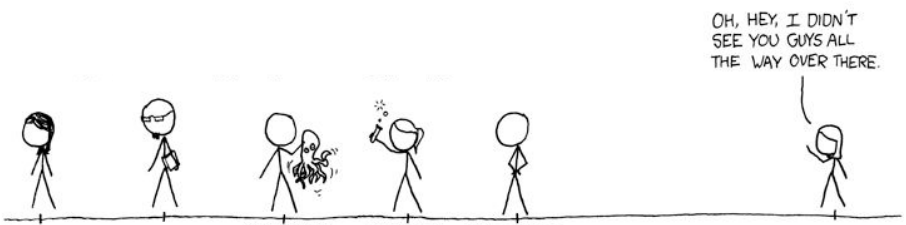
Cognitive psychometrics: ethical considerations

Joachim Vandekerckhove

Spring 2025



Psychometrics and ethics

- Psychometrics is the science of psychological measurement (intelligence, personality, aptitude)

Psychometrics and ethics

- Psychometrics is the science of psychological measurement (intelligence, personality, aptitude)
- Cognitive science provides tools for studying individual differences, underpinning research, diagnosis, and educational assessment

Psychometrics and ethics

- Psychometrics is the science of psychological measurement (intelligence, personality, aptitude)
- Cognitive science provides tools for studying individual differences, underpinning research, diagnosis, and educational assessment
- Psychometric tools impact lives

Psychometrics and ethics

- Psychometrics is the science of psychological measurement (intelligence, personality, aptitude)
- Cognitive science provides tools for studying individual differences, underpinning research, diagnosis, and educational assessment
- Psychometric tools impact lives
- Requires serious ethical scrutiny

Psychometrics and ethics

- Psychometrics is the science of psychological measurement (intelligence, personality, aptitude)
- Cognitive science provides tools for studying individual differences, underpinning research, diagnosis, and educational assessment
- Psychometric tools impact lives
- Requires serious ethical scrutiny
- Easily abused by scientific racists, eugenicists, and other pseudoscientists

The shadow of the past

The allure of quantification

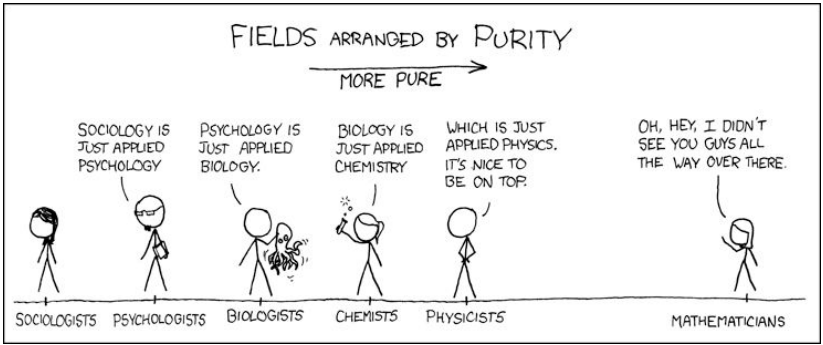
- Late 19th/early 20th century: drive to establish psychology as a "hard science."

The allure of quantification

- Late 19th/early 20th century: drive to establish psychology as a "hard science."
- Early tools to quantify mental faculties reflected societal biases of the era.

The allure of quantification

- Late 19th/early 20th century: drive to establish psychology as a "hard science."
- Early tools to quantify mental faculties reflected societal biases of the era.



Scientific racism and eugenics

- Both are based on the idea that certain groups are inferior to others in some sense.
- You can "prove" all sorts of inferiority/superiority claims if you are willing to cherry-pick data or take invalid liberties with measurement and statistics.

Definition (Scientific racism)

Use of pseudoscientific practices and claims to justify racial hierarchies or discrimination. Measurement of skull sizes (phrenology) within and between populations was used to conclude that some races are inherently less intelligent.

Scientific racism and eugenics

- Both are based on the idea that certain groups are inferior to others in some sense.
- You can "prove" all sorts of inferiority/superiority claims if you are willing to cherry-pick data or take invalid liberties with measurement and statistics.

Definition (Eugenics)

The belief or practice of improving the genetic quality of the human population through selective breeding or sterilization. Some governments have enacted policies to sterilize individuals deemed "unfit" to reproduce (e.g., based on mental health diagnoses), and make health care more easily accessible for certain subpopulations than others, etc.

