

In Search of Equity

An Update on CASPer

Prof H Reiter, MD, MEd, FRCPC, DABR - McMaster University

Disclosures:

Founder, shareholder, board member of Altus Assessments Inc - CASPer; no income

How I know my 10
month old grandson will
be a researcher...

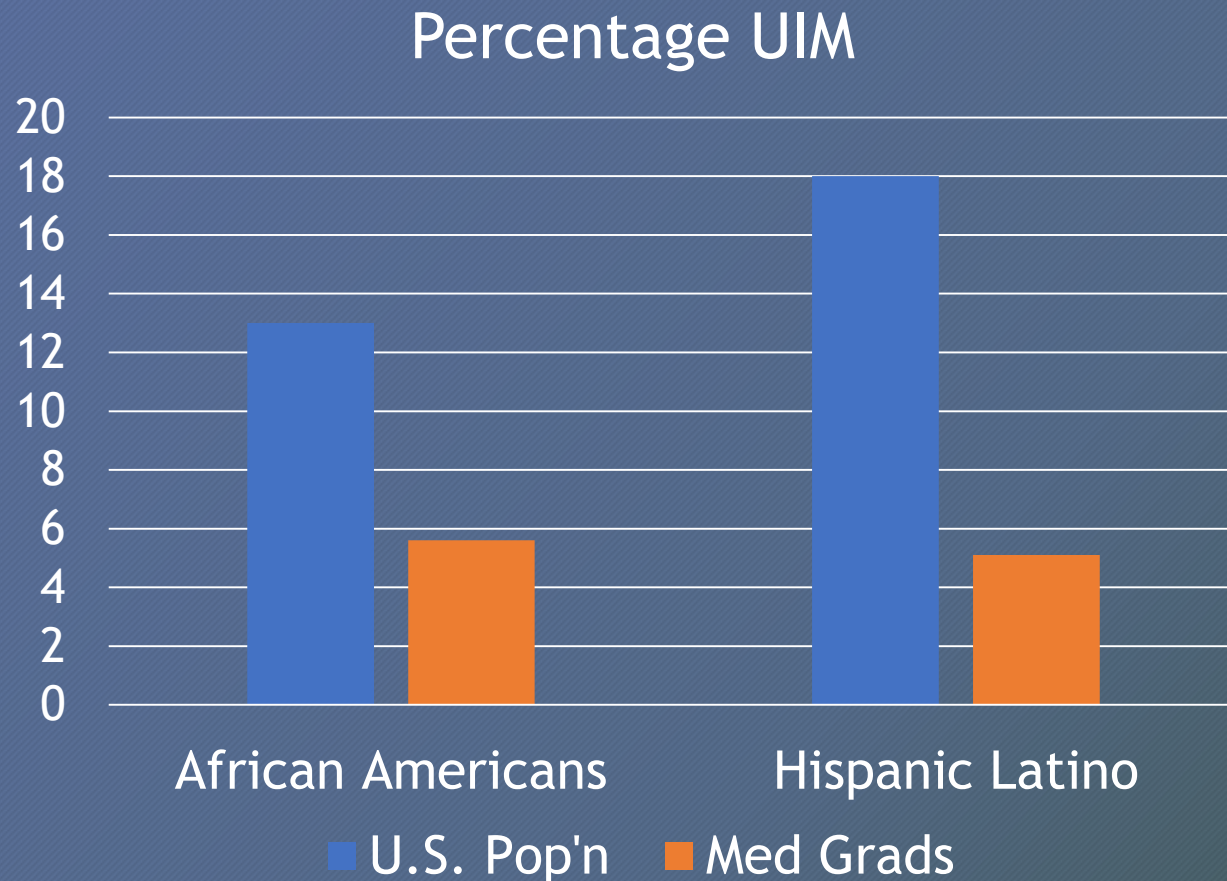
... or an appliance
repairman



Why do we care?

- Health Disparities - race, ethnicity, SES
- Patients prefer “their own”
- UIM grads gravitate more to underserved
- Diversity enhances cultural competency
- Betancourt 2003, Fiscella 2016
- Cabral 2011
- Xu 1997
- Guiton 2007

How have we done?



- Garces 2015, Smith 2015
- AAMC Databooks 2017 grads
- Parental education, income
 - Cooper 2003
- Higher SES > 50% med students
 - Grbic 2015

How have we done?

- AAMC News Press Release, Tuesday, December 04, 2018

- **Women Were Majority of U.S. Medical School Applicants in 2018**

- **Data show increasing diversity among applicants and enrollees**

African American enrollees rose by 4.6%

Native American enrollees rose by 6.3%

- African American enrollees

2017: 5.63%

2018: 5.81%

- Native American enrollees

2017: 2.38%

2018: 2.59%

How have we done?

The Supreme Court of the United States, *Grutter v Bollinger et al*, 2003, referring to the 25 years since the previous SCOTUS ruling on the subject in 1978 *Regents of the University of California v Bakke*, 1978:

“The Court expects that 25 years from now, the use of racial preferences will no longer be necessary to further the interest approved today”

Cognitive competencies assessment tools that show no subgroup differences:

As GPA, MCAT emphasis rise, UIM's decline - Heller 2014

Gender differences - MCAT - Ramsbottom-Lucier 1995

Differential educational opportunities

Personal competencies assessment tools that show no subgroup differences:

Adverse impact on minorities lower - Ballejos 2015

Multiple Mini-Interview (Terregino C, Acad Med, 2015)

Situational Judgment Test (Lievens, J Appl Psychol, 2018)

GENESIS 24

15 And it came to pass, before he had done speaking, that, behold, **Rebekah came out, who was born to Bethuel**, son of Milcah, the wife of Nahor, Abraham's brother, with her pitcher upon her shoulder.

16 And the damsel *was* very fair to look upon, a virgin, neither had any man known her: and she went down to the well, and filled her pitcher, and came up.

17 And the servant ran to meet her, and said, Let me, I pray thee, drink a little water of thy pitcher.

18 **And she said, Drink, my lord: and she hasted**, and let down her pitcher upon her hand, and gave him drink.

