
Selecting future doctors

An overview

and an evidence based model for selection

David Powis, Miles Bore, Don Munro
Newcastle Innovation

Summary

- there should be more emphasis on identifying the potentially inadequate
 - use well-researched and validated measures
 - personality tests have a valuable role to play
 - apply a sequential filtering model
 - bulk screening through to individual assessment
 - reserve interview for doing what it can do uniquely well
-

What do we know about doctors?

most are adequate,

but some

- **don't communicate appropriately**
 - junior colleagues, peers, patients, patients' families
 - **behave unethically / unprofessionally**
 - attract complaints and litigation
(53 per '000 insured doctors; Australia 2000 – 2004)
 - attract notoriety
(e.g. Dr Shipman, Dr Martin in UK; Dr Patel, Dr Reeves in Australia)
-

-
- **have high prevalence of depression, anxiety**
 - 37% of primary care physicians
high levels of ‘psychological discomfort’
associated with practice (Spain)
 - 1 in 5 hospital doctors (UK survey)
symptoms of **‘such severe depression and anxiety**
that they warranted psychiatric care, had it been sought’
-

- **high incidence of substance abuse**

- 1 in 15 doctors become dependent on alcohol or drugs during career (UK, GMC, 2005)
- 1400 doctors disciplined for substance abuse (USA 1999 – 2004)

- **high suicide rate**

- male doctors 1.4 X general population
- female doctors 2.3 X general population

(meta-analysis of studies of physician suicide rates
Schernhammer & Colditz, *Am J Psychiatry* 161, 2004)

– **high levels of burnout / distress**

- stressful work environment
- long hours
- conflict between work and personal life
- individual psychological vulnerability

What do we know about medical students?

most are adequate,
but some cause concern

- one UK medical school identified 10 – 15% of each intake (over 5 years) as ‘strugglers’

J Yates & D James, *BMJ* 332 (2006) 1009-1013

- medical school teachers have all observed ‘unsuitable’ individuals
-

A survey of professionals

clinical staff asked to list undesirable personal characteristics they *had observed* in medical students (and colleagues)

(n = 190 respondents; Newcastle)

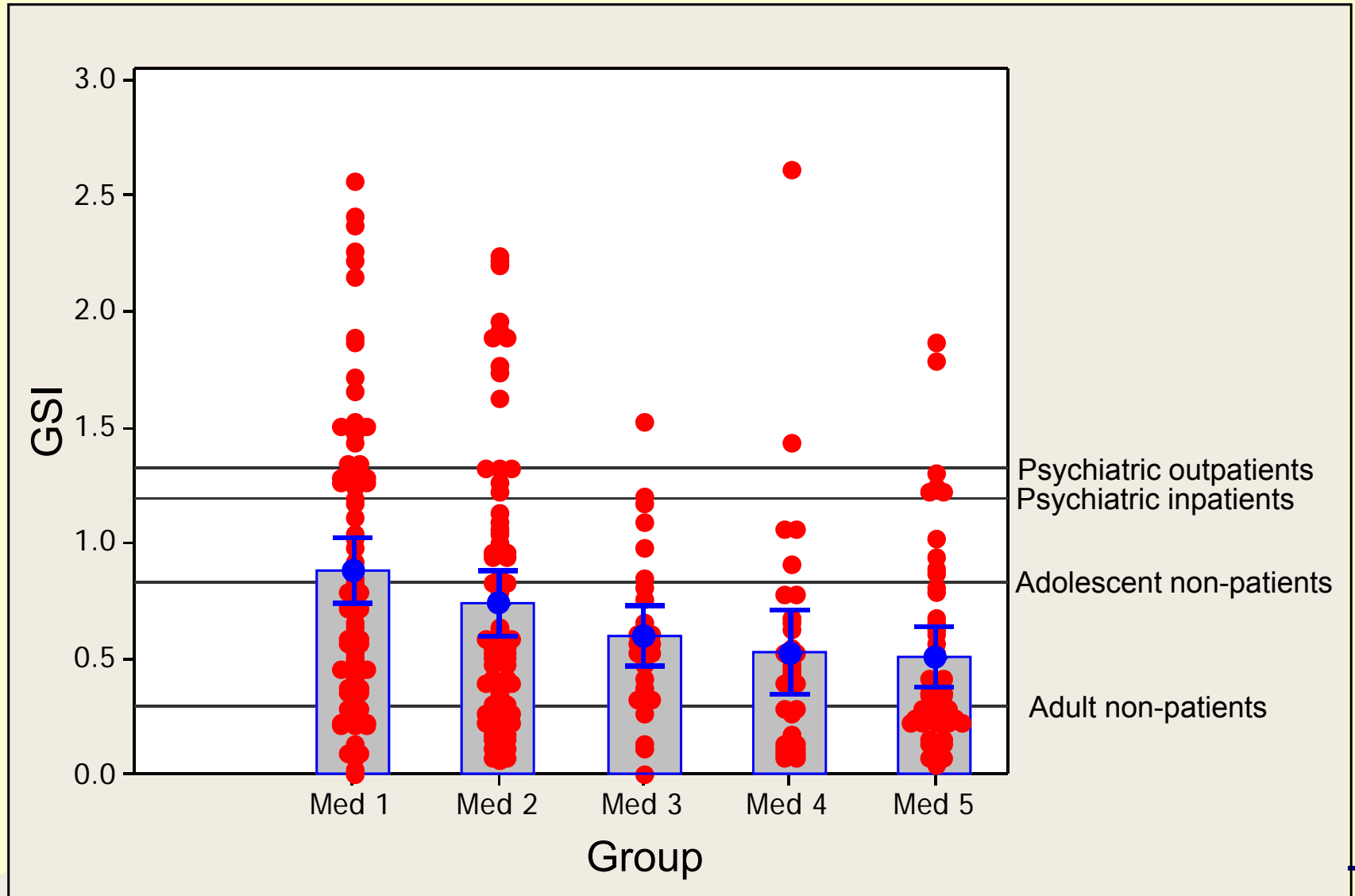
their descriptors of inappropriate medical student behaviours and attitudes:

- arrogant
- power-seeking
- inflexible
- defensive
- dishonest
- patronising
- brash
- egocentric
- isolated
- insensitive
- self-centred
- uncaring
- indifferent
- selfish
- antisocial
- amoral
- devious
- prejudiced
- flippant
- rude
- aggressive
- condescending
- rigid attitudes
- judgemental

Problems observed in medical students

- **high prevalence of psychiatric morbidity**
 - 26% in final year of one Australian medical school
 - 30% of 1st and 2nd year medical students suffered high levels of emotional distress (Spain)
 - 31% of 1st year medical students at Newcastle had symptom level score on Brief Symptom Inventory \geq BSI norm for adult psychiatric in-patients
-

Global Severity Index compared to BSI norms



Why might this be the case?