Case study 1: eugenics and intelligence testing

- Key figures: Galton, Goddard, Terman, Brigham.

Case study 1: eugenics and intelligence testing

- Key figures: Galton, Goddard, Terman, Brigham.
- Misuse of early IQ tests to argue for intellectual inferiority of certain racial/ethnic groups and the "feeble-minded."

Case study 1: eugenics and intelligence testing

- Key figures: Galton, Goddard, Terman, Brigham.
- Misuse of early IQ tests to argue for intellectual inferiority of certain racial/ethnic groups and the "feeble-minded."
- **Consequences:** discriminatory immigration policies (e.g., US Immigration Act of 1924), forced sterilization programs.

Case study 1: eugenics and intelligence testing

- Key figures: Galton, Goddard, Terman, Brigham.
- Misuse of early IQ tests to argue for intellectual inferiority of certain racial/ethnic groups and the "feeble-minded."
- **Consequences:** discriminatory immigration policies (e.g., US Immigration Act of 1924), forced sterilization programs.
- **Ethical breaches:** gross cultural bias, lack of validity, confirmation bias, justification for harmful social engineering.

Case study 1: eugenics and intelligence testing

- Key figures: Galton, Goddard, Terman, Brigham.
- Misuse of early IQ tests to argue for intellectual inferiority of certain racial/ethnic groups and the "feeble-minded."
- **Consequences:** discriminatory immigration policies (e.g., US Immigration Act of 1924), forced sterilization programs.
- **Ethical breaches:** gross cultural bias, lack of validity, confirmation bias, justification for harmful social engineering.
- **Today:** This still just happens, it's not even particularly more subtle (e.g., Richard Lynn's "The Intelligence of Nations" was published in 2019; Lynn also argued for sex differences in intelligence).

Case study 2: cultural bias in test development

- Many early tests normed on specific, often privileged, populations (e.g., white, middle-class).

Case study 2: cultural bias in test development

- Many early tests normed on specific, often privileged, populations (e.g., white, middle-class).
- Test items often reflected specific cultural knowledge, not universal cognitive ability.

Case study 2: cultural bias in test development

- Many early tests normed on specific, often privileged, populations (e.g., white, middle-class).
- Test items often reflected specific cultural knowledge, not universal cognitive ability.
- **Consequences:** mislabeling individuals from diverse backgrounds, flawed educational tracking, perpetuation of systemic inequalities.

Case study 2: cultural bias in test development

- Many early tests normed on specific, often privileged, populations (e.g., white, middle-class).
- Test items often reflected specific cultural knowledge, not universal cognitive ability.
- **Consequences:** mislabeling individuals from diverse backgrounds, flawed educational tracking, perpetuation of systemic inequalities.
- **Ethical breaches:** lack of fairness and equity, failure to consider construct-irrelevant variance.

Ethical failures are also scientific failures

Many scientific questions are really questions about values

Ethical failures are also scientific failures

Many scientific questions are really questions about values

- For example, whether an AI is “really intelligent” or “has consciousness” is really a question about how we should treat it, whether we should give it rights, etc.

Ethical failures are also scientific failures

Many scientific questions are really questions about values

- For example, whether an AI is “really intelligent” or “has consciousness” is really a question about how we should treat it, whether we should give it rights, etc.
- The same is true for classical intelligence testing of people

Ethical failures are also scientific failures

Many scientific questions are really questions about values

- For example, whether an AI is “really intelligent” or “has consciousness” is really a question about how we should treat it, whether we should give it rights, etc.
- The same is true for classical intelligence testing of people
- Very first IQ tests were to determine which children merited which education

Ethical failures are also scientific failures

Many scientific questions are really questions about values

- For example, whether an AI is “really intelligent” or “has consciousness” is really a question about how we should treat it, whether we should give it rights, etc.
- The same is true for classical intelligence testing of people
- Very first IQ tests were to determine which children merited which education
- We still give different legal rights to people based on their IQ scores

Ethical failures are also scientific failures

Many scientific questions are really questions about values

- For example, whether an AI is "really intelligent" or "has consciousness" is really a question about how we should treat it, whether we should give it rights, etc.
- The same is true for classical intelligence testing of people
- Very first IQ tests were to determine which children merited which education
- We still give different legal rights to people based on their IQ scores

Even if that is what you want to do, tests need to be valid – they need to *measure what they purport to measure* – and that has often not been the case.