19 And when she had done giving him drink, she said, I will draw *water* for thy camels also, until they have done drinking.

20 And she hasted, and emptied her pitcher into the trough, and ran again unto the well to draw *water*, and drew for all his camels.

If so, why do most personal competencies assessment tools show subgroup differences?

READING COMPREHENSION

CONSTRUCT CONTAMINATION

Imagine you are a teacher at a community college...

Christian Pers Psych 2010, Chan J Appl Psych 1997, Weekley Pers Psych 1997

RESPONSE DISTORTION

You will have 5 minutes only to provide responses to all 3 questions...

Arthur, Indust & Org Psychology 2009

TEST-TAKING SKILL

Select the one best response from the following options...

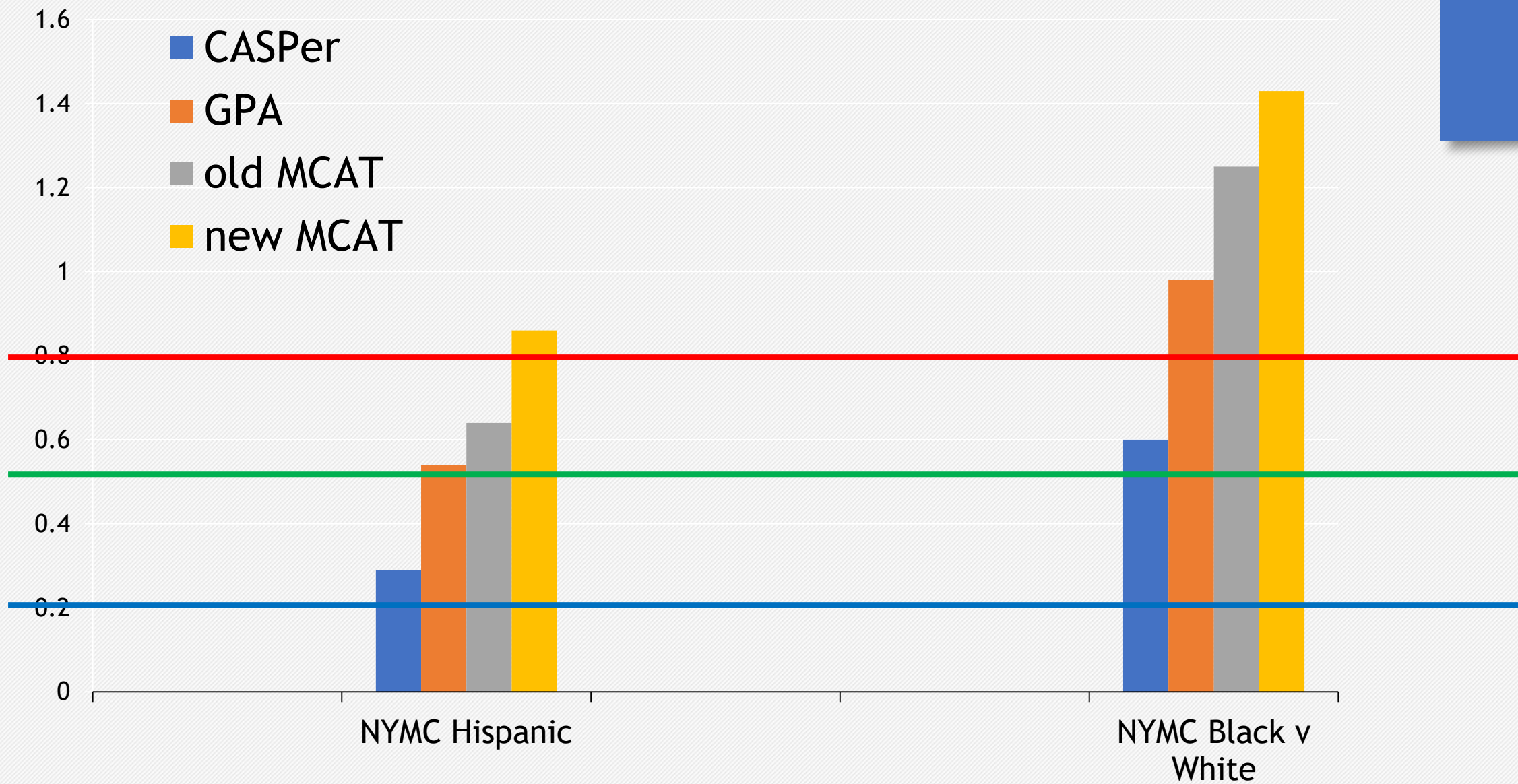
Arthur, Pers Psychol 2002

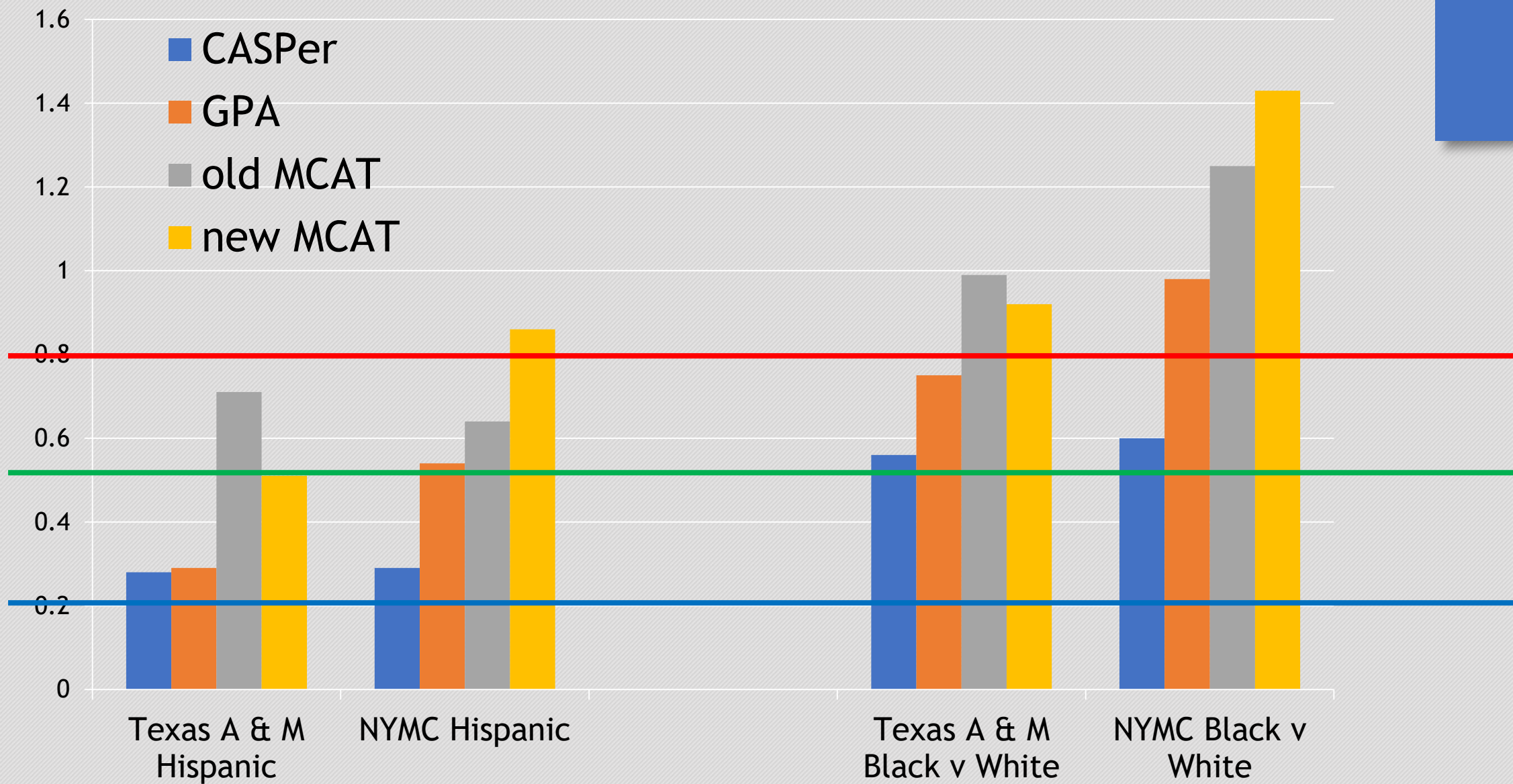
WRITING SKILL

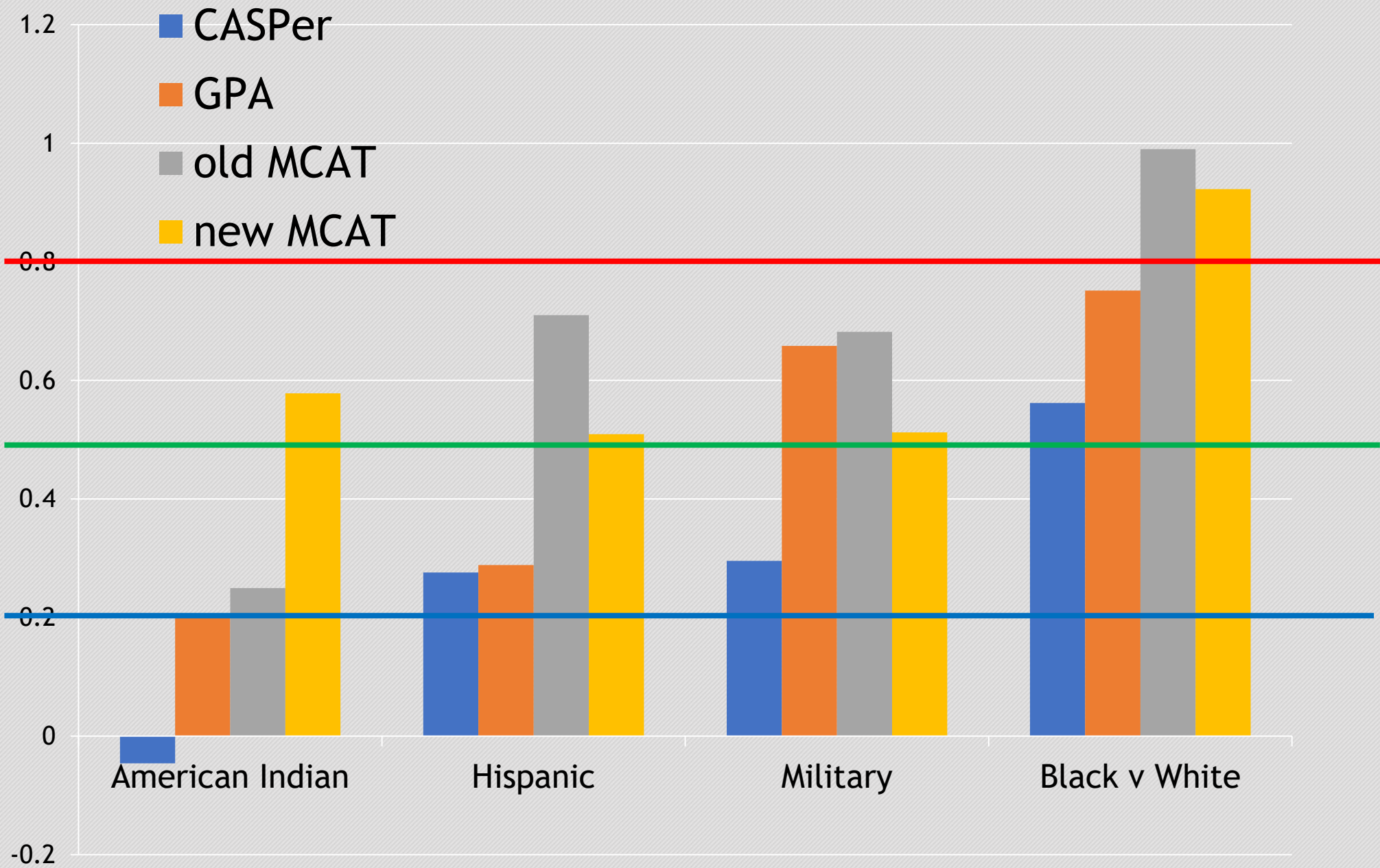
You will have 5 minutes only to type responses to all 3 questions...