- ? undue focus on positive attributes of applicants during selection
 - and failure to adequately consider the warning signs
-

Selection procedures

most procedures are focused on seeking the

- highest academic achievers
- best cognitive skills
- and (maybe) best communication skills etc

‘the brightest and best’, ‘the cream of the cream’

Nothing wrong with this, except

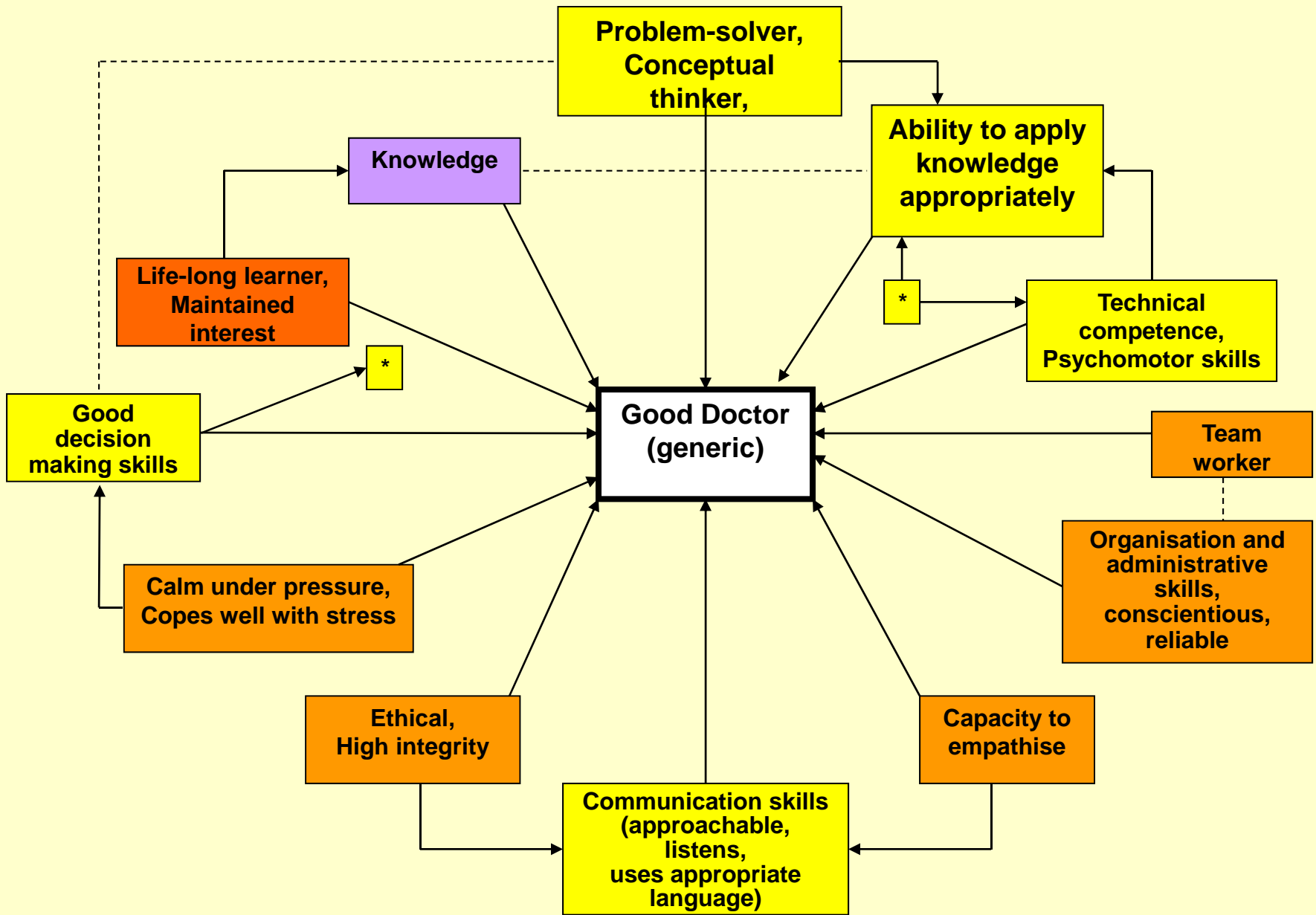
perhaps **the emphasis...**

- concentrating on trying to discriminate
the 'best' from
the 'next best' from
'above average' etc

is wasted effort if most students will be just fine

- should we expend more effort trying to
identify the potentially inadequate?
-

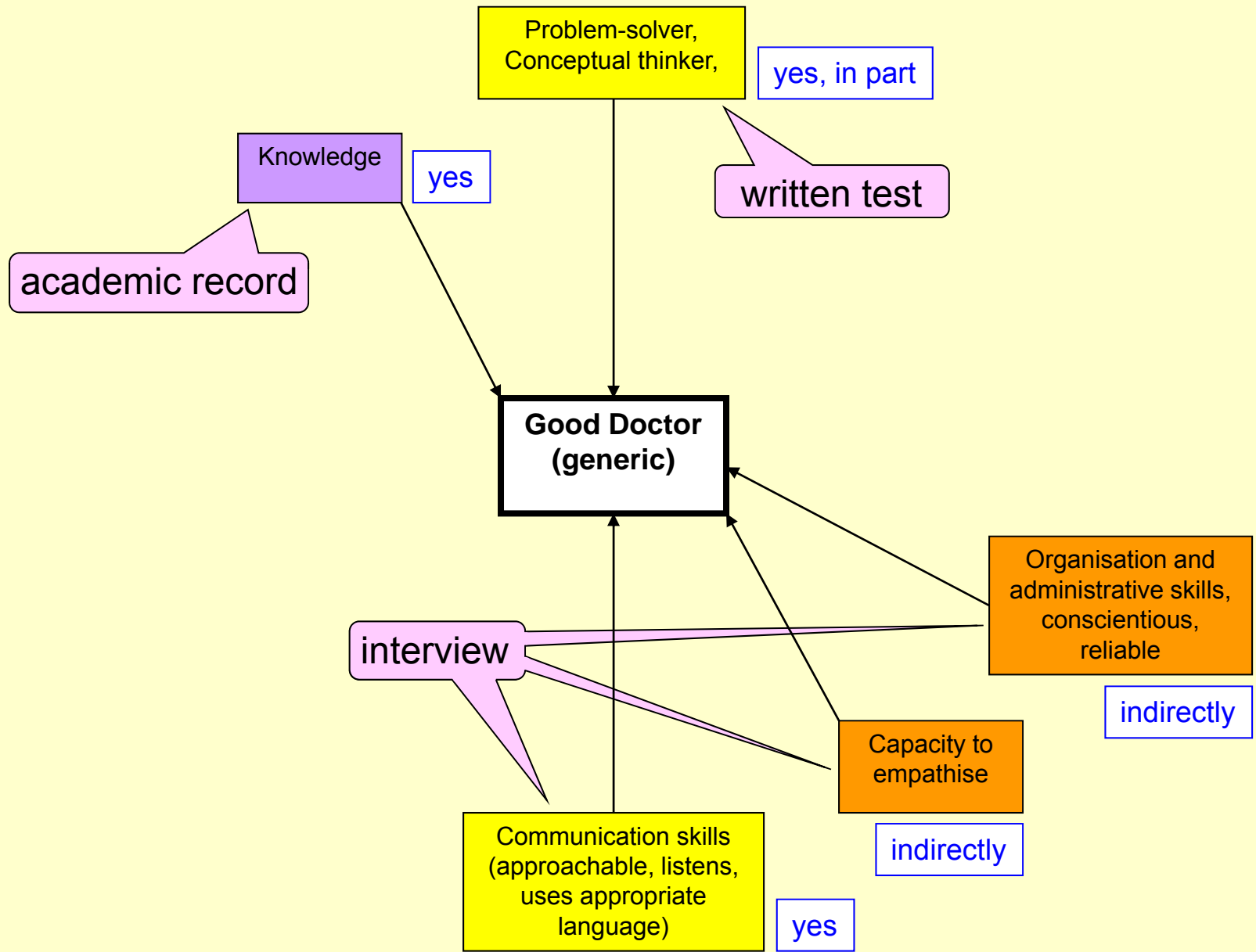
What is a good doctor?



