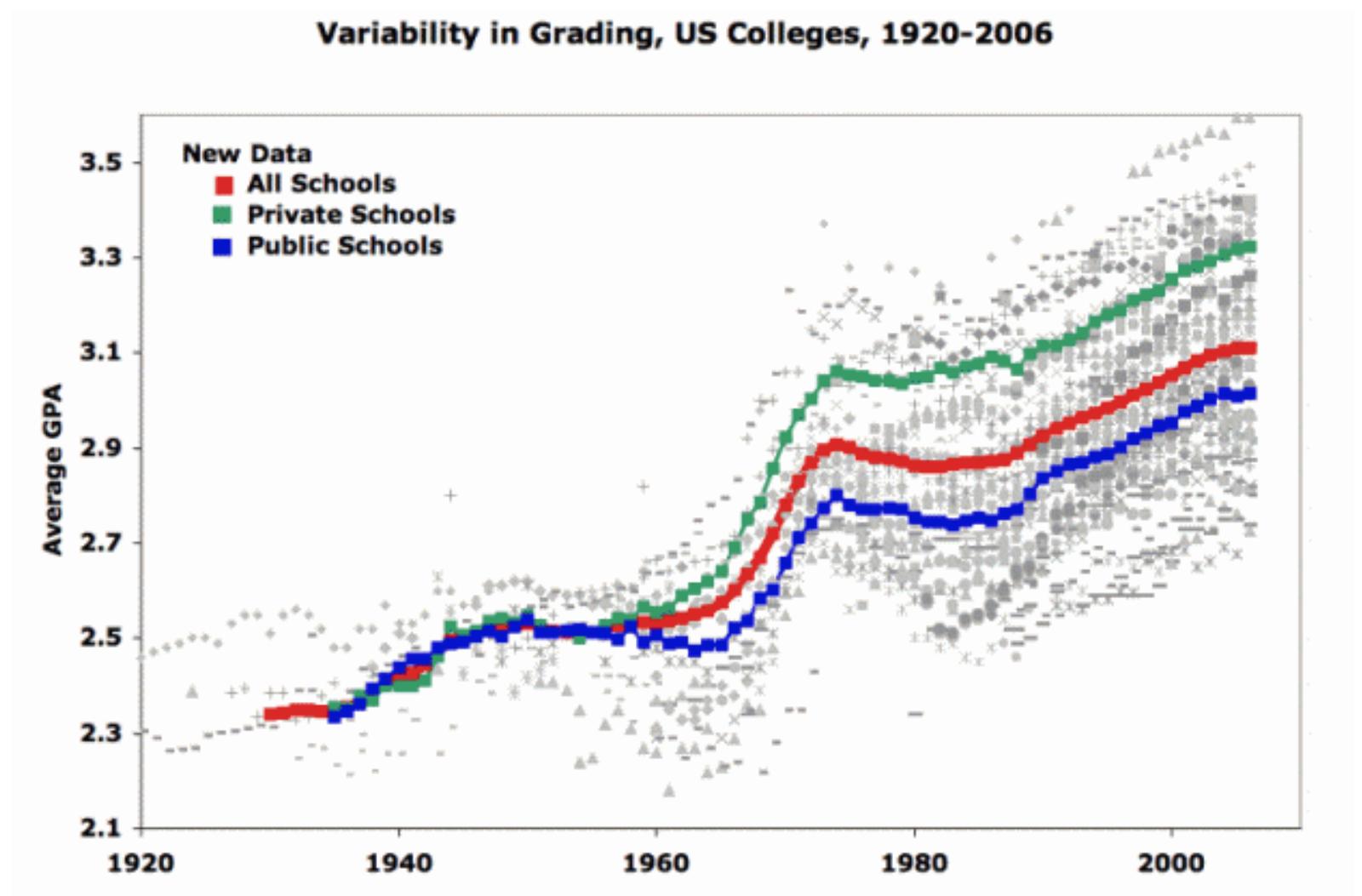
Voto di laurea magistrale medio in Italia

- Fonte: Almalaurea



Grade inflation: why weren't the instructors all giving all A's already??

Posted on July 27, 2011 9:16 AM by Andrew Gelman



“assigns grades in an unregulated environment (that is, we have no standardized tests and no rules on how we should grade), all the incentives to toward giving only A’s: When I give A’s, students are happier and complain less, I get to feel like a nice person, and I give my own students (whom I generally have somewhat warm feelings toward) a benefit in their future lives (...)"

“(...) when I used to organize a class with several different section leaders, each instructor wanted to give his or her students higher grades”

“that profs don’t give all A’s, even though they can, is interesting to me”



Andrew Gelman

Gelman, da statistico, offre una spiegazione «psicologica»: i docenti, che a scuola ottenevano mediamente voti più alti degli altri, non vogliono inflazionare «troppo» (e quindi svalutare) la moneta che li ha resi comparativamente ricchi

2025 update

Problems caused by grade inflation

Posted on January 19, 2025 9:52 AM by Andrew Gelman

- Ricevere «A» anche in tutti gli esami non è più segnale di eccellenza
- Si potrebbe «pesare» il voto in base alla difficoltà dell'esame (un «A» conta davvero se preso in un esame dove spesso si riceve «B» o «C») → ma ormai il «soffitto» è così forte che non resta più «segnale informativo»
- Un singolo voto negativo può essere eccessivamente penalizzante
- Se *i voti universitari non servono più*, chi è eccellente è costretto a trovare altri modi per «distinguersi»,

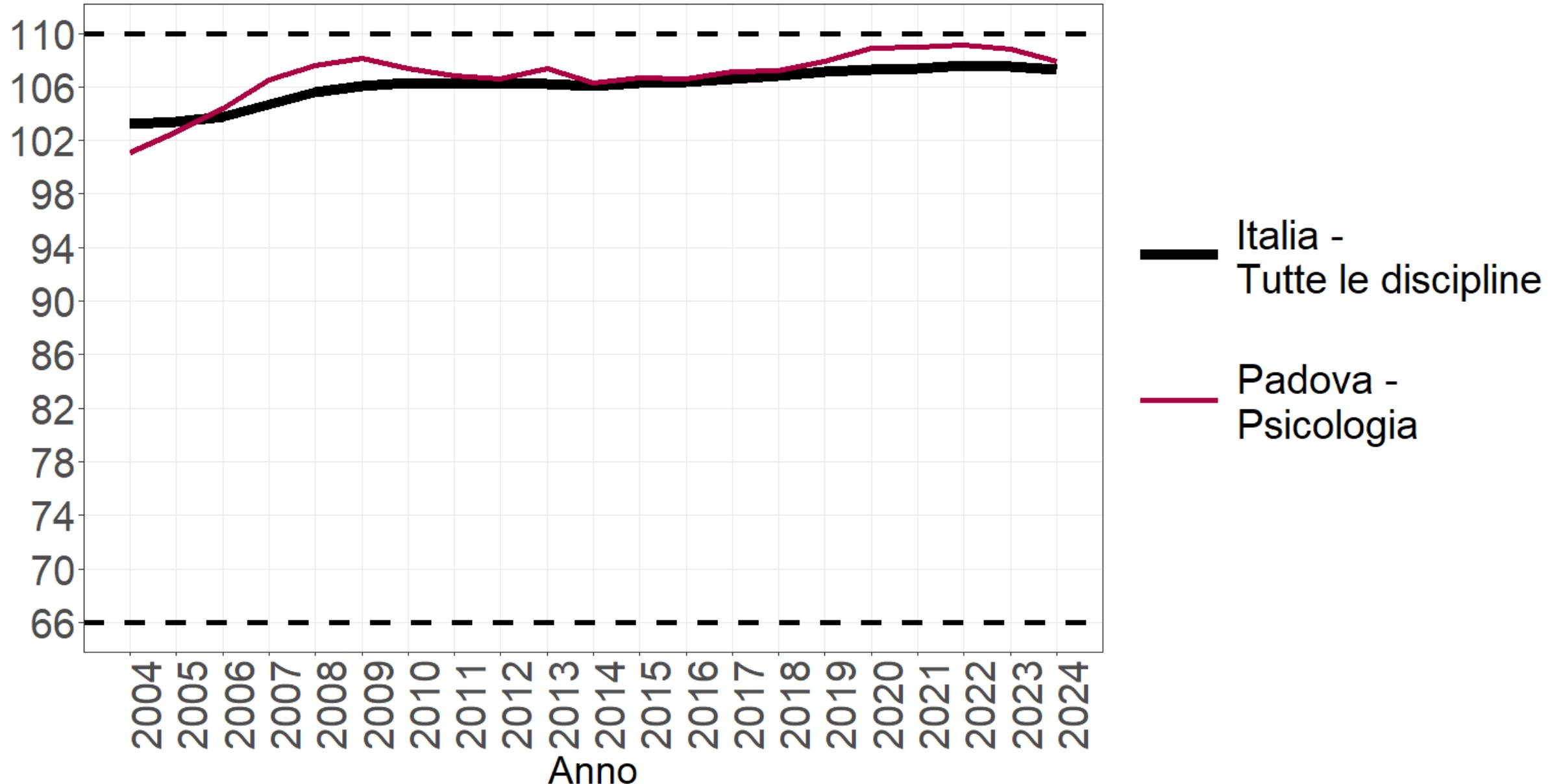


Stiamo screditando l'università

Spinta verso progetti, tirocini, pubblicazioni (*publish or perish* ancor prima del dottorato), modi per trovare raccomandazioni