Lievens, J Appl Psychol, 2018

COACHING EFFECT, PRACTICE EFFECT Koenig Acad Med, 1997, Dore AMEE, 2018







Conclusions so far...

1. MCAT-related subgroup differences are moderate to high across the board
2. GPA-related subgroup differences vary from small to moderate to high
3. CASPer-related subgroup differences vary from absent to small to moderate
4. MCAT & GPA-related subgroup differences are smaller for Texas than in NY
5. CASPer-related subgroup differences are the same for Texas and for NY

In Search of Equity: The Ideal SJT?

CONTAMINATION

- Reading Comprehension
- Response Distortion
- Test-Taking Skill
- Writing Skill
- Coaching Effect
- Practice Effect

REMOVAL

- Video stems
- Speeded testing
- Constructed response
- Video responses
- Speeded, constructed response
- Speeded, constructed response

In Search of Equity: Changes to CASPer

JUNE 2015 - SPRING 2019

- 12 SJT sections
- 8 video stems, typed response
- 4 text stems, typed response
- 5 minute response time

JUNE 2019 -

- 14 SJT sections
- 8 video stems, typed response
- 4 text stems, typed response
- 2 video stems, video response
- Shorter video response time

*shorter video response time than for typed - Dore Acad Med 2009, NITE pers comm 2017

In Search of Equity: Data Analysis & Beyond

GOALS

- Confirm feasibility
- Test reliability
- Convergent validity
- Discriminant validity
- **Subgroup differences**
- Acceptability
- Ideal response time

TIMELINE

- Changes between cycles
- June 2019 - early 2020
- Spring 2020 data analysis
- Result-generated test Jun 2020

CASPer test raters; Demographics and QA

Equity & The rating system

THE CASPer[®] ASSESSMENT

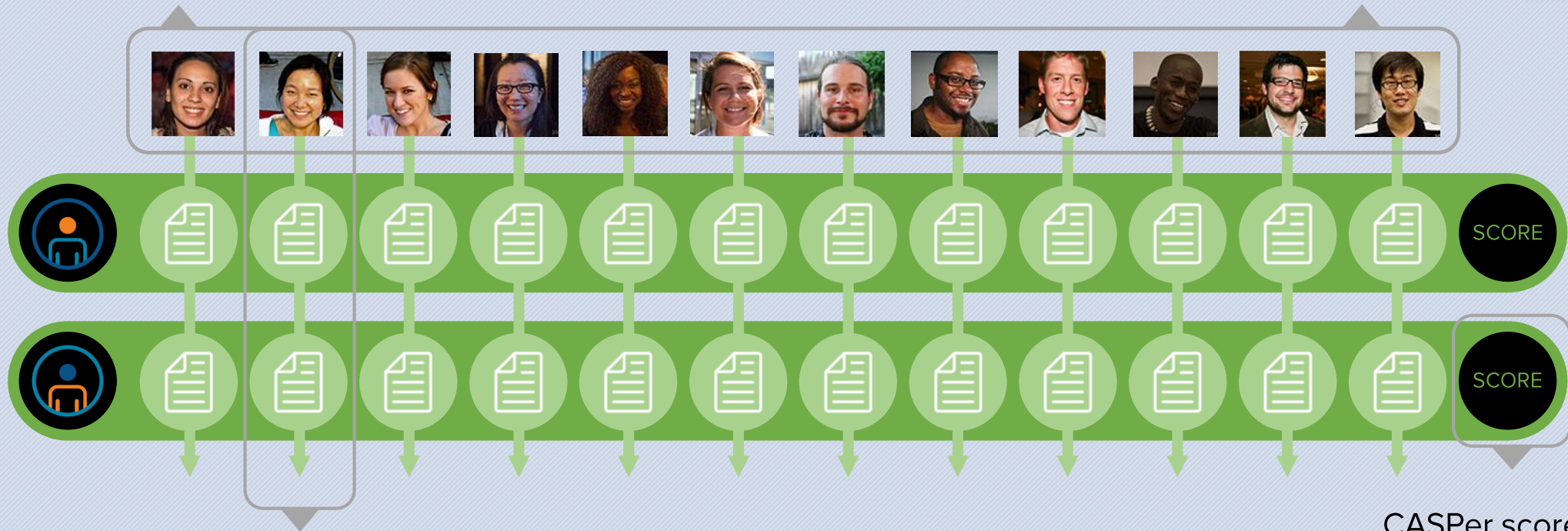
An online, video based
situational judgement test (SJT)

- 12 sections
- 3 open-ended questions per section
- 5 minutes per section



HOW IS CASPer™ RATED?

Raters are blinded



Each rater focuses on 1 scenario, gets anonymized responses, likert 1-9 RPM

CASPer score is an average of 12 unique impressions

RATER QA

- Monitored daily/real-time with structured QA system
- Responses from single section oversampled
- Rater QA metrics monitored include:
 - low inter-rater reliability (IRR)
 - overly low or overly high mean ratings (too lenient or too strict)
 - low standard deviation (is not using the full range of the scale to differentiate applicants)
 - excessively fast reading speed
 - misuse of flagged responses

IRR Responses 2017-07-09T14

Responses for USMEDICINE test on 2017-07-09T14.

The test ID: DMYZuclZKx8u3RHGR

The inter-rater reliability intraclass correlation (ICC) for this test was: **0.95**

The scenario oversampled was: **Public Shaming (Set A)** which was a video scenario.

You are a friend.

What is your opinion on this situation?

If the original social media post turns out to be inaccurate, is posting a retraction a reasonable remedy? Why, or why not?

Social media is a powerful tool. Should an individual's privacy be considered even when using social media to make a positive impact?