PPIK theory:



Which of these do
current selection procedures
look for?



Tests used in Australia

UMAT

- logical reasoning & problem solving
- ‘understanding people’
- non-verbal (spatial) reasoning

GAMSAT

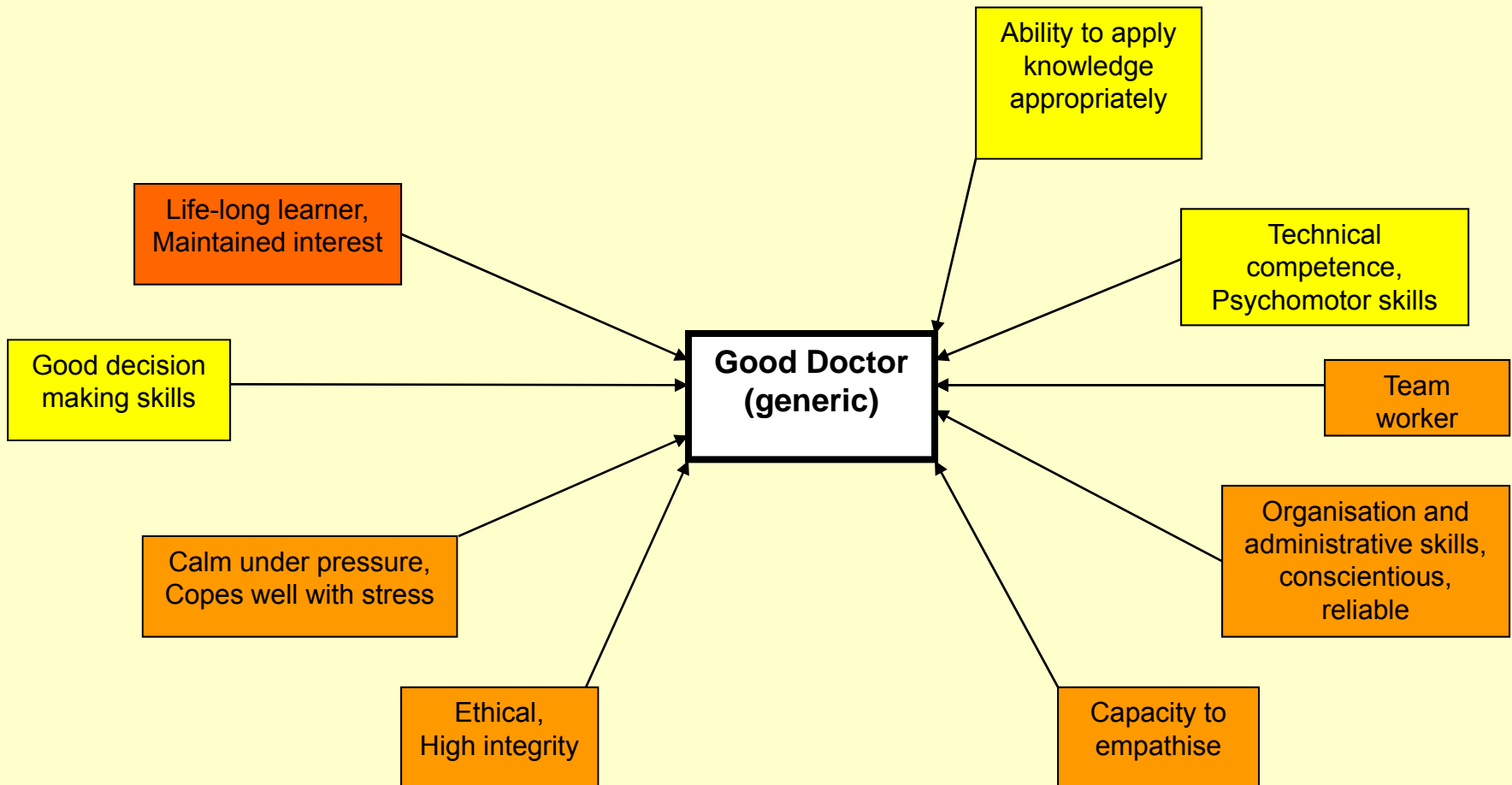
- reasoning in physical and biological sciences
 - reasoning in the humanities
 - creative writing
-

Tests used in UK

UKCAT

- decision analysis
 - abstract reasoning
 - verbal reasoning
 - quantitative reasoning
-

Existing tests DO NOT measure:



How useful are they?

outcome studies disappointing

- **UMAT** - no published studies
informal reports only
no correlation with
medical school assessment measures
 - **GAMSAT** - 2 published studies
no correlation with
medical school assessment measures
-

Interviews

most medical schools interview;

some in Australia have abandoned or downgraded the interview

- only 1 study has shown interview scores predict medical school outcome (Newcastle 1988)
 - ‘withdrawal/exclusion’ vs ‘graduation with honours’
 - other studies have shown modest correlations with clinical assessment
-

The *Personal Qualities Assessment*

PQA

PQA research base

Extensive research studies over 12 years

- > 28,000 individuals
- England, Scotland, Australia, Israel, Sweden, Japan, Taiwan, Hong Kong, Fiji, Canada

Established

- high reliability of all subtests
 - construct validity
-

What are the tests?