SITUAZIONE ITALIANA

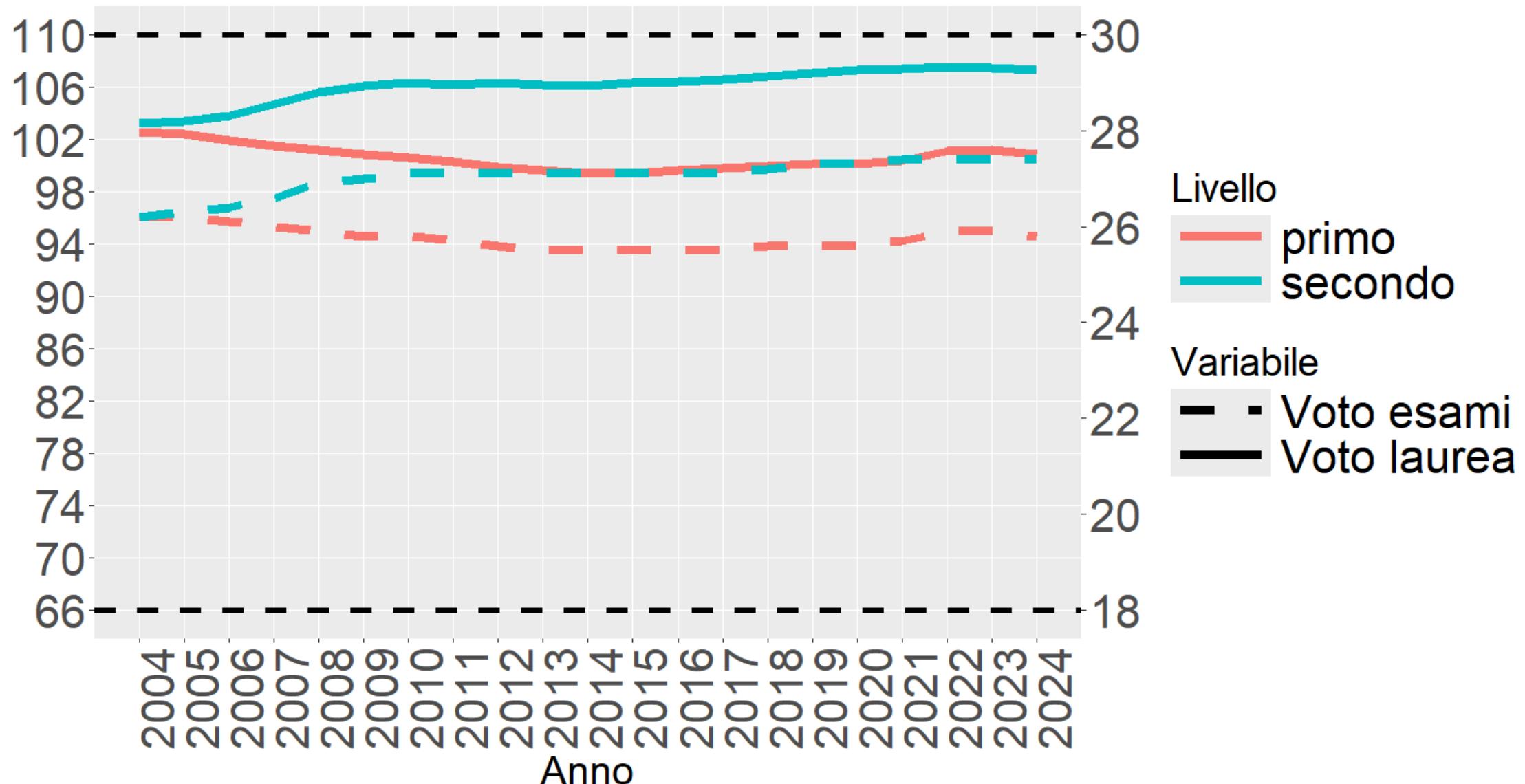
Voto di laurea magistrale medio - Fonte: Almalaurea



Tutte le discipline

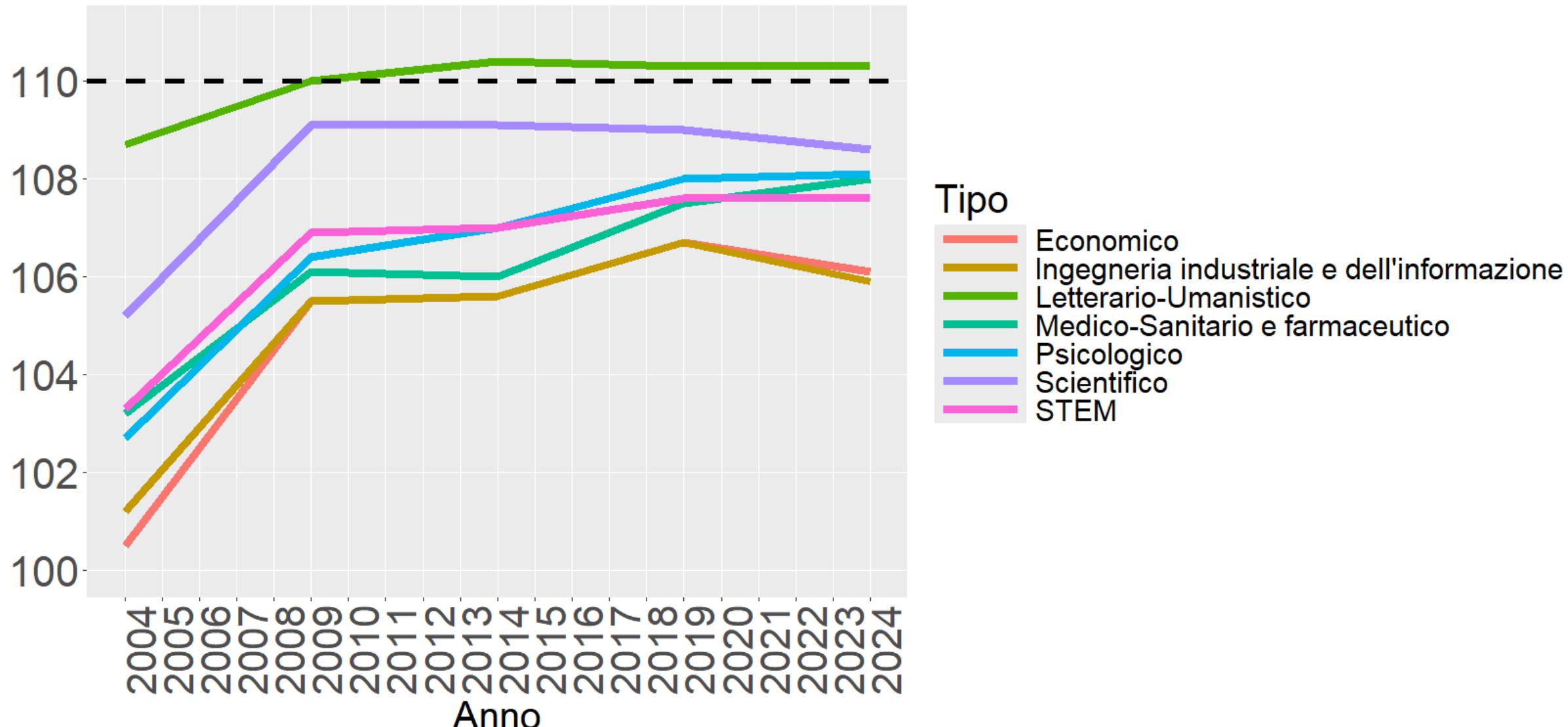
Voto di laurea medio e voto esami - Italia