	Rater Correlation	Mean	Standard Deviation
MrEqDKcRmGmrptxdQ	0.45	5.9	1.21
qSEe2kEEPziz7C5Lw	0.64	5.6	1.67
wcr3m8pbpHGwaayXp	0.68	5.7	1.26
BAfnsq5EnJeqZjtX6	0.77	5.45	1.28
336i7FHh9fJBzECmt	0.82	6.25	1.33
RAom2nirf3WiqmNXQ	0.72	5.05	0.76
eBLwNbMA3Pw44BJ6H	0.48	6.05	0.51
tXCHd5aa9qfvTDT2C	0.71	5.8	1.82
yPfHWf7teskWmNbAX	0.45	5.15	1.84
suZLCnHNuj9ri9Hjg	0.73	5.1	0.85
xPLsp7Mn2PmTS3bux	0.74	5.6	1.6
A2iw8Cb6q2DS2ZRcB	0.72	3.9	1.55
iZQrH8wJFHYqAdy3K	0.72	5.85	1.39
SoZGXSyupCiAYXNjE	0.61	6.6	1.76
nm28adLt5RgBnxopX	0.52	5.2	1.2
g3EPpyrtzwtu9ZrD	0.4	4.7	1.53
ygSWgJFQRmdCSCua8	0.66	5.05	0.94

Sample printout of rater descriptions (raters are anonymized for this document)

*Roughly 60 raters were removed from our platform last year for a variety of reasons

What's required of raters to get started?

- Raters submit a resume/CV as a part of the application process. We assess for:
 - Relevant experience
 - Healthcare knowledge
 - General work experience
 - Lack of connection to medical school applicants
- A shortened version of the CASPer test made specifically for raters; Serving to:
 - Help familiarize raters with the content/format.
 - keep an eye out for very strong opinions or anything “red flag” worthy.
- An online video training module
- A set of practice responses + Pre-benchmarked responses to rate.

Where do we recruit from?

- Conferences & Networking with Universities)
- Job Boards (indeed and others)
- Schools (not exclusively healthcare)
- Special Interest groups
- Professional associations
- Email campaigns, flyers, Posting to specific websites.

Rater Pool Goals

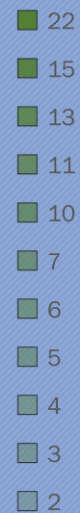
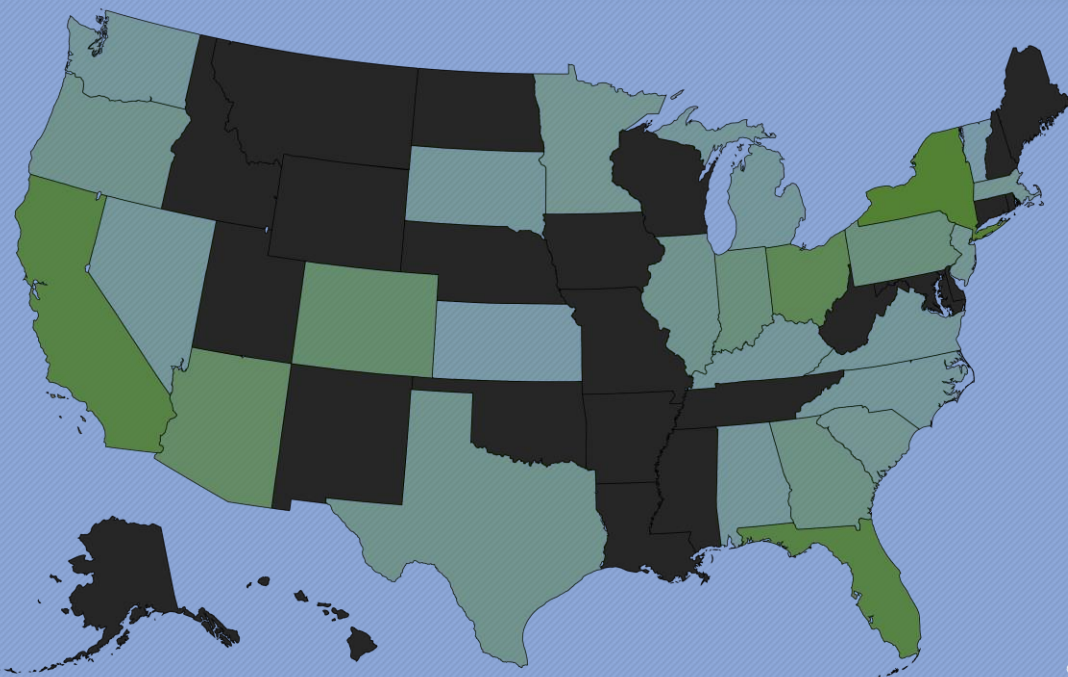


1. The rating pool is localized (nationally) with the tests that will be rated. (American raters for American healthcare programs).
2. The rating pool is as diverse as the population that these professionals will one day serve (based on national demographics).
3. The raters have an interest the future of healthcare professions or education.

Why is diversity in the rating pool important?

- Different perspectives based on background, abilities, and experience.
- Wisdom of the Crowd VS. Experts?
- Do we want exclusively Deans of Medical Schools?

Where do our raters live?

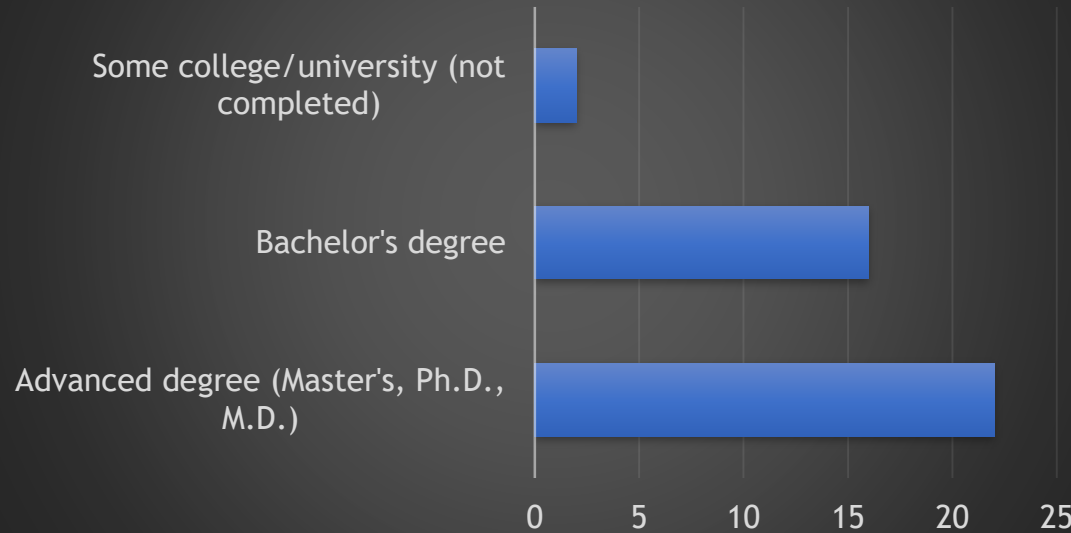


There are approximately 200 active raters in the United States, spread across 28 different States.

