

Measurements of performance gaps are sensitive to the level of test stakes: Evidence from PISA and a Field Experiment

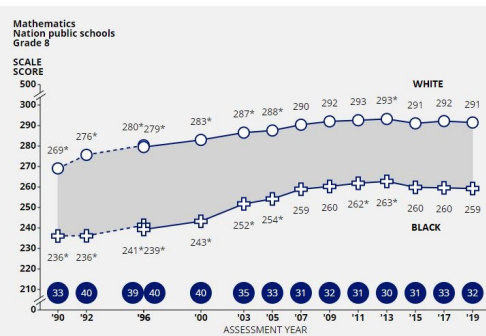
Yuval Ofek-Shanny

Friedrich-Alexander-Universität Erlangen Nuremberg

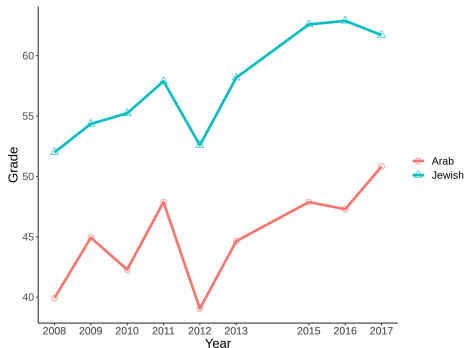
August 2022

Minority Students Underperform on Assessment Tests

United States

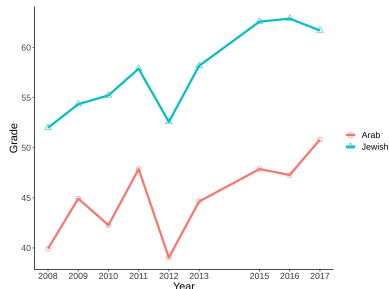


Israel



Why Do Minorities Underperform?

- Assessment tests are low-stakes
- Performance determined by proficiency and intrinsic motivation (and other factors)
- Heterogeneity in intrinsic motivation can distort performance comparisons



This Paper

- **Are minority students less engaged on low-stakes tests?**
 - Use engagement measures on PISA 2015
- **Does lower minority engagement lead to overestimated performance gaps?**
 - Experimentally compare performance in a low and high-stakes test of 8th grade Arab and Jewish students, Israel
- **Preview of the results -**
 - Minority students in the US and Israel are less engaged
 - 60% of the Jewish-Arab performance gap in Israel can be attributed to lower engagement

Contribution

- **Evaluation of minority performance** - show that minority students are less engaged, leading to overestimation of performance gaps.
Brey et al. (NCES, 2019), Schleicher, (OECD, 2019), RAMA (2017)
- **Field experiments on assessment tests** - combine within-subject design, natural-incentives setting, population of interest - age and group.
Gneezy et al. (2019), Schlosser et al. (2019)
- **Motivation filtering** - explain why current methods for do not capture the full effect of lower engagement.
Akyol et. al (2021), Zamarro et. al. (2019), Soland (2018)

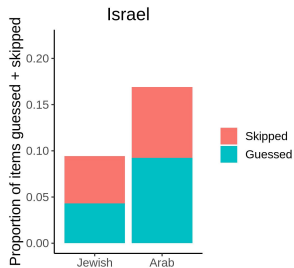
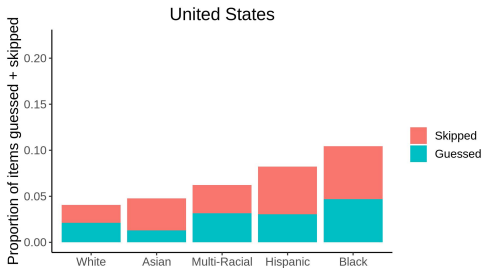
Programme for International Student Assessment

- 500K participants from 72 countries
 - 15 years old, no feedback (low stakes)
 - Ethnicity data, response time data
- Do we see evidence to lower engagement of minority students?
- Two measures for engagement -
endurance, probability of skipping/guessing

Skipping/Guessing - Methodology

- Time limit is usually not binding
- Choice not to answer or rapid guess → lower motivation
- Use response time measures to identify rapid guessing (Wise & Kong, 2005).
- Measure - proportion of items skipped/guessed per student.

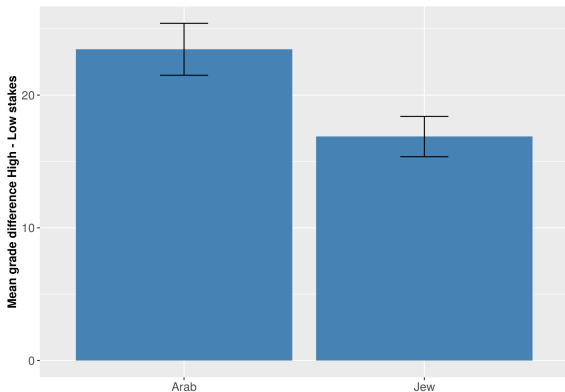
Minority students skip and guess twice as much



Experimental Design

- 600 8th grade students from 7 schools
 - Three Arab
 - Four Jewish
- Each student writes 2 similar tests in math:
 - First – Low-Stakes: no grade, no feedback
 - Second – High Stakes: with grade, about 30% of final year grade, a week later