Fonte: Almalaurea



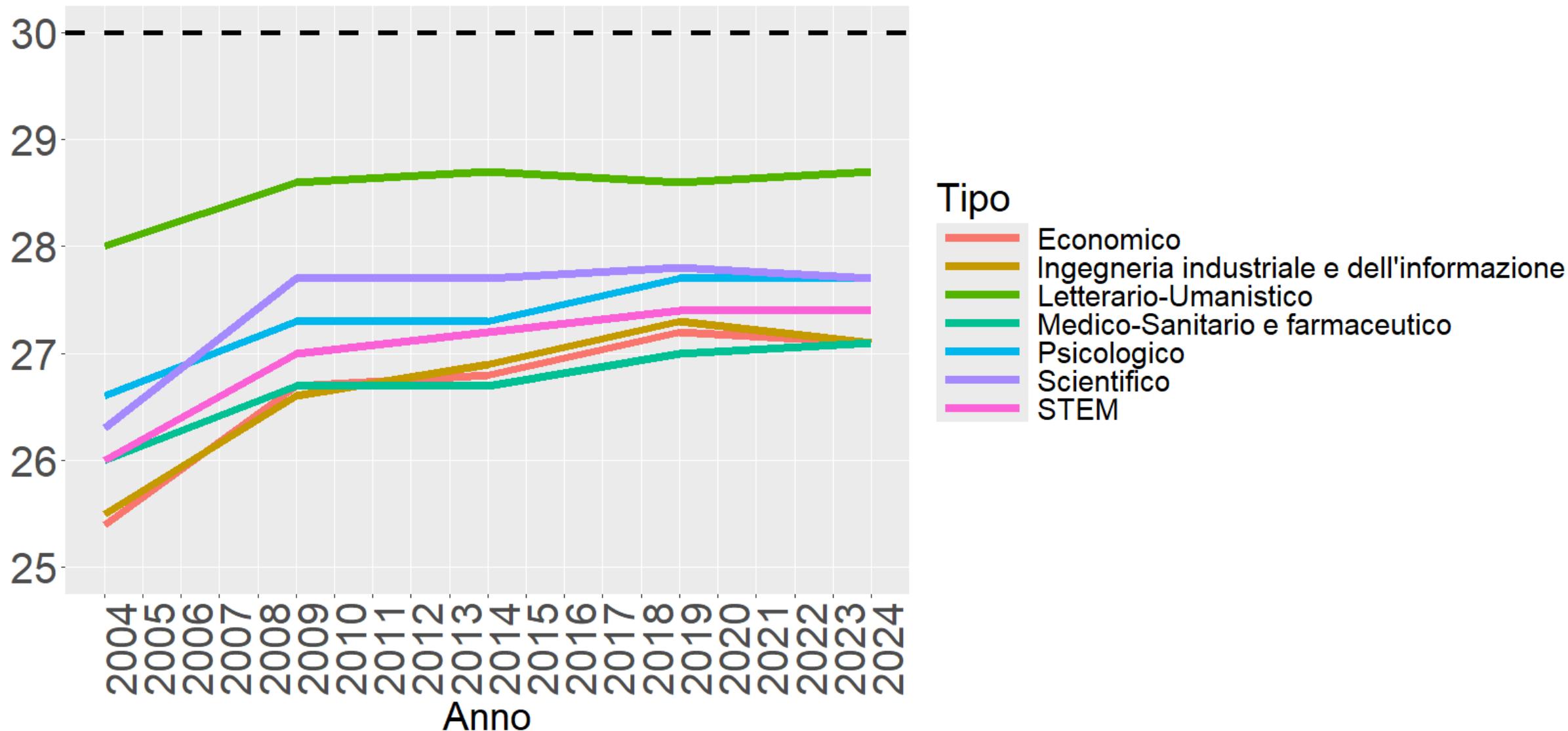
Voti laurea magistrale - Italia

Fonte: Almalaurea



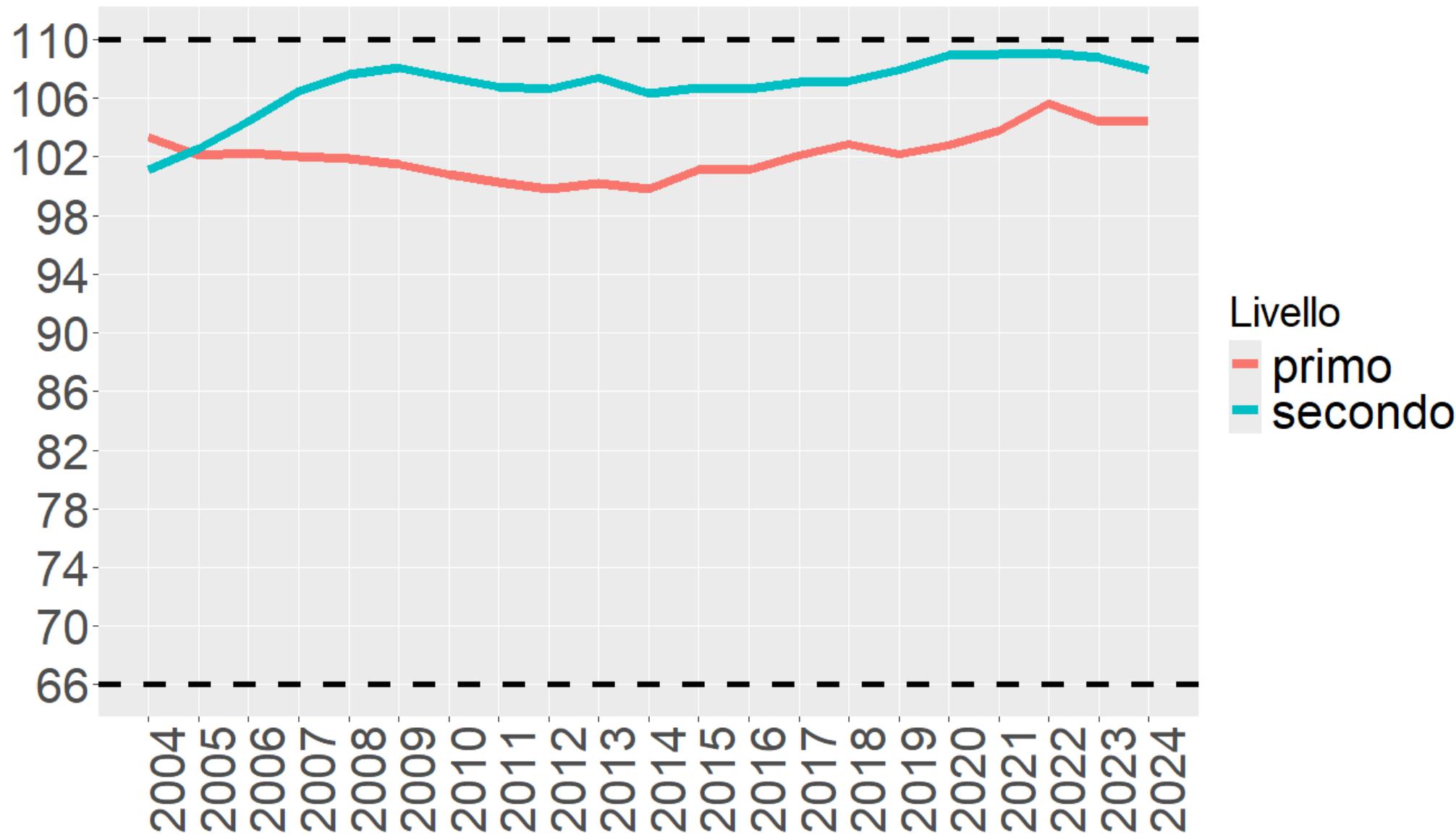
Voti esami magistrale - Italia

Fonte: Almalaura



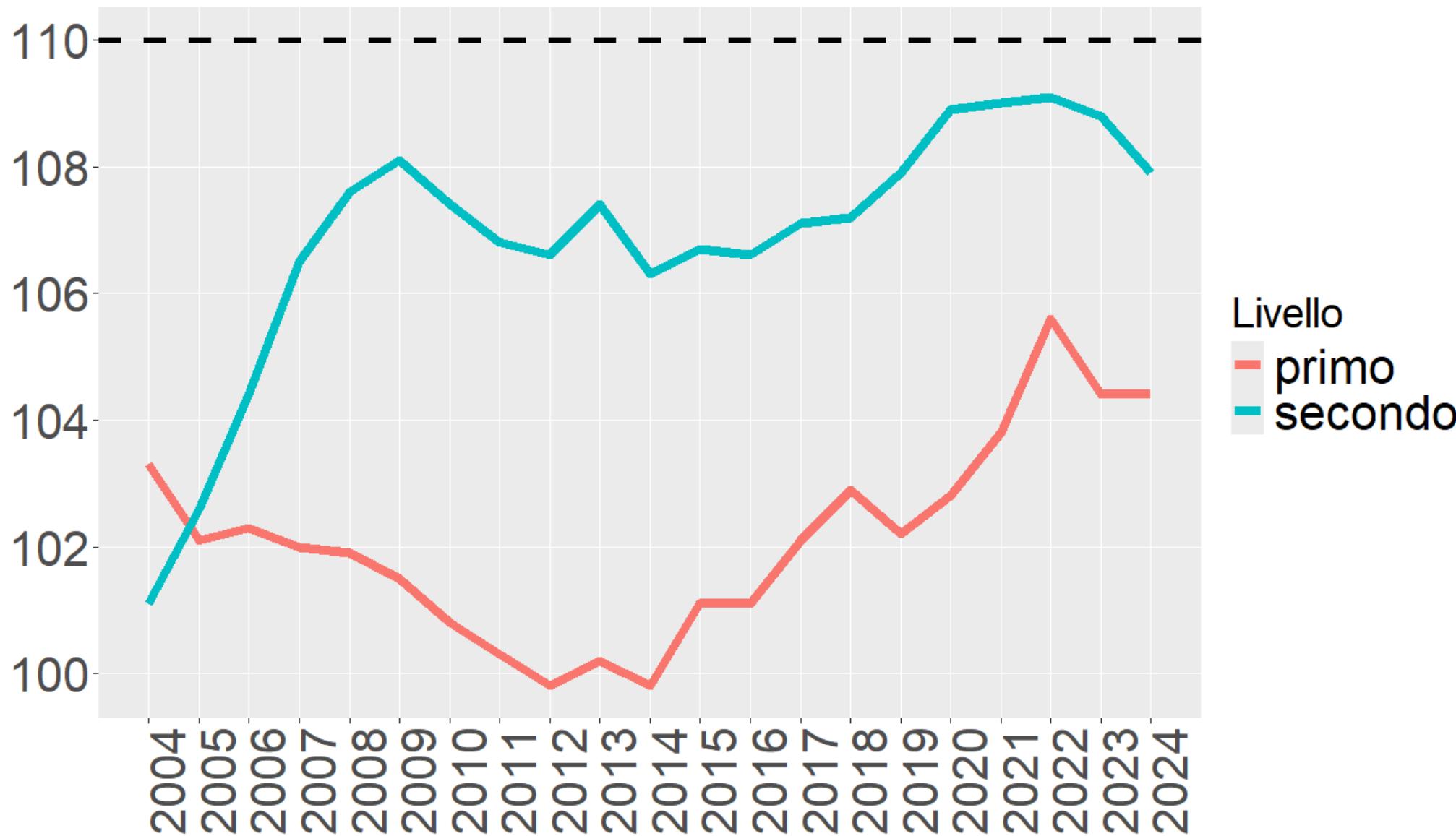
Voto di laurea medio - Psicologia - Università di Padova

Fonte: Almalaurea



Voto di laurea medio - Psicologia - Università di Padova

