

# Evaluarea educațională componentă a gândirii pedagogice moderne

Dragos Iliescu, Universitatea din Bucuresti

Istoricul testării  
educationale: de unde  
ne vine și de ce?

# Un scurt istoric al testarii psihometrice

- Cand a aparut primul test?

# Judecatori (12:4-6)

- Un test dezvoltat de Gileaditi pentru a-i descoperi pe Ephraimitii ascunsi in mijlocul lor
- Lungimea testului: 1 item
- Sarcina: sa pronunte cuvantul “Shibboleth”
- Studii de validitate: nici unul
- Consecinta: 42 000 de morti (!)

# Primul test “adevarat”: approx. 2200 ien, China

- Rafinat apoi in timpul dinastiei Quing (approx. 1115) si mai mult in timpul dinastiei Sui (581-618):
- Proceduri de testare formale pentru selectia functionarilor publici
- Un examen de admitere in servciul statului foarte sever: dura zile intregi, uneori candidatii mureau in timpul examinarii
- Evalua 3 mari arii:
  - Cunostinte culturale
  - Planificare si administrare
  - Arte martiale
- Sarcini specific, grupate pe aceste arii: aritmetica, scriere, muzica, calarie, tras cu arcul, manuirea sabiei etc.

# China: inventatorii principiilor din spatele testarii moderne

- Esantionul de comportament:
  - comportamente anume, specific, foarte reduse in amploare, pot fi predictive pentru comportamentul mai larg al unei persoane
- Conditii controlate:
  - se asigura standardizarea conditiilor de evaluare
- Anonimitate pentru asigurarea obiectivitatii
  - Numele candidatilor erau ascunse fata de evaluatori
  - Raspunsurile candidatilor erau copiate de scribi sau alti candidate, astfel incat evaluatorii sa nu poata recunoaste scrisul

# China: influenta globala in testare

- 1791: Un program similar a fost adoptat in Franta, sub influenta lui Voltaire si Quesnay
- 1833: Imperiul britanic trece la un program similar de selectie in “British civil service”, in special pentru functionarii din India
- 1868: SUA adopta in Congres examinarea pentru slujbasii publici, pe aceleasi principii
  - Raportul “Civil Service in the United States” (Jenckes)
  - Civil Service Act (1883)

Cum arata testarea la  
acest moment?

(Psihologica, Educationala)



# 2 mari zone de aplicare a testarii

- Testare **psihologica** vs. Testare **educationala**
  - Ambele masoara constructe intangibile: variabile psihologice vs. cunostinte
  - Ambele au la baza aceleasi principii de psihometrie, de masurare robusta, riguroasa si de “matematizare”
- Dar cele 2 se dezvoltă in modalitati destul de diferite
  - Se adreseaza unor sisteme diferite (clinica/munca vs. scoala)
  - Au utilitate diferita (dizabilitate/performanta vs. educabilitate)
  - Au utilizatori diferiti (“initiati” vs. “publicul larg”)
  - Au consecinte diferite (“decizie” vs. “formare”)

# Cateva lucruri despre testarea educationala

CollegeBoard

The **ACT**<sup>®</sup>

**GRE**<sup>®</sup>



**TOEFL**<sup>®</sup>

SAT<sup>®</sup>



**GMAT**<sup>®</sup>

**NCLEX-RN**<sup>®</sup>



**OCR**<sup>®</sup>  
RECOGNISING ACHIEVEMENT

**AQA**

**edexcel**  
advancing learning, changing lives

Medical College  
Admission Test

**MCAT**



# Inger sau demon?

- Testarea educationala este principalul instrument de management educational in multe state
  - Daca nu masori, nu poti gestiona
  - Masori si compari: copii, profesori, scoli, districte etc
- Testarea educationala este principalul acuzat de starea educatiei ... in SUA si alte state
  - “Teaching to the test”
  - “Overtesting”

# Cateva lucruri despre testarea standardizata in general

# Haideti sa discutam!

- Ar trebuie elevii sa sustina un examen serios pentru a obtine diploma de Bacalaureat?