PQA

Test 1 Mental Agility Test

high level cognitive skills

Test 2 'Mojac' Questionnaire

balancing individuals' freedoms with society's rules

Test 3 NACE Questionnaire

emotional involvement versus detachment

Test 4 Personal Characteristics Inventory

self-control and resilience

PQA Mental Agility Test (MAT)

A **general** cognitive skills test of high difficulty

1 hour; 45 items; MCQ format

- verbal
 - numerical
 - spatial
-

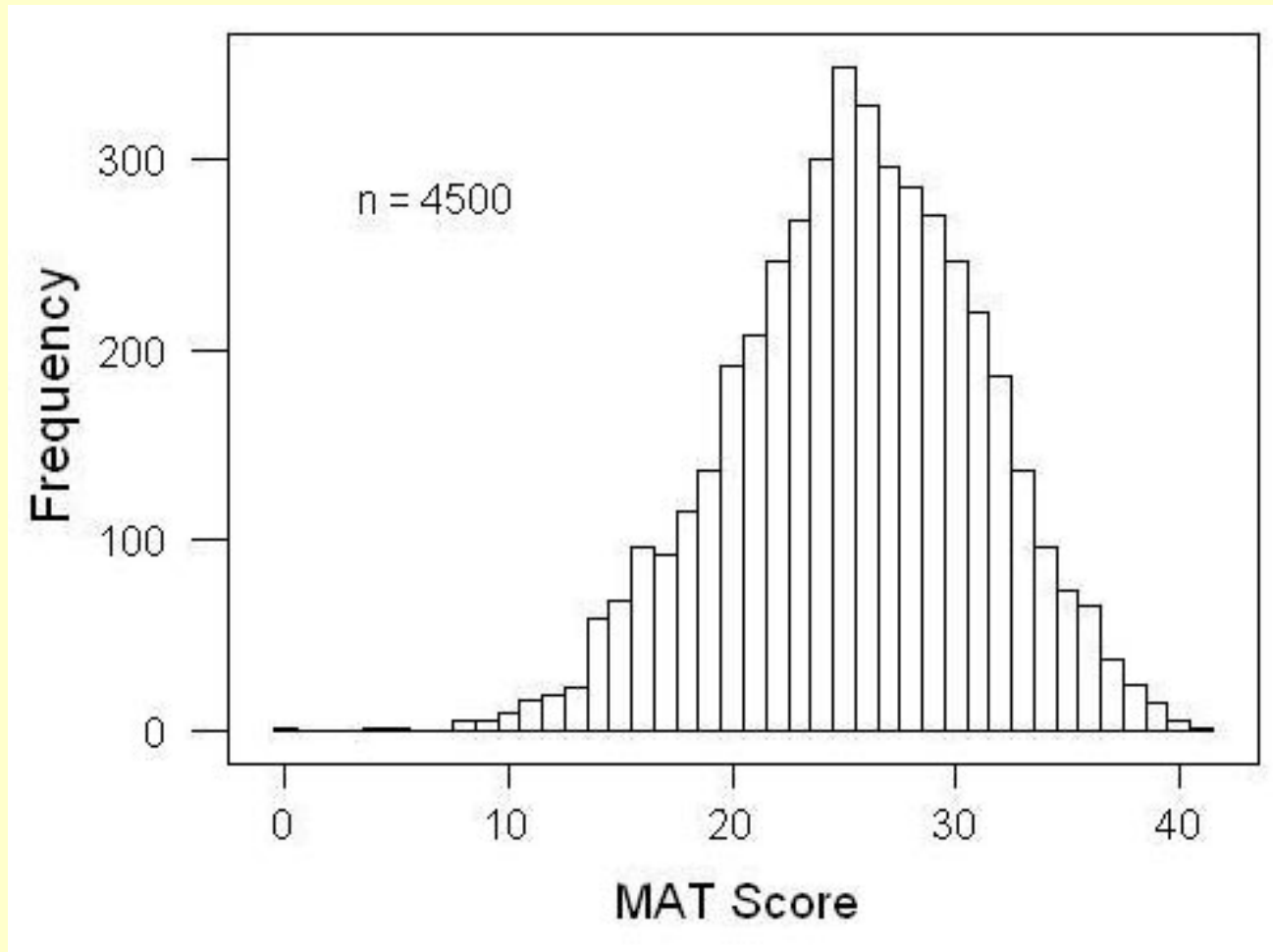
General Cognitive Ability (GCA)

- **moderate to strong predictor of occupational attainment and performance within occupations***

*Schmidt FL, Hunter J. General mental ability in the world of work: Occupational attainment and job performance. *J Pers & Soc Psych*, 2004; **86**; 162-73.

*Brown KG, Le H, Schmidt FL. Specific aptitude theory revisited: Is there incremental validity for training performance? *Int J Selection and Assessment* 2006; **14**: 87-100.

PQA MAT score distribution



45 items; mean score 25.5 ± 5.6 S.D; range 0 - 41

PQA MAT reliability

Good internal consistency:

Cronbach's alpha 0.73

PQA ‘Mojac’ questionnaire

a series of moral dilemmas requiring balancing individual freedoms against society’s rules

- 49 items; approx. 30min
- a dissonance model of ethical decision making

Bore, 2001; Bore *et al.*, 2005

'Ethics' – nurture or nature?

- knowledge
 - attitudes and values
 - sensitivity
 - reasoning
 - intrinsic orientation
 - decision making style
 - behaviour
-

'Ethics' – nurture or nature

- knowledge
- attitudes and values
- sensitivity
- reasoning
- intrinsic orientation
- decision making style
- behaviour