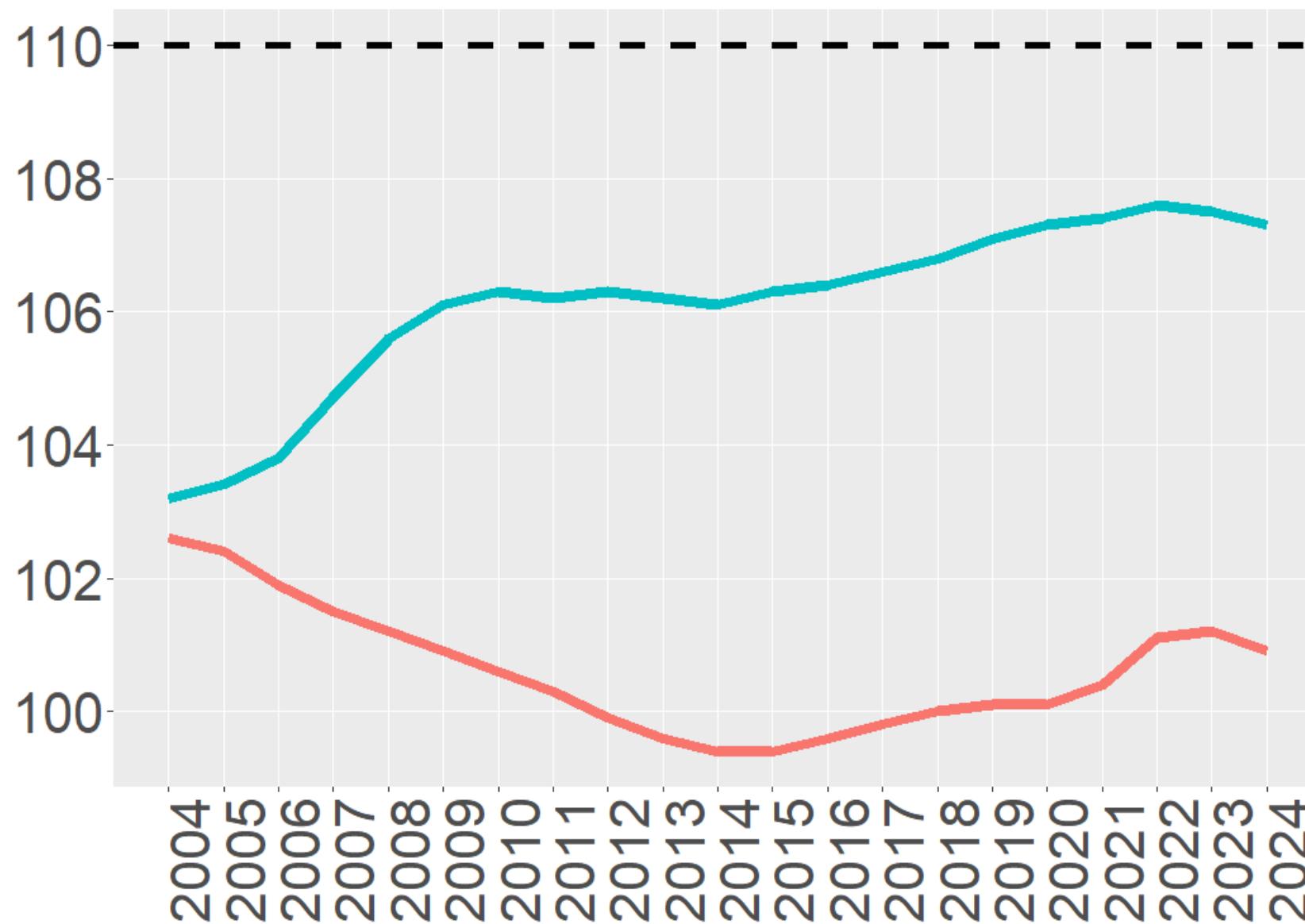
Fonte: Almalaurea



Voto di laurea medio - Italia

Fonte: Almalaurea

Tutte le discipline



Livello

primo
secondo

Grade Inflation:
A Crisis in College
Education

Valen E. Johnson

Springer

Grade Inflation

A Crisis in College Education

Book | © 2003

1st edition | [View latest edition](#)