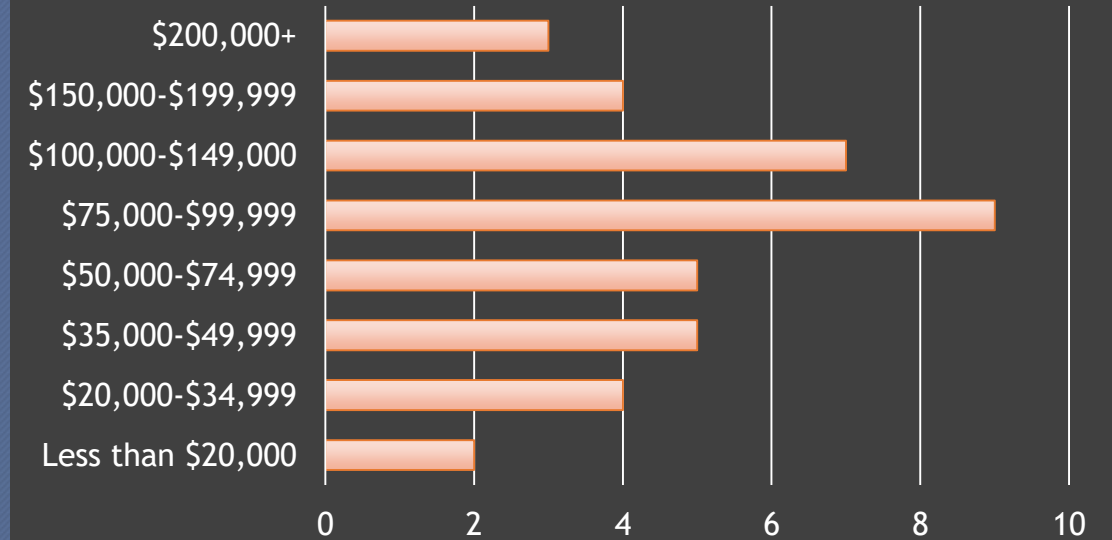
The five states with the most raters are: New York, California, Florida, Ohio, and Arizona

Education Level & Income of the Raters

Education Level



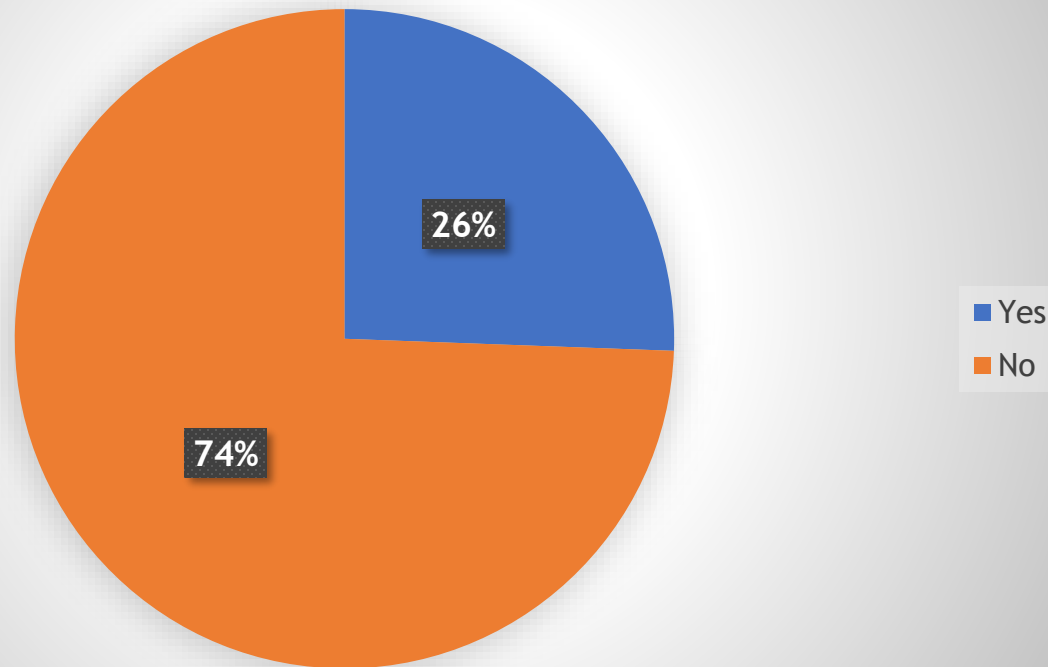
House Hold Income



- Vast Majority of raters possess a post-secondary degree (87% have a degree)
- Most raters have a “Middle class” income - Though there’s a wide range.