learned on course

measured by Mojac

desired outcome

'Mojac': individuals' freedom in the context of society's rules

**individual
orientation**



**dual
orientation**



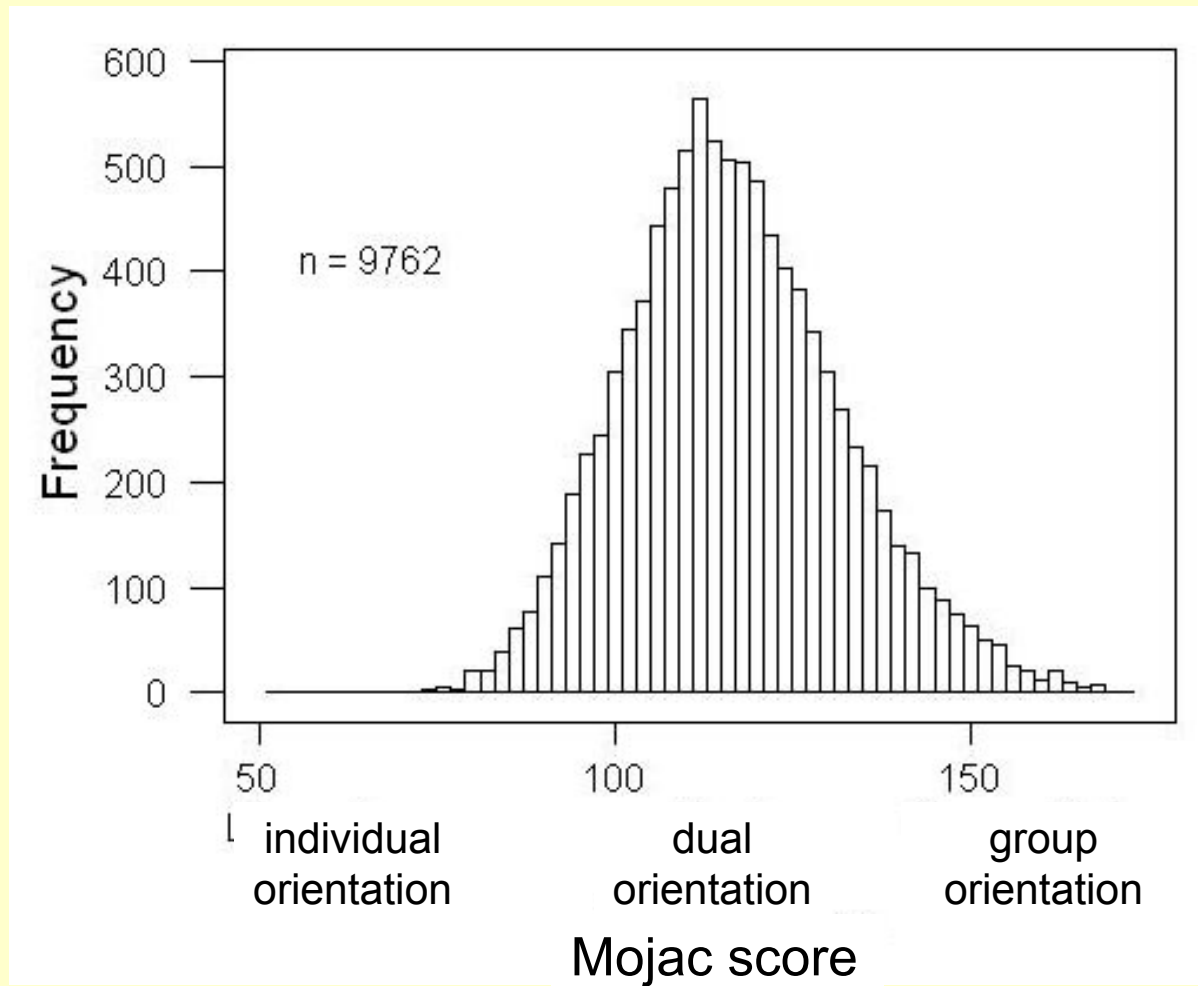
**group
orientation**

values individuals' freedom much higher than duty to the group

values equally individuals' freedom and duty to group

values duty to the group much higher than individuals' freedom

'Mojac': distribution of scores



PQA

'Mojac'

18,000+ individuals tested

- gender differences
 - age effects
 - effects of socio-economic status
 - effects of ethnic background and culture
Australia UK Sweden Japan Singapore HK Taiwan Israel
 - stable cohort mean score during medical education
-

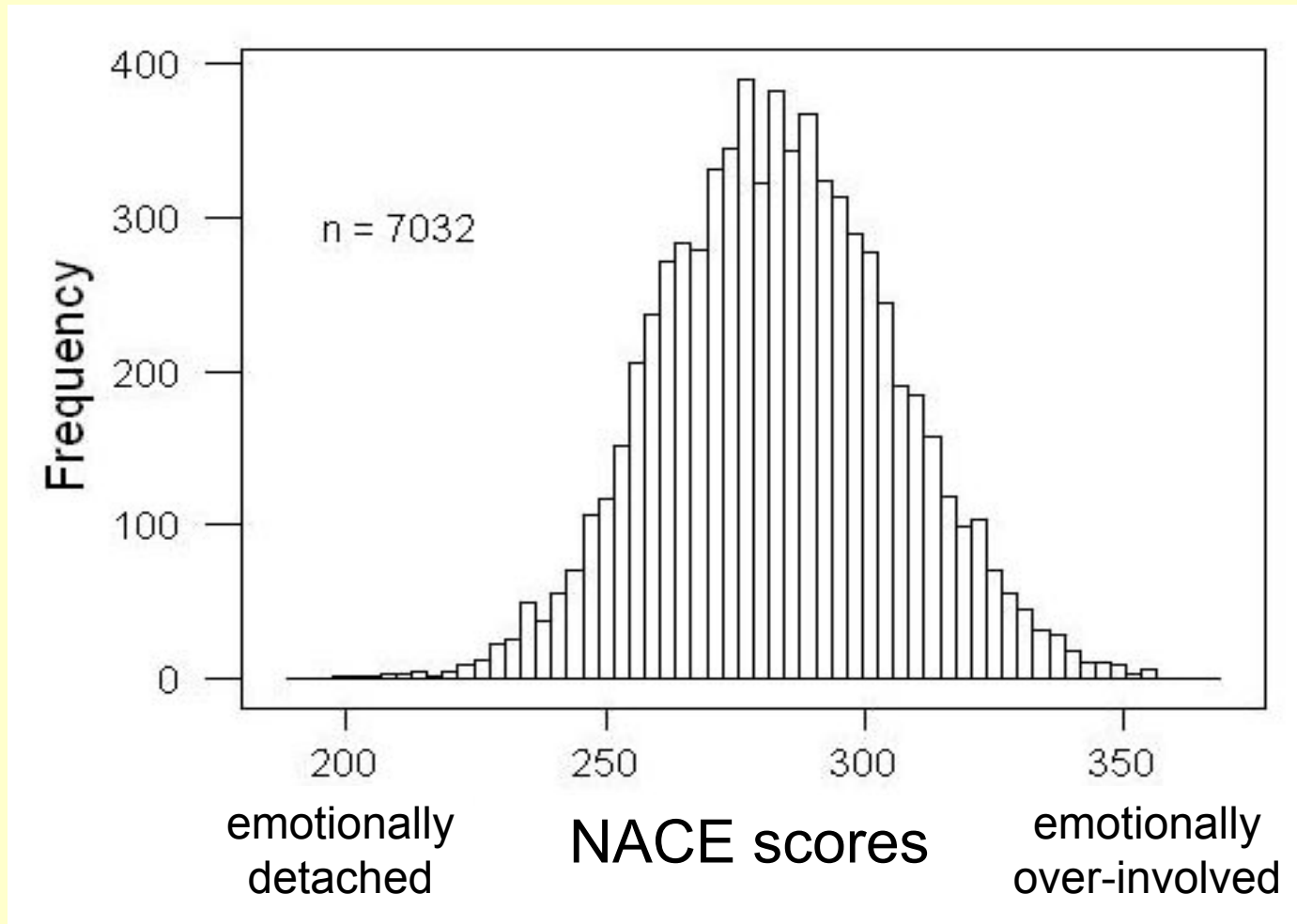
PQA

NACE Questionnaire

- 100 item personality inventory (Munro *et al.* 2005)
- approx. 30 minutes to complete
- factor analysis indicates 4 factors
- factors condense to two pairs on a continuum:

Involved with others vs Detached from others

NACE



PQA NACE Questionnaire

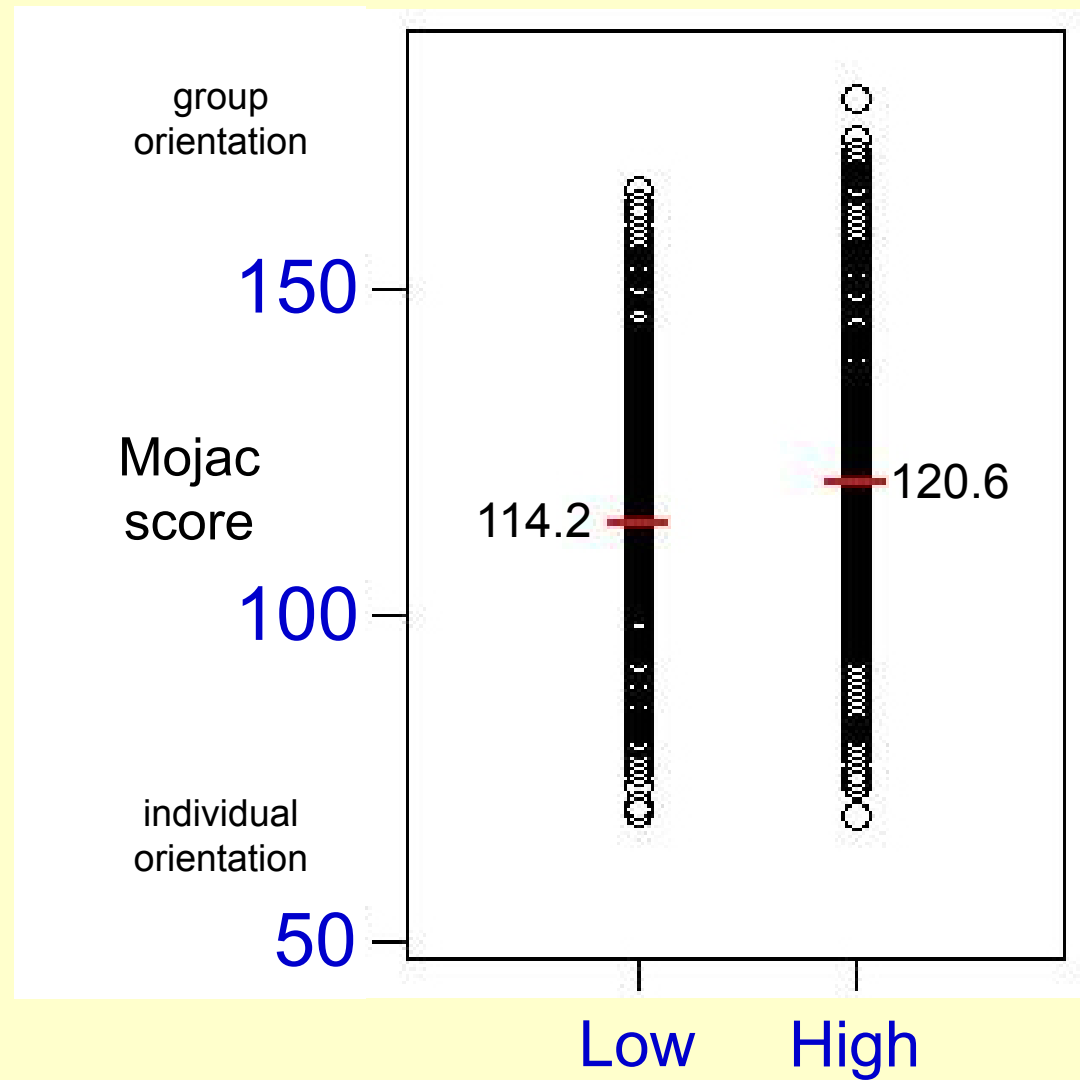
Data on 20,000+ individuals

- gender differences
 - age effects
 - effects of socio-economic status
 - effects of ethnic background and culture
 - stable cohort mean score during medical education
-

Mojac and NACE are reliable

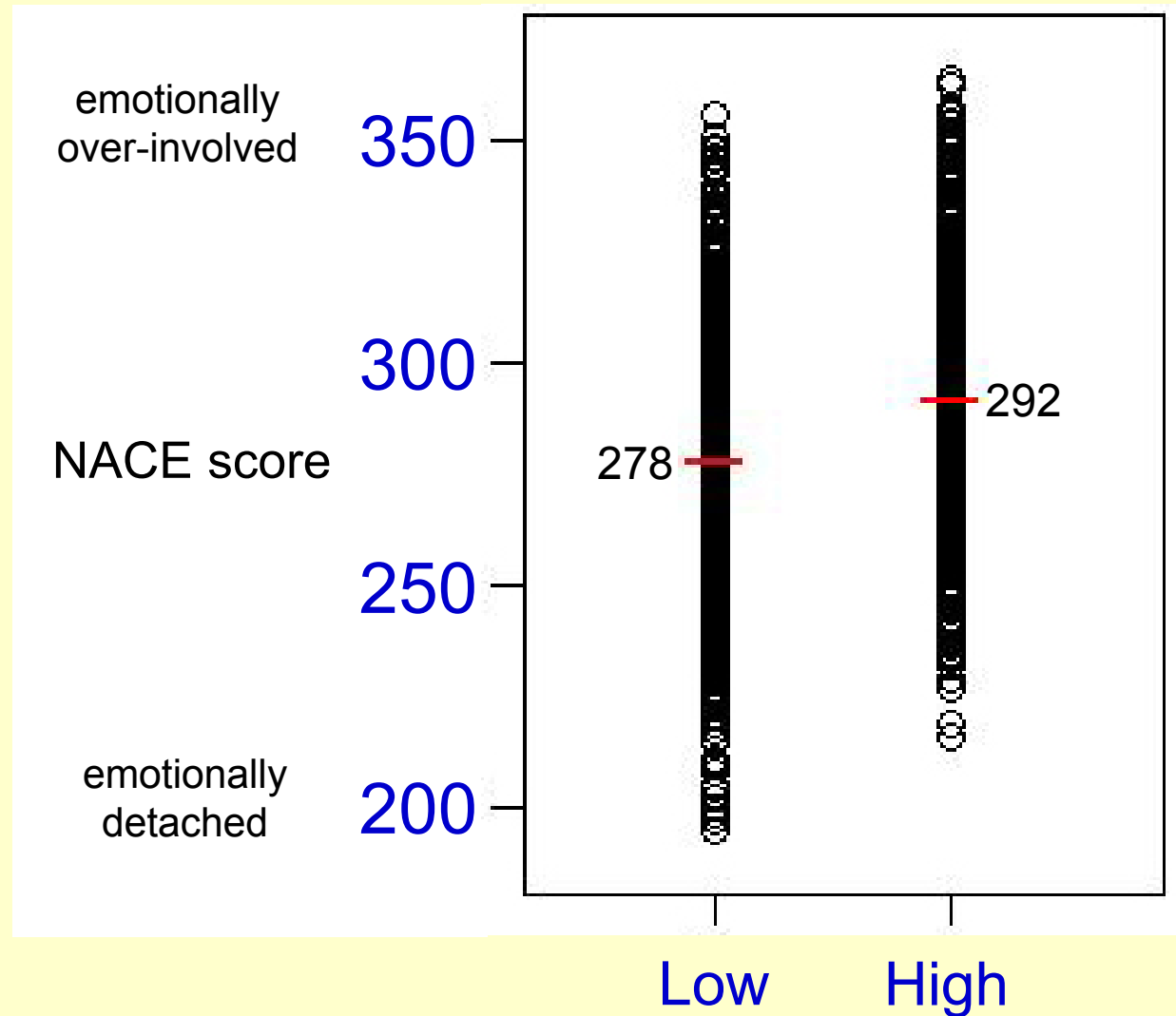
- high internal consistency
(*Cronbach alpha reliability coefficients >.9*)
 - test-retest reliability (over time) satisfactory
 - no (subjective) rater judgements used
 - discriminate well within cohorts
 - difficult to fake ‘good’ in either test
-

Mojac



N = 4412 4668

NACE



N = 4267

4570

Do the tests have construct
validity?

do they measure what
purport to measure?