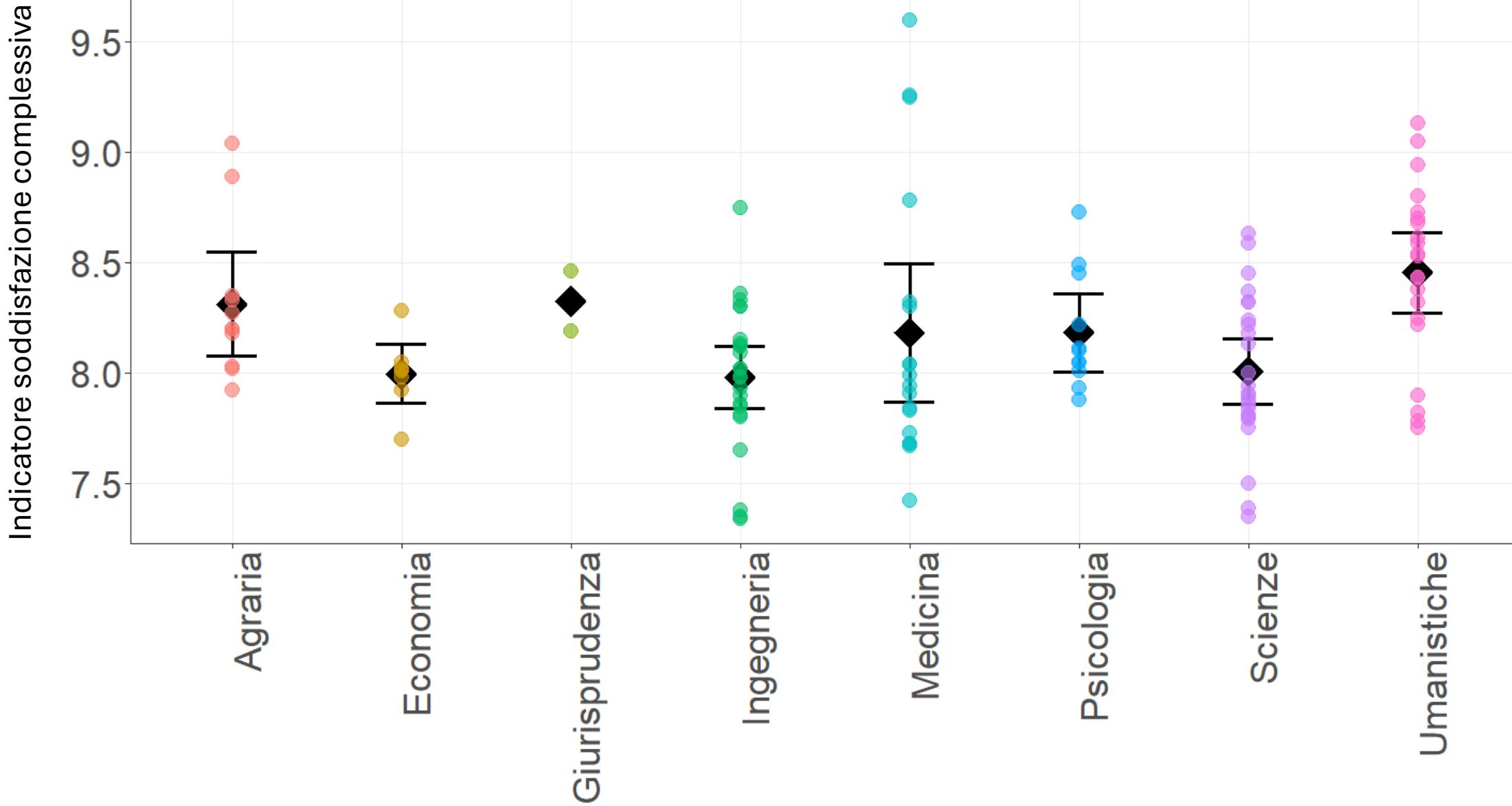
“the correlation between student grades and student evaluations of teaching is shown to result primarily from grade attribution”

“Analyses based on these data demonstrate that the effects of grades on teacher-course evaluations are both substantively and statistically important, and suggest that

instructors can often double their odds of receiving high evaluations from students simply by awarding A's rather than B's or C's”

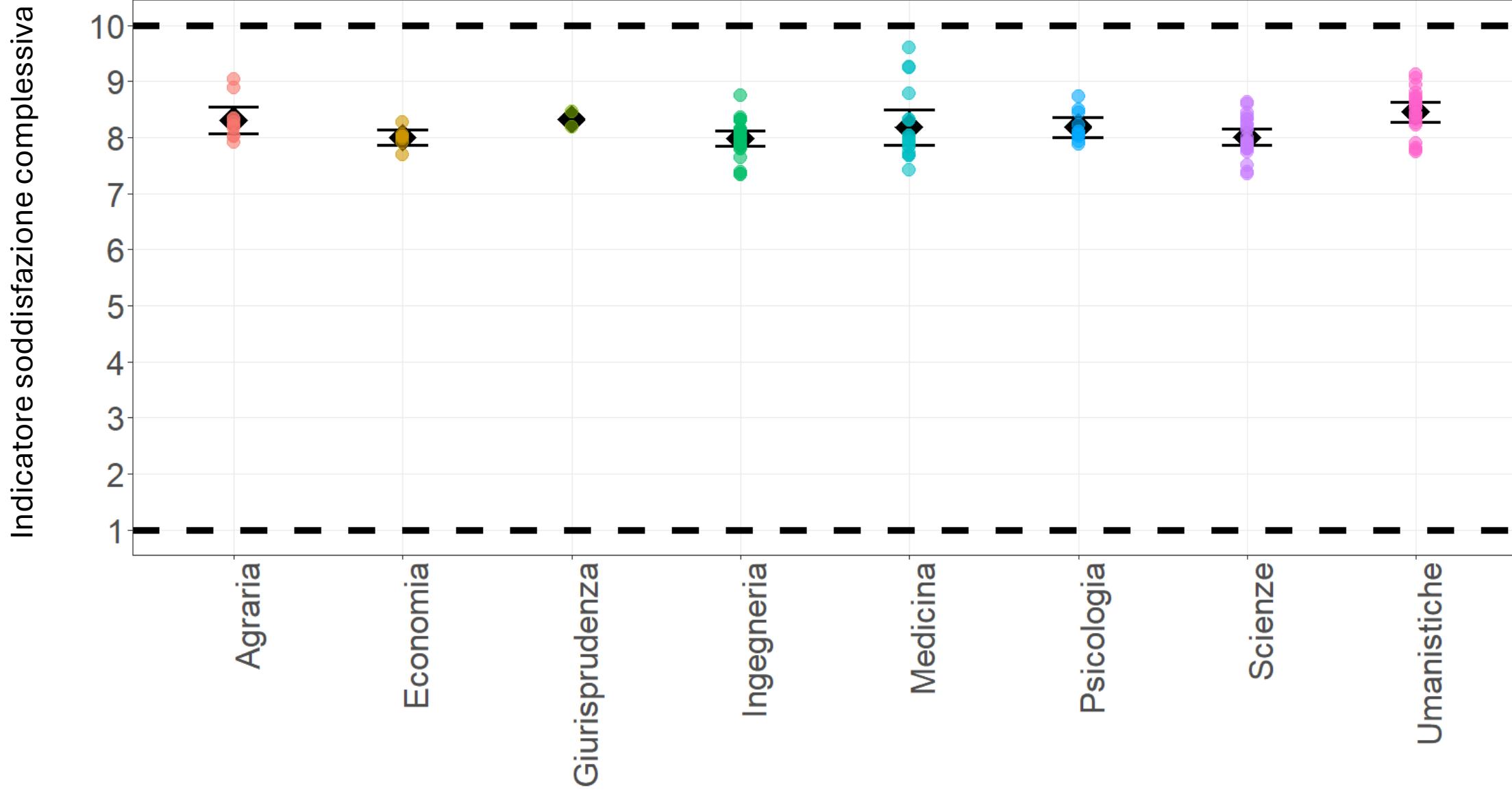
**SODDISFAZIONE
STUDENTESSE E
STUDENTI**

Soddisfazione Media Corsi Magistrali Padova 2024/2025- Italia
Fonte: <https://opinionistudenti.didattica.unipd.it/>



Soddisfazione Media Corsi Magistrali Padova 2024/2025- Italia
Fonte: <https://opinionistudenti.didattica.unipd.it/>

*Comunque meno a
soffitto dei voti di laurea!*



“Quando un indicatore diventa un target da massimizzare, smette di essere un buon indicatore”

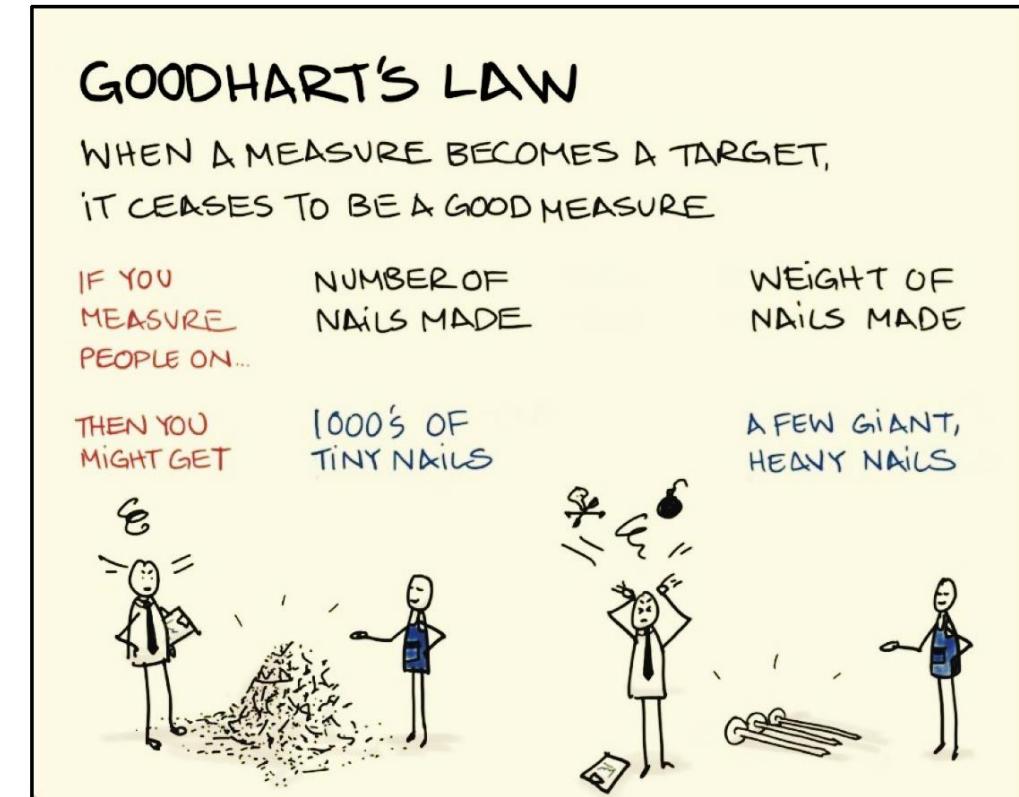
- Legge di Goodhart

“Studenti, docenti e personale amministrativo sembrano impegnati in un gioco individualmente razionale, ma socialmente distruttivo” - Langbein ([2008](#))



