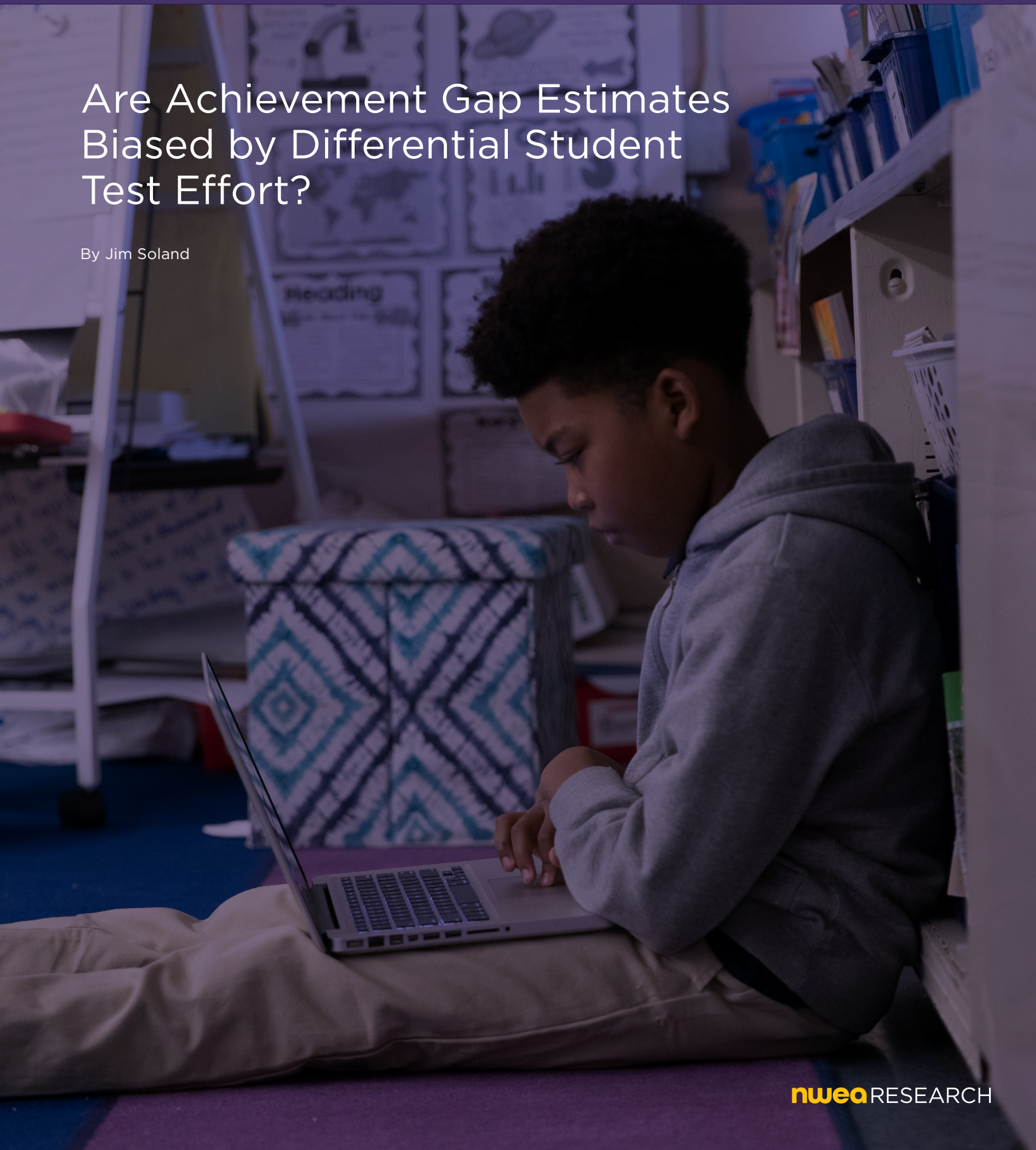


# Are Achievement Gap Estimates Biased by Differential Student Test Effort?

By Jim Soland



## KEY FINDINGS

- Test effort differs substantially by subgroup. Male students rapidly-guess nearly twice as often as female students in later grades, and African American students rapidly-guess more often than white students.
- However, these differences in rapid guessing generally do not change most interpretations of achievement gaps in a substantive way.
- Some important conclusions about gaps may change when models account for test effort.
  - The male-female gender gap in mathematics increased by around 200 percent in upper grades when test effort was considered. This shift was large enough to reverse the direction of this gap, calling into question whether apparent recent progress in narrowing the gap in math may actually reflect differential effort rather than gains by female students.
  - The gap in reading achievement between African American students and white students decreased substantially, by more than a tenth of a standard deviation in later grades, when test effort was considered.
  - Because rates of rapid guessing increased at varying rates for different student subgroups as they progressed through school, changes in gaps across grades appears sensitive to test effort.