Construct validity of Mojac and NACE

- Correlate with other standard measures as expected from theoretical model
 - 16PF modified (Cattell, 1998)
 - IPIP Five-Factor Test - 'Big 5' (Goldberg, 1999)
 - Right Wing Authoritarianism (Altemeyer, 1982)
 - Emotional Intelligence (Schutte *et al.* 1998)
 - Eysenck Personality Questionnaire (Eysenck, 1985)
 - Depression, Anxiety & Stress Scales (Lovibond, 1995) [modified]
 - Horney-Coolidge Type Indicator (Coolidge, 2001)
-

PQA

Personal Characteristics Inventory

- discussion with interested parties indicated need for a test of '*integrity*' and '*mental robustness*'
 - personality trait correlates
 - Conscientiousness / Self-Control
 - Emotional Stability / Resilience
 - Agreeableness / Involvement
 - these traits good predictors of general workplace outcomes
-

PQA

Personal Characteristics Inventory

PCI designed to measure

- **Self-Control** (vs Disorderliness)
- **Resilience** (vs Emotional Reactivity)
- **Lie Scale**

Questions derived from other published instruments used in PQA validity research, e.g.,

- Self-Control Scale (Tangney *et al.* 2004)
 - IPIP Five-Factor Test (Goldberg, 1999)
 - Brief Symptom Inventory (Derogatis *et al.* 1982)
 - 16PF (Cattell, 1994 edition)
 - EPQ (Eysenck, 1976)
-

PCI research

- samples of Psychology and Medicine students at three Australian universities
 - internal consistency (reliability) (n = 605)
 - Self-Control $\alpha = .85$
 - Resilience $\alpha = .89$
 - rest-retest reliability (9 months, n = 54)
 - Self-Control $r = .74$
 - Resilience $r = .74$
-

PCI construct validity

- Correlates of PCI Self-Control

IPIP Conscientiousness	.51
------------------------	-----

Tangney Self-Control	.63
----------------------	-----

16PF Rule-Consciousness	.52
-------------------------	-----

16PF Perfectionism	.60
--------------------	-----

16PF Self-Control	.74
-------------------	-----

- Correlates of PCI Resilience

IPIP Neuroticism	-.66
------------------	------

PQA research base

Extensive research studies over 12 years

- > 28,000 individuals
- England, Scotland, Australia, Israel, Sweden, Japan, Taiwan, Hong Kong, Fiji, Canada

Established

- high reliability of all subtests
- construct validity

Predictive Validity?

PQA

- not widely used for selection, **hence** has greater power to predict student outcomes than UMAT or GAMSAT, because initial data not restricted in range
 - with some exceptions still no large correlation between any of the tests and most medical school student outcomes
-

predictive validity

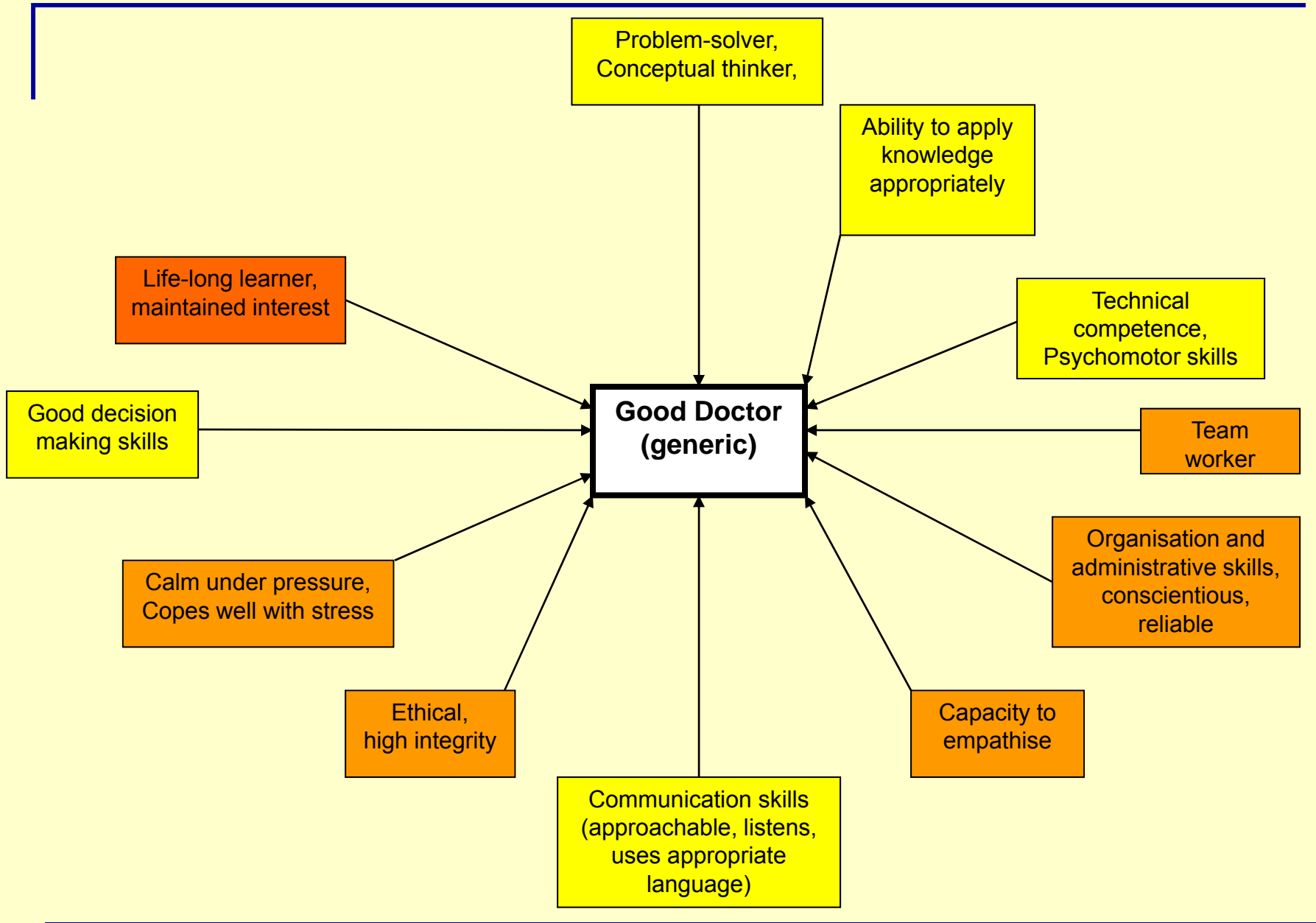
- of tests, including *PQA*
 - and interviews
 - has not been decisively demonstrated
-

What can we conclude?