Anni '70 - Epoca del monetarismo: banche centrali notano la correlazione tra certi *indicatori* di aggregati monetari (quantità e crescita quantità di moneta, M1/M3), inflazione e crescita economica → fissano questi indicatori come *target* da controllare per governare l'economia. Charles Goodhart (1975) nota che dopo aver fissato questi *target*, molti attori (banche, imprese, famiglie) agiscono diversamente e la correlazione *indicatori-target* sparisce

Un *indicatore* può anche essere un buon *predittore* di un *outcome* di interesse, ma solo finché gli attori in gioco non vengano incentivati a ottimizzarlo (*target*) → il principio è famoso come «*Legge di Goodhart*» – e ha risvolti non solo **economici**, ma anche **psicologici ed etici**



Original Investigation

FREE

 Cite  Permissions  Metrics

The Cost of Satisfaction

A National Study of Patient Satisfaction, Health Care Utilization, Expenditures, and Mortality

Joshua J. Fenton, MD, MPH; Anthony F. Jerant, MD; Klea D. Bertakis, MD, MPH ; [et al](#)

Results Adjusting for sociodemographics, insurance status, availability of a usual source of care, chronic disease burden, health status, and year 1 utilization and expenditures, **respondents in the highest patient satisfaction quartile** (relative to the lowest patient satisfaction quartile) had lower odds of any emergency department visit (adjusted odds ratio [aOR], 0.92; 95% CI, 0.84-1.00), higher odds of any inpatient admission (aOR, 1.12; 95% CI, 1.02-1.23), 8.8% (95% CI, 1.6%-16.6%) greater total expenditures, 9.1% (95% CI, 2.3%-16.4%) greater prescription drug expenditures, **and higher mortality** (adjusted hazard ratio, **1.26**; 95% CI, 1.05-1.53)

**Maggiore soddisfazione associata NEGATIVAMENTE a sopravvivenza,
presumibilmente perché massimizza prescrizioni, ricoveri,
trattamenti inutili, e spese per comfort anziché cure reali** (soddisfazione
clienti nel contesto USA è richiesta da agenzie finanziarie pubbliche, e concorre a compensi medici)

La Soddisfazione degli studenti è un
indicatore della qualità della formazione?

Anche se lo fosse, sarebbe davvero
opportuno trasformarla in un *target* da
massimizzare? *Legge di Goodhart!*

In Italia l'ANVUR «costringe» a raccogliere i dati sulla soddisfazione di studentesse e studenti, ma per fortuna NON usa (?) direttamente l'indicatore per suddividere soldi o punti organico

L'incentivo rimane soprattutto sociale (è molto discusso nei *Consigli di Corso di Studio*)

TUTTAVIA...

Didattica, didattica integrativa e servizio agli studenti	Punti 10
Per il volume e la continuità degli insegnamenti universitari (I, II e III livello) e dei moduli di cui si è assunta la responsabilità (es. per ciascun insegnamento universitario pertinente con il SSD max punti 1/per anno)	Max Punti 4
Per il volume e la continuità dell'attività didattico integrativa universitaria di I, II e III livello e di servizio agli studenti (per ciascuna attività documentata di tipo seminariale pertinente al SSD M-PSI/08 punti 0,5; per ciascuna attività di tutoraggio e supporto alla didattica punti 0,5)	Max Punti 5
Per le valutazioni degli studenti ove presenti per tutti i candidati	Max Punti 1

Cose che si leggevano nei “Verbale 1” delle procedure selettive per RTD fino a qualche anno fa

Didattica, didattica integrativa e servizio agli studenti	Punti 10
Per il volume e la continuità degli insegnamenti e dei moduli di cui si è assunta la responsabilità (per ciascun insegnamento universitario pertinente con il SSD punti 1/per anno)	Max Punti 3
Per il volume e la continuità dell'attività didattico integrativa e di servizio agli studenti (per ciascuna attività di cosupervisione tesi di laurea-laurea magistrale punti 0.5; per ciascuna attività seminariale punti 0.5; per ciascuna attività di tutoraggio punti 1, per ciascuna attività di didattica integrativa es. punti 0.5/per anno)	Max Punti 6.5
Per le valutazioni degli studenti ove presenti per tutti i candidati	Max Punti 0.5

Per fortuna
oggi sembrano
sparite dai
verbali per RTT

**AMMINISTRAZIONE CENTRALE
AREA RISORSE UMANE
UFFICIO PERSONALE DOCENTE**



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Decreto Rep. Prot. n.
Anno 2024 Tit. VII Cl. 4 Fasc. All. n.

OGGETTO: Premialità a favore dei Ricercatori a tempo determinato per risultati rilevanti conseguiti nell'attività di didattica per l'a.a. 2023/2024, ai sensi dell'articolo 3, comma 4 del Regolamento per la Premialità di Ateneo.

favorevole al riconoscimento di un attestato di merito ai Ricercatori a tempo determinato, in regime di impegno a tempo pieno, che hanno conseguito per l'indicatore di sintesi "Soddisfazione complessiva" una valutazione "ottima" o "molto buona" e non hanno ottenuto in nessuna ulteriore attività didattica del docente valutata un valore dell'indicatore inferiore alla mediana del proprio Corso di Studio;

Considerato che per l'anno accademico 2023/2024 il premio è stato quantificato in € 500,00.=lordo percipiente, corrispondente ad un importo di € 671,55.=lordo ente, per un totale di € 71.855,85.=lordo ente, che trova copertura nel Fondo di Ateneo per la Premialità;

Attestato che la somma totale di € 71.855,85.=l'ordine ente per l'anno accademico 2023/2024, necessaria per procedere all'erogazione del premio e al pagamento degli oneri a carico dell'Ateneo, trova copertura nel Fondo di Ateneo per la Premialità;

<https://www.stat.berkeley.edu/~stark/Seminars/setUCBEECS20.htm>

Stark (2020)

- Valutazioni della didattica da parte degli studenti (SET) sono un metodo usato da un secolo, e negli ultimi decenni rilevanti anche per assunzioni, licenziamenti, tenure, promozioni
- SET sono quantitative, semplici da raccogliere, rapide ed economiche

Student Evaluations of Teaching: Incomplete, or Fail?

University of California, Berkeley
24 April 2020

Philip B. Stark
Department of Statistics
University of California, Berkeley
<http://www.stat.berkeley.edu/~stark> | @philipbstark

Joint work with Anne Boring, Richard Freishtat, Kellie Ottoboni

Proprietà psicometriche?

- Campione **non** rappresentativo (molti studenti rispondono al minimo necessario o evitano se possibile)
- Scala ordinale → non sarebbe lecita la media, eppure si usa universalmente (1 vs 2 è uguale a 5 vs 6?)
- Docenti «polarizzatori» vs «mediocri» (stessa media, varianza diversa): sono uguali?! Eppure si usa universalmente la media...
- Nessuna evidenza di invarianza tra discipline, corsi, istruttori, gradi di istruzione, tipi di corso, modalità di istruzione... eppure si confrontano direttamente le medie come fossero su metrifica universale

<https://www.stat.berkeley.edu/~stark/Seminars/setUCBEECS20.htm>

Stark (2020)

- Valutazioni della didattica da parte degli studenti (SET) sono un metodo usato da un secolo, e negli ultimi decenni rilevanti anche per assunzioni, licenziamenti, tenure, promozioni
- SET sono quantitative, semplici da raccogliere, rapide ed economiche

Student Evaluations of Teaching: Incomplete, or Fail?