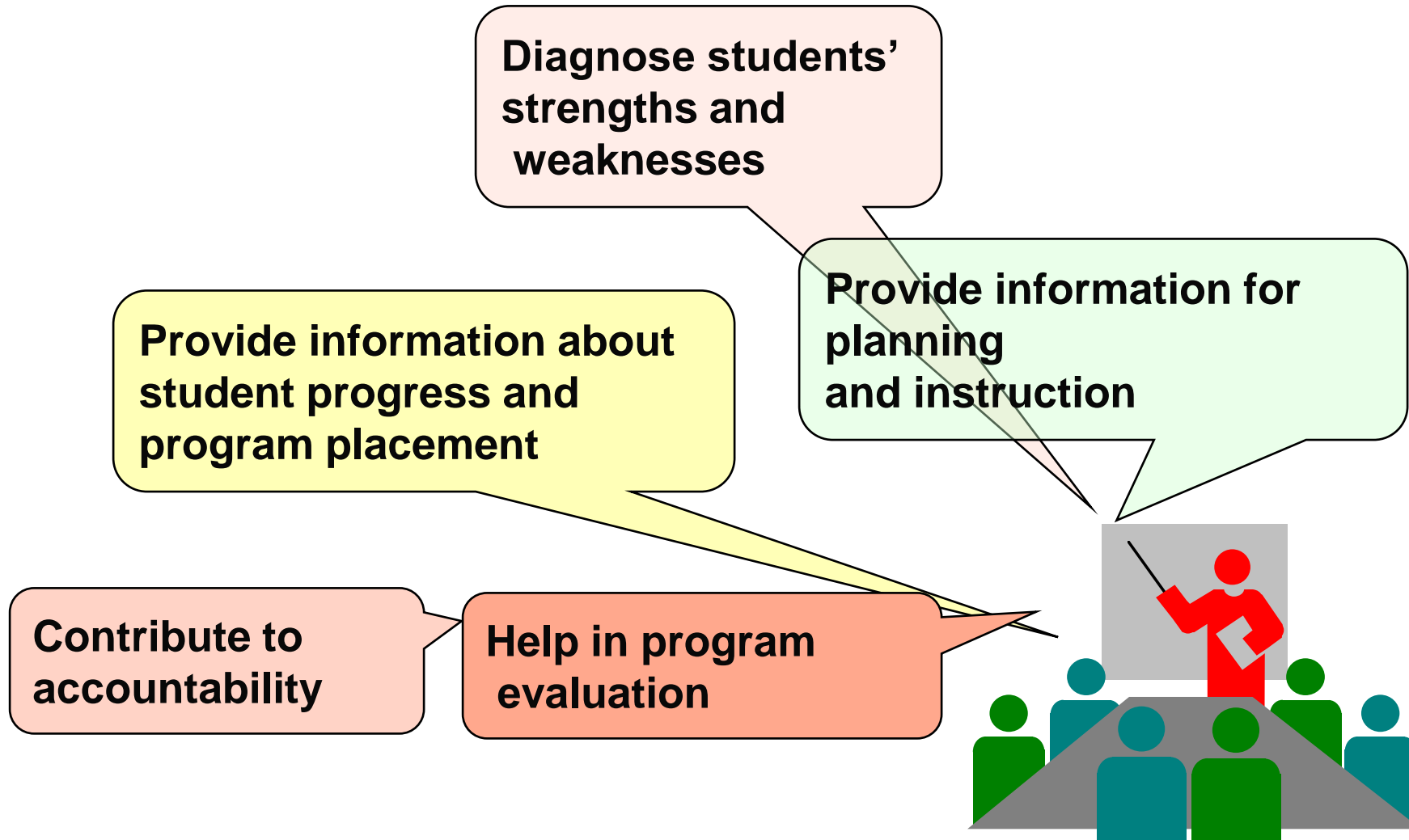
• DA (de ce?):

- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_

• NU (de ce?):

- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_

# Scopurile unui test standardizat



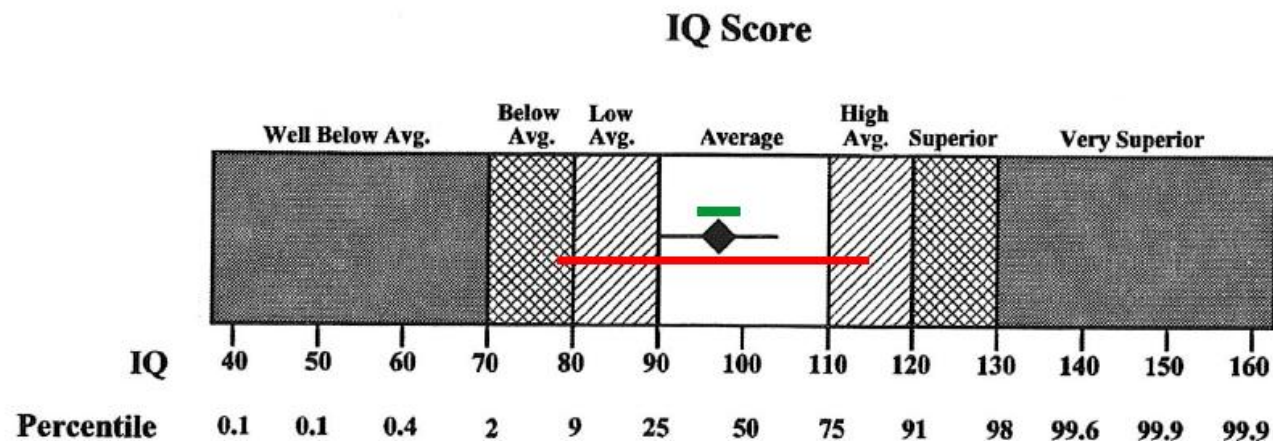
# Totusi, testarea e masurare: Principii psihometrice

- Fidelitate (reliability): in ce fel si de ce?
- Validitate (validity): pentru ce?
- Utilitate (utility): eficient din punctul de vedere al costurilor
- CORECTITUDINE (FAIRNESS):



# Pe scurt despre fidelitate

- Obiectivitate
- Lipsa de eroare: **True score = Observed score + Error**
- Repetabilitate a scorurilor
  - la re-masurare
  - la re-scorare prin alt expert
- De ce e important?
  - (Test de inteligenta: cu 90% confidence interval pentru  $r_{tt}=.50, .70$  si  $.95$ )



# Pe scurt despre validitate

- Valid pentru ce?
- **SAT/GRE/GMAT: valid pentru ce?**
  - Academic achievement (GPA)?
  - Drop-outs?
  - Post-graduation academic achievement (PhD)?
  - Post-graduation income level (\$)?



# Pe scurt despre utilitate

- Cand este un test util?
- Utilitatea te indeamna la o analiza de cost
- O judecam prin prisma a ce obtinem, relative la cost
  - Testul ne ajuta?
  - Dar cu ce costuri?
  - Luam mai multe decizii corecte
  - Daca nu am avea aceasta informatie cate decizii mai gresite am lua?
  - Care ar fi costul acestor decizii gresite?
    - financiare, de imagine etc. etc.
- Controlul costurilor:
  - Testarea are costuri out-of-pocket, pe cand avantajele sunt costuri estimate si posibile...
  - Solutii:
    - Administrare in grup?
    - Scorare automata?
    - Testare online?