- perhaps that medical schools are failing to assess in program outcomes the specific qualities and behaviours measured by selection tests and interviews
-

Implication of assessment failure

- medical schools may be graduating doctors deficient in the generally agreed important qualities tests and interviews are designed to measure
-



Some positive exceptions

PQA predicts 'clinical interview' scores

- medical students trained in clinical interview
- then video recorded
- videos scored objectively

Good interview performers - higher NACE scores

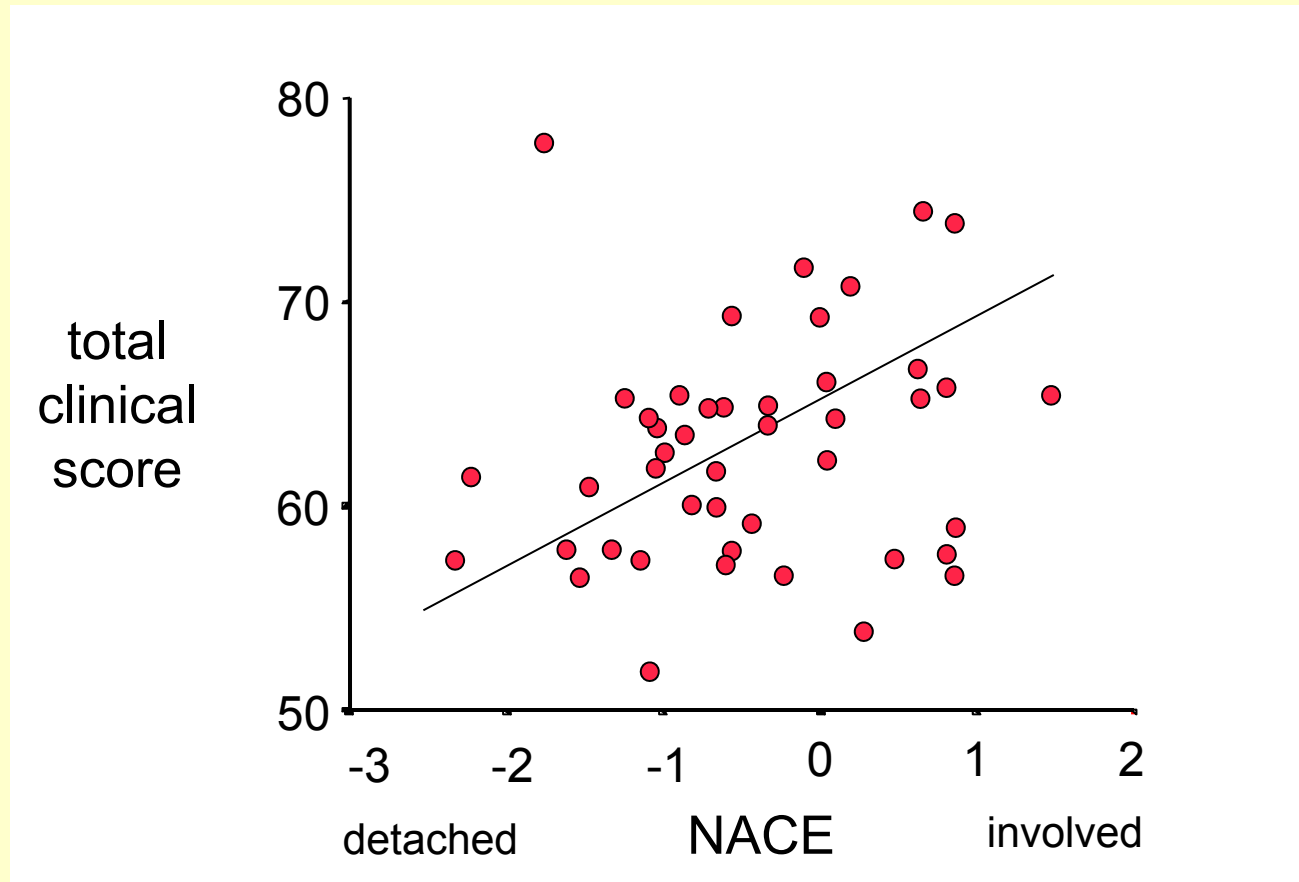
($t = 2.5$, $p = .02$)

Poor interview performers - lower NACE scores

($t = 2.7$, $p = .01$)

PQA predicts Final BMBS score

- those who are 'more involved' do better



PQA and job satisfaction in GPs

high scorers for

Neuroticism (vs Emotional Stability)

found their job

- more demanding ($r = .31, p < .05$)
- more stressful ($r = .35, p < .01$)

(N = 57 GP volunteers)

PQA Final year pharmacy students

predicted performance on communications skills tasks

- ability to handle ‘difficult customer’
correlated with *Confidence* ($r = .556$)
negatively correlated with *Aloofness* ($r = -.479$)
- clinical scenario viva
correlated with *Confidence* ($r = .434$)
- simulated patient seeking advice
correlated with *Confidence* ($r = .458$)

($n = 28$)

A UK medical school year 1 students, 2007-2008

- 143 newly enrolled medical students at a UK Medical School
 - completed PCI, Mojac and NACE
 - were appraised by tutors at twice weekly PBL sessions and during personal mentoring throughout year
 - at end of year sat for summative examinations
-

- tutor appraisal

- attends punctually
 - treats tutors with appropriate respect
 - demonstrates appropriate attitude
 - completes given tasks on time
 - integrates themselves into group
 - takes responsibility for group learning
 - contributes work for the group
 - treats peers with respect
 - listens effectively
 - willing to learn from others
 - undertakes PBL role appropriately
 - communicates appropriately with peers
 - communicates appropriately with tutors
 - manages conflict appropriately
-

-
- end of year examinations
 - Theme A: life sciences; clinical sciences
 - Theme B: clinical techniques and skills; person-centred care
 - OSCE stations: communication skills (7)
 - OSCE stations: practical skills (7)
 - Theme C: evidence-based decision making; population health & medicine; managing resources
-

significant correlations ($p < 0.05$)

SAI
ITP
IVP

non-cognitive measure	tutor rating	OSCE practical skills	OSCE communication skills	examination theme	distinguishes top 20% from bottom 20% of exam achievers
self-control	appropriate attitude + completes tasks on time + listens effectively +				
impulsive	completes tasks on time – listens effectively -				
permissiveness	listens effectively -		all -		
anti-social tendencies	manages conflict appropriately -				
moodiness					
conscientiousness					
neuroticism					
unreal thoughts			all –		
lie scale				Theme B –	
narcissism	manages conflict appropriately -				$p = 0.025$
aloofness	manages conflict appropriately -			Theme A –, Theme B –	$p = 0.001$
confidence	contributes work to group + treats peers with respect – listens effectively – manages conflict appropriately -	all +	all +	Theme A +, Theme B +	$p = 0.001$
empathy	manages conflict appropriately -			Theme C +	
ECAN				Theme B +	$p = 0.005$
social responsibility	attends punctually + treats peers with respect +			Theme B +, Theme C +	$p = 0.004$

Conclusion

- *PQA* does appear to be measuring qualities that predict actual behaviours
-