University of California, Berkeley
24 April 2020

Philip B. Stark
Department of Statistics
University of California, Berkeley
<http://www.stat.berkeley.edu/~stark> | @philipbstark

Joint work with Anne Boring, Richard Freishtat, Kellie Ottoboni

Cosa dovrebbe essere davvero un insegnamento «efficace»?

- Dovrebbe facilitare l'apprendimento misurato in modo oggettivo (non i voti dati dal docente, ovviamente) → pochi studi, ma ci sono, e mostrano universalmente che Soddisfazione studenti ≠ Insegnamento efficace
- Studenti generalmente poco capaci di valutare quanto abbiano realmente imparato
- Studenti e Docenti NON hanno lo stesso concetto di buon insegnamento [Lauer \(2017\)](#)

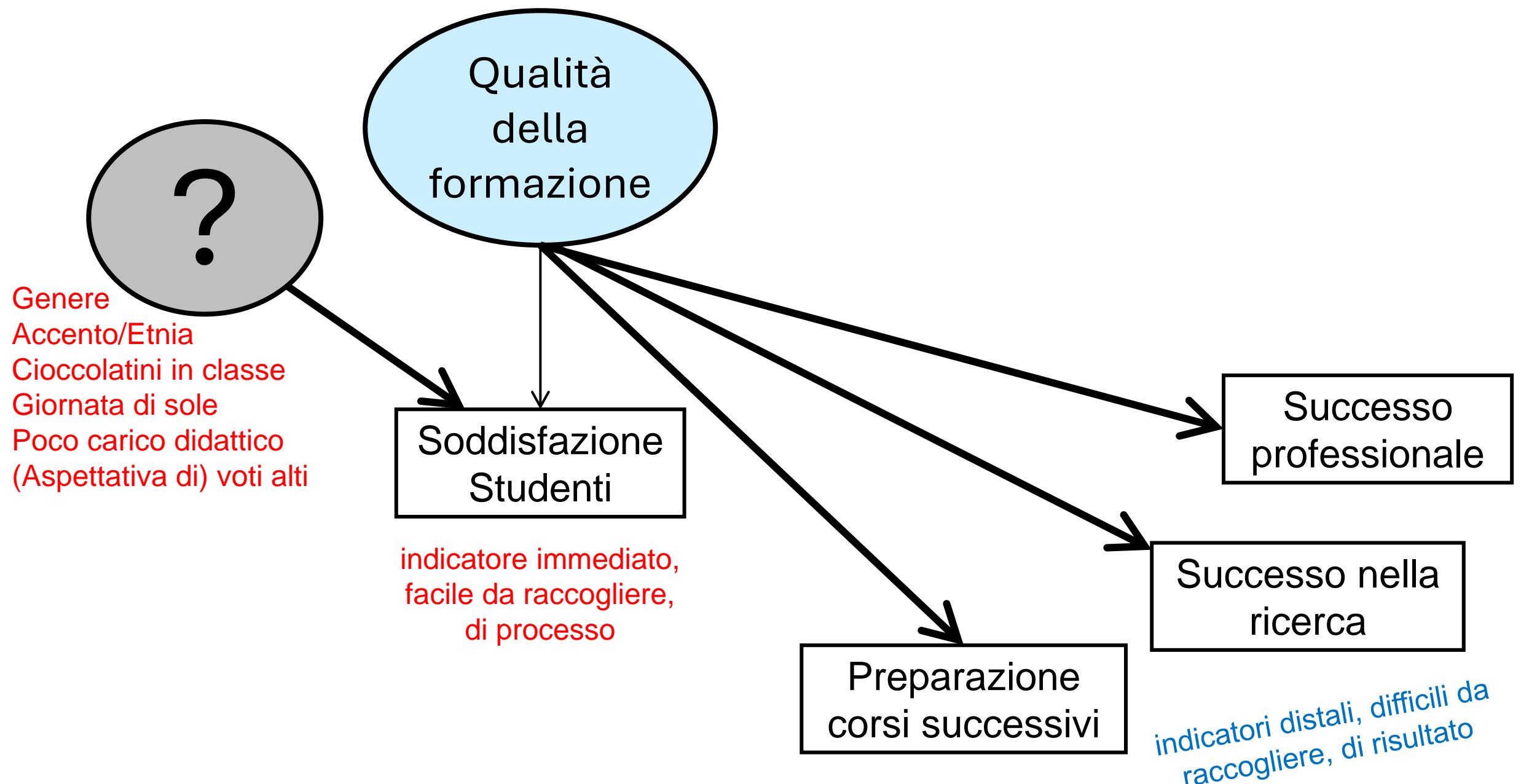
Studenti e Docenti NON hanno lo stesso concetto di buon insegnamento

Table 13.2. What Not Fair Means to Students.

Cosa gli studenti pensano vs cosa i docenti pensano che gli studenti pensino

[Lauer \(2017\)](#)

Comments	Students		Faculty	
	Number	Percentage	Number	Percentage
Plays favorites	119	45.8	20	31.7
Grading problematic	58	22.3	31	49.2
Work is too hard	33	12.7	0	0
Will not “work with you” on problems	32	12.3	0	0
Other	18	6.9	12	19.0
Total	260	100	63	100



Devo massimizzare la Qualità della formazione?

Cerco di migliorare la chiarezza dell'esposizione,
tengo alto il livello delle richieste

Devo massimizzare la Soddisfazione degli studenti?

Cerco di migliorare la chiarezza dell'esposizione,
abbasso il livello delle richieste (aspettativa voti alti è
correlata a soddisfazione), porto cioccolatini in classe, ecc.

Uttl (2021)

4.2 *Chocolates and Cookies*



Two randomized studies demonstrate the power of chocolates and cookies in improving SET ratings. In one of the earlier randomized studies, Youmans and Jee (2007) examined whether providing small chocolate bars would result in higher SET ratings in two statistics and one research methods class. Students who were offered chocolate bars rated their instructor substantially higher than students who were not offered chocolate bars ($d = 0.33$). In another randomized study, Hessler et al. (2018) conducted a single-center randomized control group trial to determine whether the availability of chocolate cookies affects SET ratings. Relative to the no-cookie groups, the cookie groups rated teachers as well as the course material much higher, $d = 0.68$ and $d = 0.66$, respectively. Accordingly, at minimum, chocolates and chocolate cookies are both very effective ways to increase one's SET ratings.

per confronto: Meta-analisi Utzl et al. (2017) suggerisce che rating medio soddisfazione di un corso spieghi al più 1% varianza preparazione accademica oggettiva... $R^2 = 0.01 \rightarrow r = 0.10 \rightarrow$ Cohen's $d = 0.20$

Se anche offrire formazione migliore fosse utile ad aumentare la soddisfazione, l'effetto sarebbe più piccolo che offrire biscotti o cioccolatini

ripeto...

**Se offrire formazione migliore fosse
anche utile ad aumentare la
soddisfazione, il suo effetto sarebbe
più piccolo di quello ottenuto
offrendo biscotti o cioccolatini**

**Ma da dove nasce questa volontà
di «soddisfare» gli studenti?**

**si tratta davvero solo di un modo pratico,
economico, e *rassicurante* di ottenere cattivi
indicatori quantitativi?**

Teaching Sociology, Vol. 30, No. 1 (Jan., 2002), pp. 100-107

**“WE’RE THE CUSTOMER—WE PAY THE TUITION”:
STUDENT CONSUMERISM AMONG
UNDERGRADUATE SOCIOLOGY MAJORS***

MICHAEL DELUCCHI

University of Hawaii—West Oahu

KATHLEEN KORGEN

William Paterson University

BRITISH JOURNAL OF SOCIOLOGY OF EDUCATION, 2017
VOL. 38, NO. 4, 450–467
<http://dx.doi.org/10.1080/01425692.2015.1113856>



Student perceptions of themselves as 'consumers' of higher education

Michael Tomlinson

Southampton Education School, University of Southampton, Southampton, UK

TEACHING IN A CONSUMERISTICALLY CHARGED ENVIRONMENT

William L. Smith
Georgia Southern University

ABSTRACT

This paper argues that student consumerism is currently a driving force in U. S. higher education. Several problems associated with the rise of student consumerism are discussed, including the erosion of the intellectual authority of classroom instructors. The discussion concludes with an appeal to professors to reclaim their authority and thus reshape the organizational culture of the university.

CONVERSATION

A POSTMODERN EXPLANATION OF STUDENT CONSUMERISM IN HIGHER EDUCATION*

MICHAEL DELUCCHI
University of Hawaii-West Oahu

WILLIAM L. SMITH
Georgia Southern University

Fosse anche 100% stipendio docenti pagato dalle rette, se il titolo ha **valore legale** ed è riconosciuto dalla società, sarebbe giusto considerare studenti = clienti?