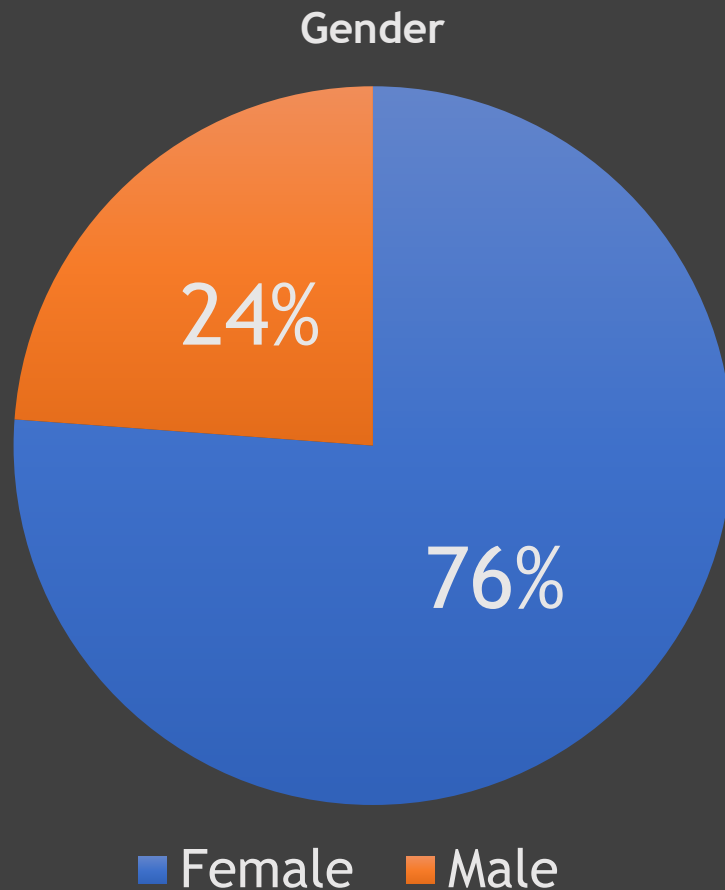
Rater Occupation (Healthcare vs. Other)

Work in Healthcare?



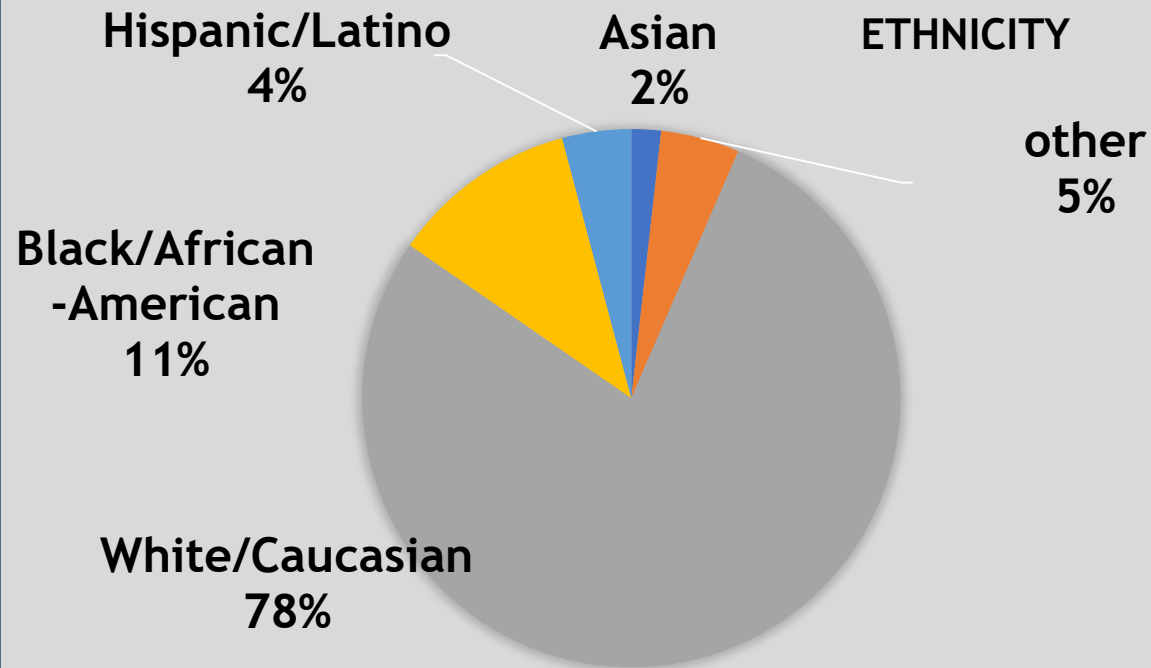
- Over $\frac{1}{4}$ of the rating pool works in a healthcare related field.
- Lots of diversity in other professions (ranging from flight attendant to retired).

Gender & Age of Raters



- Males underrepresented
- Average age of raters is 37, with a wide range of ages.

Ethnicity of Raters



- Hispanic/Latino very underrepresented - Could be a product of where previous recruiting efforts have taken place.
- Asian American significantly underrepresented

What's next for rating?

Rater QA

- Implementing “Rater Report Cards”
- Developing more robust methods for intervening when raters aren't performing well. This includes more automated messages.
- “Re-training process” and upkeep for current raters.

Rater Recruitment

- We're doing OK, but there's room for improvement!
- More targeted outreach campaigns, including more associations and reaching out to communities within specific geographies.
- Seeking assistance/guidance from programs



THE ADMISSIONS SUMMIT

**Navigating Healthcare
Admissions Challenges
Together**

June 10-11, 2019 | Toronto, ON | Globe and Mail Centre | www.theadmissionssummit.com