Ethical failures are also scientific failures

Many scientific questions are really questions about values

- For example, whether an AI is "really intelligent" or "has consciousness" is really a question about how we should treat it, whether we should give it rights, etc.
- The same is true for classical intelligence testing of people
- Very first IQ tests were to determine which children merited which education
- We still give different legal rights to people based on their IQ scores

Even if that is what you want to do, tests need to be valid – they need to *measure what they purport to measure* – and that has often not been the case.

(E.g., Lynn's work compared children in remedial education programs to children in regular high schools.)

Other historical ethical issues

- **Over-generalization and labeling:** treating scores as definitive, immutable measures of an individual's worth or potential.

Other historical ethical issues

- **Over-generalization and labeling:** treating scores as definitive, immutable measures of an individual's worth or potential.
- **Lack of informed consent:** individuals often tested without clear understanding of the test's purpose or how results would be used.

Other historical ethical issues

- **Over-generalization and labeling:** treating scores as definitive, immutable measures of an individual's worth or potential.
- **Lack of informed consent:** individuals often tested without clear understanding of the test's purpose or how results would be used.
- **Transparency issues:** limited access to test content or understanding of scoring for test-takers.

Contemporary ethical challenges

Transition: lessons learned or repeated?

- While the field has evolved, many historical concerns persist in new forms.

Transition: lessons learned or repeated?

- While the field has evolved, many historical concerns persist in new forms.
- New technologies also introduce novel ethical dilemmas.

Transition: lessons learned or repeated?

- While the field has evolved, many historical concerns persist in new forms.
- New technologies also introduce novel ethical dilemmas.
- Any number of computer image processing studies are basically phrenology (physiognomy).

> J Pers Soc Psychol. 2018 Feb;114(2):246-257. doi: 10.1037/pspa0000098.

Deep neural networks are more accurate than humans at detecting sexual orientation from facial images

Yilun Wang [1], Michal Kosinski [1]

Affiliations – collapse

Affiliation

1 Graduate School of Business, Stanford University.

Fairness, equity, and bias today

- Ensuring tests are fair across diverse cultural, linguistic, and socioeconomic groups remains a major challenge.

Fairness, equity, and bias today

- Ensuring tests are fair across diverse cultural, linguistic, and socioeconomic groups remains a major challenge.
- Tools like differential item functioning (DIF) analysis help detect bias.

$$\theta_{pi} = \beta_0 + \beta_A \times \text{Ability}_p + \beta_G \times \text{Group}_p \times \text{DIF}_{Gi} + \varepsilon_p$$

Fairness, equity, and bias today

- Ensuring tests are fair across diverse cultural, linguistic, and socioeconomic groups remains a major challenge.
- Tools like differential item functioning (DIF) analysis help detect bias.

$$\theta_{pi} = \beta_0 + \beta_A \times \text{Ability}_p + \beta_G \times \text{Group}_p \times \text{DIF}_{Gi} + \varepsilon_p$$

- Challenges in test translation and cross-cultural adaptation.

Informed consent and test user competence

- Test-takers (or guardians) must understand:

Informed consent and test user competence

- Test-takers (or guardians) must understand:
  - Purpose and nature of the test.

Informed consent and test user competence

- Test-takers (or guardians) must understand:
  - Purpose and nature of the test.
  - How results will be used and their limitations.

Informed consent and test user competence

- Test-takers (or guardians) must understand:
  - Purpose and nature of the test.
  - How results will be used and their limitations.
  - Rights (e.g., to receive feedback, refuse testing in some contexts).

Informed consent and test user competence

- Test-takers (or guardians) must understand:
  - Purpose and nature of the test.
  - How results will be used and their limitations.
  - Rights (e.g., to receive feedback, refuse testing in some contexts).
- Critical for vulnerable populations (minors, patient groups, the incarcerated).

Informed consent and test user competence

- Test-takers (or guardians) must understand:
  - Purpose and nature of the test.
  - How results will be used and their limitations.
  - Rights (e.g., to receive feedback, refuse testing in some contexts).
- Critical for vulnerable populations (minors, patient groups, the incarcerated).
- **Common ethical principles:** autonomy, beneficence, non-maleficence.

High-stakes testing

- Major life decisions (admissions, employment, licensure) remain heavily reliant on single test scores.

High-stakes testing

- Major life decisions (admissions, employment, licensure) remain heavily reliant on single test scores.
- Potential for "teaching to the test," (e.g., SAT prep), anxiety, narrowing of curricula.