Lo studente può essere solo un cliente da soddisfare?
O la formazione scolastica-universitaria implica *anche* una **responsabilità verso la società** nel suo complesso?

NEL DETTAGLIO

LE VALUTAZIONI DELL'INSEGNAMENTO SONO ACCURATE?

Measuring the Accuracy of Student Evaluation of Teaching Stanfel, Larry E Journal of Instructional Psychology;
Jun 1, 1995; 22, 2; ProQuest pg. 117

Measuring the Accuracy of Student Evaluations of Teaching

Larry E. Stanfel

Q1 – “*The instructor’s objectives for the course have been made clear*”

→ Quiz intra-corso: **TUTTI** gli studenti identificano correttamente gli obiettivi

Q4 – “*The instructor explains clearly to students how they are evaluated*”

→ Quiz intra-corso: **TUTTI** gli studenti identificano correttamente i criteri

Q5 – “*Tests and written assignments are graded and returned in a reasonable period of time*”

→ **TUTTI** gli esiti erano restituiti alla lezione immediatamente successiva (internet non era ancora una prassi), documentato con firme datate

Measuring the Accuracy of Student Evaluations of Teaching

Larry E. Stanfel

Table 1
Evaluation Results from Class T

	Strongly agree SA	Agree A	Uncertain U	Disagree D	Strongly disagree SD	
Chiarezza obiettivi	Q1 1 6	2.56% 15.38%	7 19	17.95% 48.72%	6 15.38%	82%
Chiarezza criteri valutazione	Q4 1 9	2.56% 23.08%	11 11	28.21% 28.21%	7 17.95%	74%
Restituzione esami	Q5 5 24	12.82% 61.54%	8 1	20.51% 2.56%	1 2.56%	26%

Table 2
Evaluation Results from Class I

	Strongly agree SA	Agree A	Uncertain U	Disagree D	Strongly disagree SD	
Chiarezza obiettivi	Q1 0 2	0% 6.25%	7 8	21.88% 25%	15 46.88%	94%
Chiarezza criteri valutazione	Q4 0 2	0% 6.25%	9 8	28.13% 25%	13 40.63%	94%
Restituzione esami	Q5 1 26	3.13% 81.25%	2 2	6.25% 6.25%	1 3.13%	13%

NEL DETTAGLIO

LE VALUTAZIONI
DELL'INSEGNAMENTO
RISPECCHIANO LA BONTÀ
DELLA PREPARAZIONE?

Student evaluations of teaching (mostly) do not measure teaching effectiveness

Anne Boring^{1,2}, Kellie Ottoboni³, and Philip B. Stark*³

¹OFCE, SciencesPo, Paris, France

²PSL, Université Paris-Dauphine, LEDa, UMR DIAL, Paris, France

³Department of Statistics, University of California, Berkeley, CA, USA

*Corresponding author's e-mail address: pbstark@berkeley.edu

- 5 anni accademici - Institut d'Études Politiques de Paris, Francia
- 6 macro-corsi (*storia, macroeconomia, microeconomia, istituzioni politiche, scienze politiche, sociologia*)
- 1 solo docente “titolare” per ciascun macro-corso: tiene alcune plenarie, prepara esame uguale per tutti → esame valutato in modo anonimo → misura oggettiva dell'apprendimento
- 1194 sezioni totali da 10-24 studenti
- 379 docenti di sezione
- 23001 valutazioni soddisfazione (SET) da 4423 studenti

Boring, Ottoboni, & Stark (2016)

- **SET–Prestazione esame: $r = 0.04$ ($p = 0.09$)**
- Genere docente–Prestazione esame: $r = -0.06$, $p = 0.07$
- Genere docente–SET: $r = 0.09$, $p < 0.001$

L'efficacia didattica operazionalizzata come performance media a **esame finale anonimo e uniforme**

- (i) SET e reale apprendimento **debolmente o per nulla associati**
- (ii) gli studenti di docenti maschi **non** ottengono risultati migliori all'esame,
- (iii) vantaggio maschile nelle SET **non** riflette maggiore efficacia didattica

NEL DETTAGLIO

LE VALUTAZIONI
DELL'INSEGNAMENTO
PRIVILEGIANO VOTO
IMMEDIATO O PREPARAZIONE
PROFONDA?

Carrell & West (2010)

Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors

Situazione simile alla precedente (esami anonimizzati, corsi paralleli e punteggi confrontabili tra corsi)
>> 10,534 studenti (2000-2007)

Value-added = quanto un/a docente prepara per un voto maggiore all'esame immediato (breve termine) o per gli esami successivi (lungo termine)

Studenti apprezzano *value-added* a breve termine = **massimizzare voto all'esame immediato**: $B_{SET} = +0.50$ a $+0.70$ (*stavolta troviamo un effetto!*), che però predice negativamente il *value-added* a lungo termine "profondo", $B_{SET} = -0.40$ a -0.60 , che garantisce successo nei corsi successivi (i due *value-added* correlano $r = -0.14$)

Docenti junior vs senior: +0.79 SD su Value-added a breve termine
MA

Docenti junior vs senior: - 0.66 SD su Value-added a lungo termine

NEL DETTAGLIO

L'INDAGINE DELLE VALUTAZIONI
DELL'INSEGNAMENTO PERMETTE
DI DISTINGUERE ASPETTI
RILEVANTI VS IRRILEVANTI DELLA
QUALITÀ FORMATIVA?

Student Feedback on Teaching in Schools

Using Student Perceptions for the
Development of Teaching and Teachers

OPEN ACCESS

Springer

Valutazioni studenti «indifferenziate»

- *Chiarezza, Gestione classe, Equità valutazione, Interesse per la materia, Cura dell'insegnamento, variabili Socio-affettive (calore, disponibilità, comprensione, percezione di essere dalla parte degli studenti)* sono quasi UNIDIMENSIONALI: $R^2 > 0.70$
- Dimensioni socio-affettive guidano l'impressione totale → sono le più centrali del «fattore comune» ($B = 0.84$ su *overall impression*)

→ Studenti premiano fortemente stili «piacevoli»

On the Validity of Student Evaluation of Teaching: The State of the Art

Pieter Spooren, Bert Brockx, and Dimitri Mortelmans
University of Antwerp

Tra i predittori più sistematicamente associati ai SET: *Course difficulty* (negativamente) e *Grading leniency* (positivamente)

$$r \approx 0.10 - 0.47$$

“...teachers can buy good evaluations by giving high grades; see, e.g., Isely & Singh, 2005;
Langbein, 2008; McPherson, 2006; McPherson & Todd Jewell, 2007”

NEL DETTAGLIO

I SET SONO COMPLETAMENTE
INUTILI?

Wolfram Rollett · Hannah Bijlsma ·
Sebastian Röhl *Editors*

Student Feedback on Teaching in Schools

Using Student Perceptions for the
Development of Teaching and Teachers

OPEN ACCESS

 Springer

Meta-analisi di Röhl (2021) su 20 studi di intervento in contesti scolastici (non universitari) dove si introducono i SET

Apparentemente, introdurre i SET causa un effetto positivo sull'apprendimento oggettivo degli studenti dopo l'introduzione $d = +0.20$ MA è fortemente moderato dal supporto offerto ai docenti: Alto supporto, $d = +0.52$; medio o basso supporto, $d \approx 0$

Alto supporto → presenza di coach formatore per riflessione sui SET, suggerimenti di azioni didattiche pratiche, osservazione in classe

**Ulrich Schimmack**

Amministratore Esperto del gruppo

Persona super attiva

...

· 9 gennaio alle ore 17:43 ·

P.S. Sorry, most students are not stupid. They are lazy. I get it that they want a degree with minimal effort but should we let them evaluate teachers. Do we blame physical trainers when people do not show up in the gym?

Student evaluations are a joke, so let's have a laugh.

"just not a good class, it made no sense to learn from profs own
textbook, which is filled with subjective biases. didn't help
me learn, made me feel more stupid than smart"

Feeling is knowing, dude, feeling is knowing.

Why are we asking stupid students to evaluate teachers. I would also not understand a lecture on quantum physics, but realize that I don't have the ability to evaluate it.

Does your university care about this exercise in futility?

**Ulrich Schimmack**

Amministratore Esperto del gruppo

Persona super attiva

...

· 10 gennaio alle ore 20:38 ·

Follow up on student evaluations.

I am actually recording all my lectures and I asked ChatGPT to rate my lectures on applicable objective dimensions (not students' experiences or satisfaction). No comparison standard, but this is clearly cheaper and better than students' subjective ratings.

Summary Table

Item	Mean (1-5)
Clarity of explanations	4.35
Appropriate assumed knowledge	4.18
Critical evaluation of ideas	4.55
Organizing information	4.28
Evaluating multiple sources	4.12