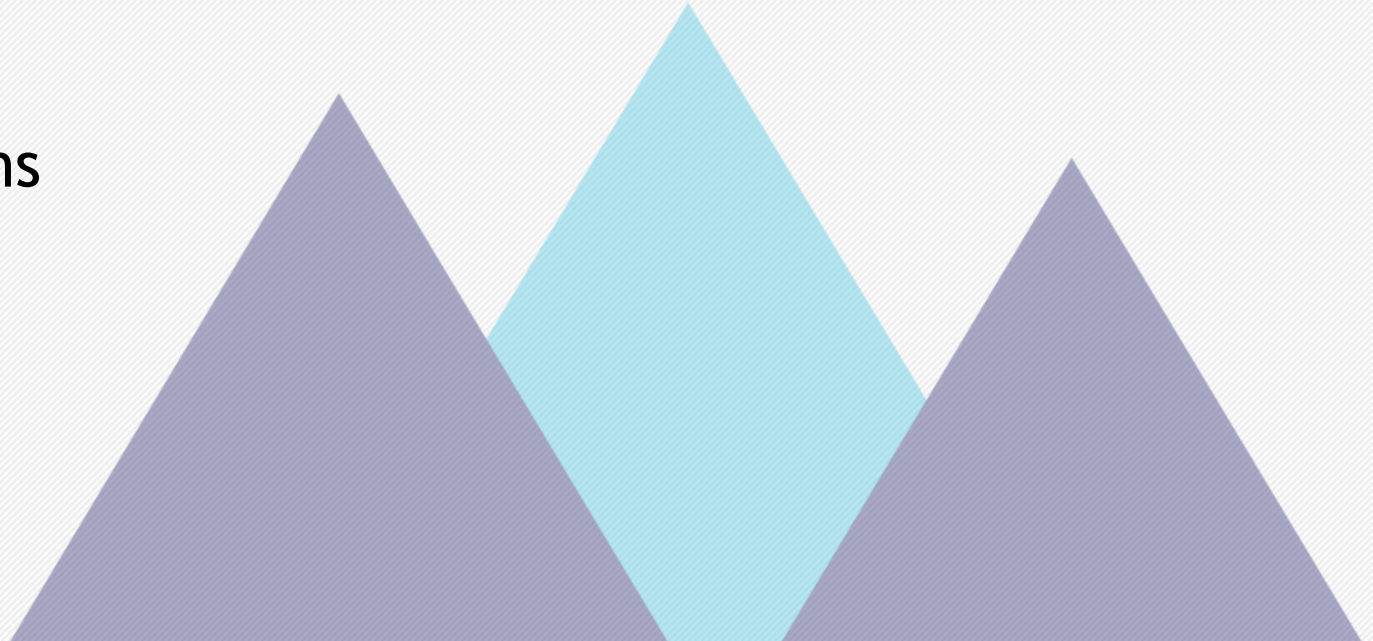
Who should attend

Open to all healthcare programs:

- Medicine, Osteopathic Medicine, Physician Assistant, Nursing, Pharmacy, Veterinary, Optometry, Dentistry, Physiotherapy and Occupational Therapy

Designed for admissions professionals:

- Deans of Admissions
- Associate Deans of Admissions
- Directors of Admissions
- Admissions Staff
- Program Directors



What to expect



Navigating Healthcare
Admissions Challenges
Together



Build a Network of Healthcare
Admissions Professionals



Benefit from the
Mentor/Mentee Program



Discuss Challenges you face
and work through them with
Peers and Experts



A Unique Conference
Experience



Explore the exciting City of
Toronto

Speakers and Themes

The Holistic Story

Data-Driven
Selection

Diversity &
inclusion

The Future of
Admissions



Allison Massari
The Holistic Story



Dr. Nathan Kuncel
Data-Driven Selection



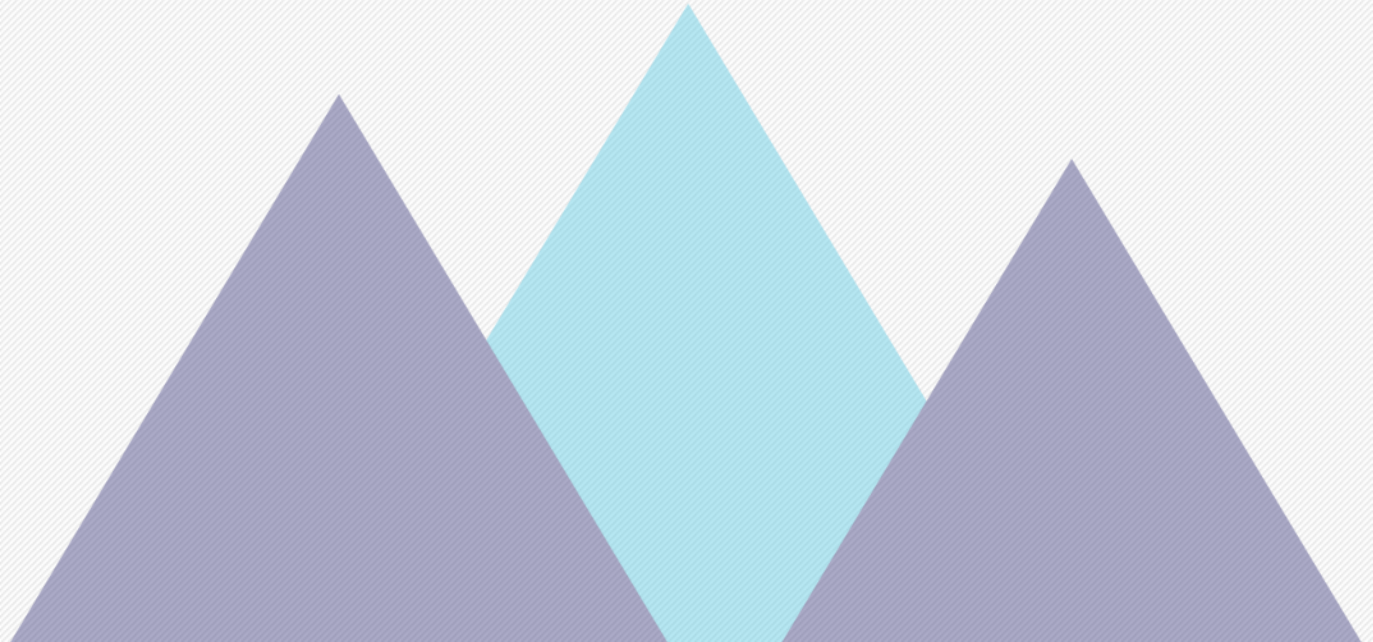
Dr. Jessica Remedios
Diversity & Inclusion



Scott E. Page
Diversity & Inclusion

Register now!

- Visit www.TheAdmissionsSummit.com
- Early Bird ticket pricing ends March 31st
- Limited tickets available
- Call for Submissions - we want to hear from you!



In Search of Equity: Can We Do Better?

H Reiter

reiterh@mcmaster.ca

J Morrow

jmorrow@altusassessments.com

T Qadir

tqadir@altusassessments.com