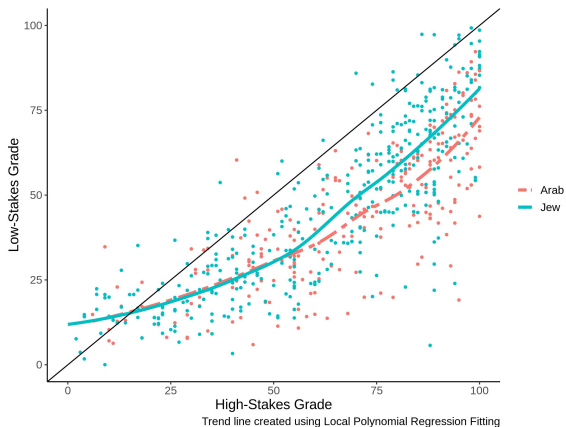
HS - LS Difference Larger for the Arab Minority



0.95 confidence level error bars

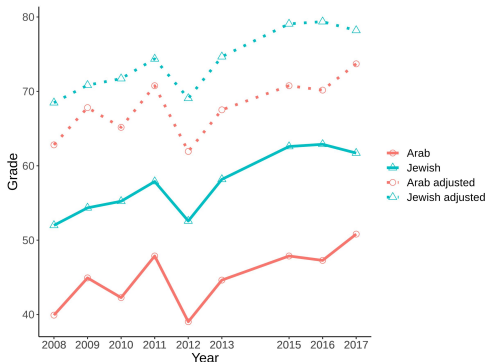
► Raw Grades

The Difference - Above-Median Students



Performance Gaps decrease by 60% When Scaled

- Performance of minority students is underestimated
- Assuming experimental results are representative
- Scaling Israeli National Assessments Tests (GEMS):



Conclusion

- Minorities are less engaged on assessment tests
- Evidence from Israel - 60% of the performance gap due to differential engagement in test, not difference in proficiency
- Choose the proper tool to measure the variable of interest to guide a well designed policy and research

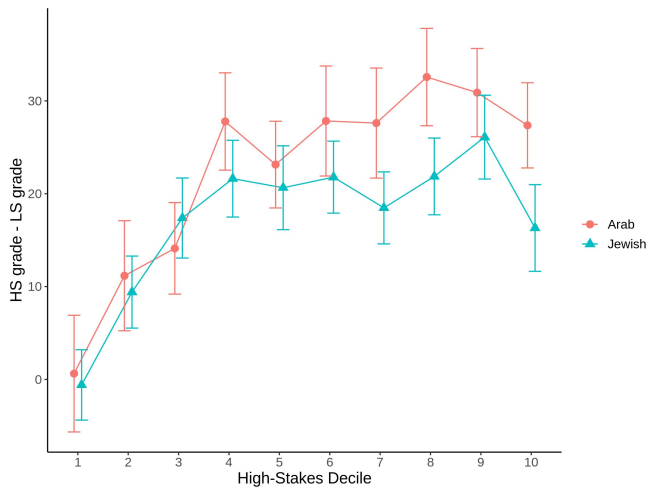
Thank You!

yuval.ofek-shanny@fau.de



Friedrich-Alexander-Universität
Erlangen-Nürnberg

The Difference - Above-Median Students



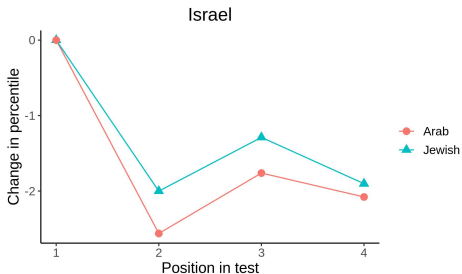
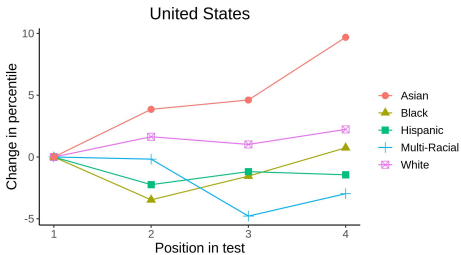
Endurance - Methodology

- Performance declines during the test
- Magnitude of performance decline is a proxy for motivation
- Heterogeneous Endurance → Heterogeneous Motivation
- Measure - percentile change relative to first part

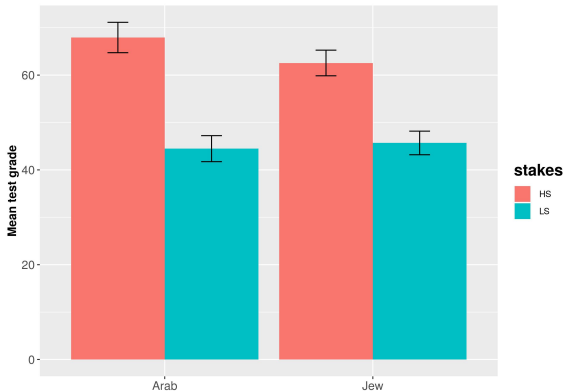
A: (80, 78, 75, 75) → (0, -2, -5, -5)

B: (50, 52, 55, 58) → (0, 2, 5, 8)

Evidence from PISA – Endurance



Grades in High and Low-Stakes Tests by Ethnicity



0.95 confidence level error bars

← Return