High-stakes testing

- Major life decisions (admissions, employment, licensure) remain heavily reliant on single test scores.
- Potential for "teaching to the test," (e.g., SAT prep), anxiety, narrowing of curricula.
- Balancing test utility with potential for individual and societal harm.

High-stakes testing

- Major life decisions (admissions, employment, licensure) remain heavily reliant on single test scores.
- Potential for "teaching to the test," (e.g., SAT prep), anxiety, narrowing of curricula.
- Balancing test utility with potential for individual and societal harm.
- **Ethical principle:** justice, scrutiny of consequences.

Big data and machine learning in assessment

- Large-scale collection of behavioral and performance data

Big data and machine learning in assessment

- Large-scale collection of behavioral and performance data
- Privacy concerns with data aggregation and sharing

Big data and machine learning in assessment

- Large-scale collection of behavioral and performance data
- Privacy concerns with data aggregation and sharing
- Potential for automated profiling and discrimination

Big data and machine learning in assessment

- Large-scale collection of behavioral and performance data
- Privacy concerns with data aggregation and sharing
- Potential for automated profiling and discrimination
- Risk of reinforcing existing societal biases

Big data and machine learning in assessment

- Large-scale collection of behavioral and performance data
- Privacy concerns with data aggregation and sharing
- Potential for automated profiling and discrimination
- Risk of reinforcing existing societal biases
- Need for transparency in algorithmic decision-making

The rise of AI in psychometrics

- AI for automated scoring, adaptive testing, profiling from digital footprints.

The rise of AI in psychometrics

- AI for automated scoring, adaptive testing, profiling from digital footprints.
- Risk of **algorithmic bias**: AI learning and amplifying existing societal biases present in training data.

The rise of AI in psychometrics

- AI for automated scoring, adaptive testing, profiling from digital footprints.
- Risk of **algorithmic bias**: AI learning and amplifying existing societal biases present in training data.
- The "black box" problem: lack of transparency in AI-driven assessment decisions.

Ethical measurement practice

A responsible path

- The field of psychometrics actively works to address ethical concerns.

A responsible path

- The field of psychometrics actively works to address ethical concerns.
- This involves professional guidelines, rigorous standards, and ongoing research.

A responsible path

- The field of psychometrics actively works to address ethical concerns.
- This involves professional guidelines, rigorous standards, and ongoing research.
- There is a lot of work left to do.

A responsible path

- The field of psychometrics actively works to address ethical concerns.
- This involves professional guidelines, rigorous standards, and ongoing research.
- There is a lot of work left to do.
- Cognitive modelers need to keep track.

Test development and validation

- Rigorous validation for specified purposes and populations is crucial.

Test development and validation

- Rigorous validation for specified purposes and populations is crucial.
- Continuous research into bias detection and reduction methodologies.

Test development and validation

- Rigorous validation for specified purposes and populations is crucial.
- Continuous research into bias detection and reduction methodologies.
- Promoting transparency in test development and scoring processes.

Your role: critical consumer and future professional

As cognitive scientists, you may use, develop, or critique psychometric tools.

Your role: critical consumer and future professional

As cognitive scientists, you may use, develop, or critique psychometric tools.

It is **your job** to critically evaluate claims made about tests and their uses:

Your role: critical consumer and future professional

As cognitive scientists, you may use, develop, or critique psychometric tools.

It is **your job** to critically evaluate claims made about tests and their uses:

- If you develop a test or measurement tool, consider if it measures what it claims to measure *even outside the context of your studies*?

Your role: critical consumer and future professional

As cognitive scientists, you may use, develop, or critique psychometric tools.

It is **your job** to critically evaluate claims made about tests and their uses:

- If you develop a test or measurement tool, consider if it measures what it claims to measure *even outside the context of your studies*?
- Understand the limitations inherent in any measurement instrument.

Your role: critical consumer and future professional

As cognitive scientists, you may use, develop, or critique psychometric tools.

It is **your job** to critically evaluate claims made about tests and their uses:

- If you develop a test or measurement tool, consider if it measures what it claims to measure *even outside the context of your studies*?
- Understand the limitations inherent in any measurement instrument.
- Advocate for ethical practices and consider societal impact of your contributions and omissions.

Cognitive psychometrics: ethical considerations

Joachim Vandekerckhove

Spring 2025