# Pe scurt despre corectitudine

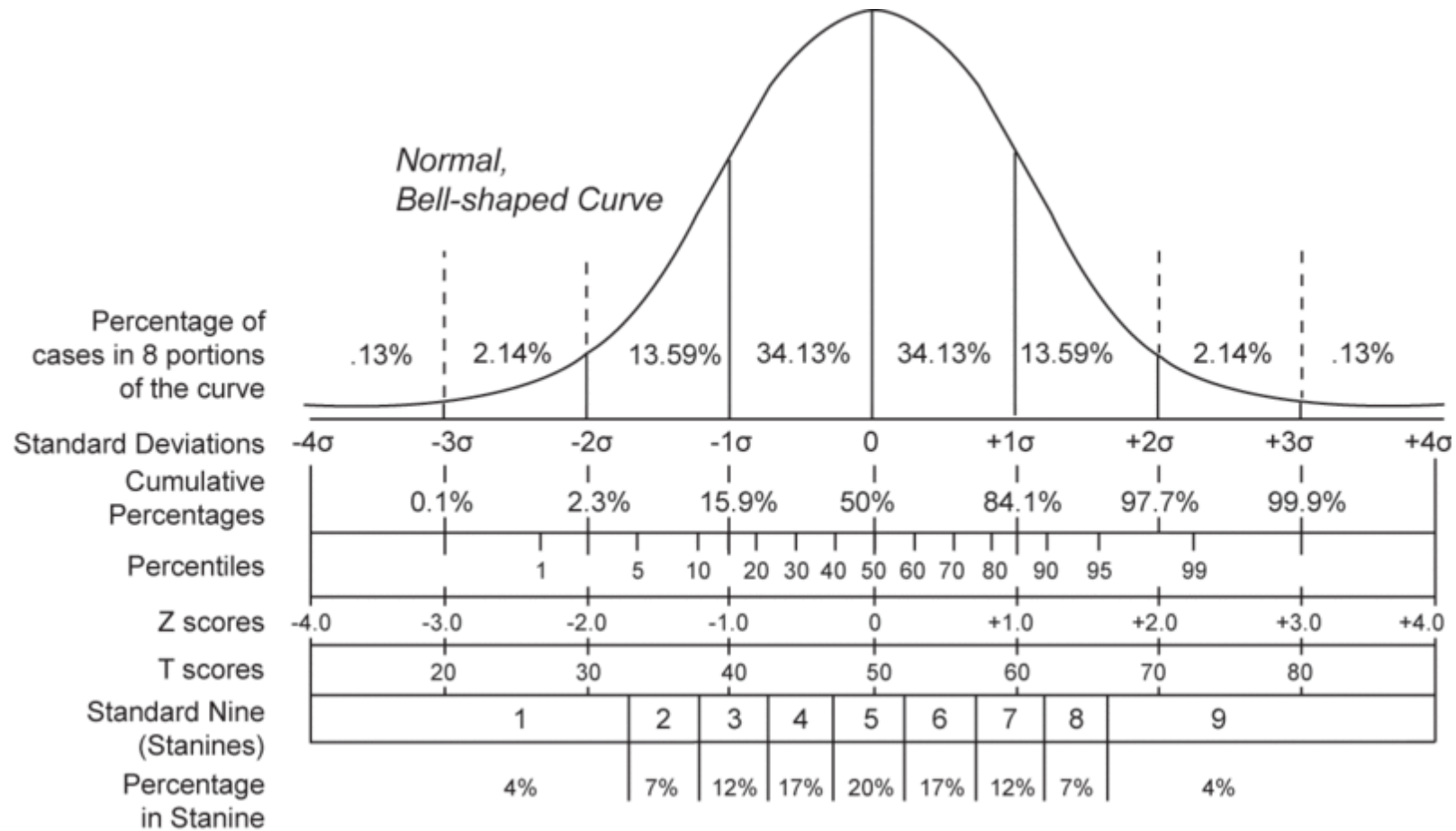
- Si caracteristicile discutate anterior tin de corectitudine
  - Fidelitate: eroarea de masurare nu influenteaza sansele mele si nota sau pozitia mea in ierarhie
  - Validitate: deciziile luate pentru mine, pe baza scorurilor, sunt decizii care chiar decurg din aceste scoruri
- Dar si:
  - sanse/opportunitati egale
  - acomodari rezonabile pentru dizabilitati
  - ...
- Ce este "fair treatment"?
  - "applicant reactions"
  - procesul de testare (drepturi, obligatii, tratament perceput drept corect)

Scurta exemplificare  
a unor principii  
moderne in testarea  
educationala

# Cum se reflecta principiile moderne ale testarii in testarea educationala

- Testare standardizata
- Eforturi de a asigura teste cat mai fidele, valide etc.
- Eforturi de a descuraja “trusul” (cheating)
- Dorinta de a beneficia de tehnologii moderne

# Introducere in masurarea moderna (CTT vs. IRT)



# Beneficii ale adoptarii modelelor IRT

- La nivel de item
  - Item quality / Distractor analysis
  - Item Information Function
  
- La nivel de test
  - Test Information Function
  
- La nivel de program
  - Item banking
  - Test equating
  - Forensic data analysis
  - **Mai multa incredere, mai mult control, mai mult “fairness”**



# PISA/OECD



- Statele care scoreaza bine in PISA au si sisteme bune de evaluare educationala
- *“Research shows that the earlier we measure progress toward meeting educational standards, the more we can strengthen a student’s chances to stay on target to succeed, and, ultimately, be ready for college and work after high school.”*

	Mean score in PISA 2012
OECD average	494
Shanghai-China	613
Singapore	573
Hong Kong-China	561
Chinese Taipei	560
Korea	554
Macao-China	538
Japan	536
Liechtenstein	535
Switzerland	531
Netherlands	523
Estonia	521
Finland	519
Canada	518
Poland	518
Belgium	515
Germany	514
Viet Nam	511
Austria	506
Australia	504
Ireland	501
Slovenia	501
Denmark	500
New Zealand	500
Czech Republic	499
France	495
United Kingdom	494
Iceland	493
Latvia	491
Luxembourg	490

Conclusie

Intrebari la:  
*dragos.iliescu@fpse.unibuc.ro*