Achievement gaps are one of our nation's most important policy metrics. For example, we study racial and ethnic achievement gaps to measure progress in reducing longstanding educational inequities rooted in segregation and economic disparity. The male-female gap in mathematics is also considered a factor in the lower numbers of women who pursue quantitative disciplines, and, because jobs requiring these skills are often more lucrative, is a contributor to earnings inequality. Educators and policymakers use gaps between races and genders to measure the effectiveness and fairness of the US education system, and progress in those areas over the course of decades. But, what are these gaps really measuring?

Achievement gap studies often use results from assessments to measure and compare average

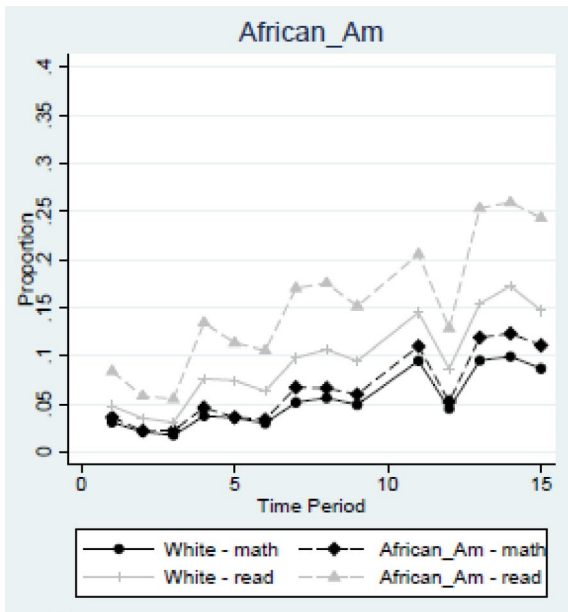
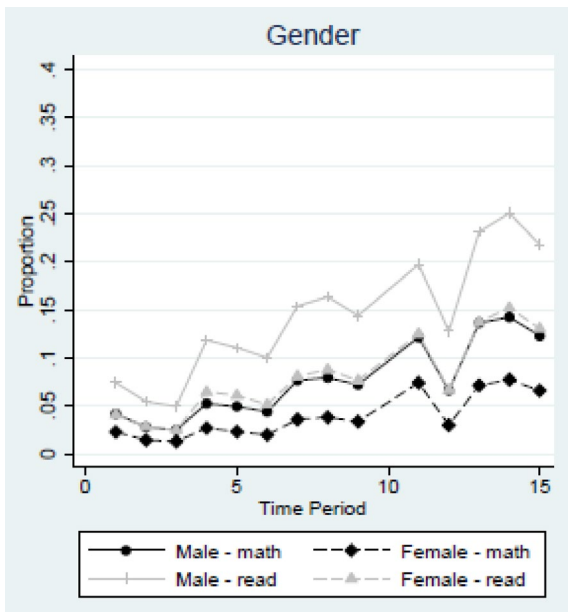
**Response-Time Effort (RTE): a measure of test taking engagement reflecting how often a student answered test questions effortfully rather than rapidly-guessing over the course of a test. Scores from tests with RTE values of .90 or below, in which students rapidly-guessed on 10 percent or more of questions, may be less reliable and biased downward**

scores for different groups of students. These studies assume that differences in observed test scores across subgroups measure differences in content mastery. However, students often disengage during tests, and rapidly-guess, answering questions so quickly that they could not have understood the content. Substantial research shows that low test effort is common and can bias observed test scores downwards, sometimes substantially. What research does not demonstrate is whether test effort differs by student subgroup and thus may bias estimates of achievement gaps.

Using data from MAP® Growth™ assessments from students in grades 5 through 9 from seven states, this study compared response time effort across student gender and racial subgroups, then estimated achievement gaps in ways that did, and did not, account for rapid guessing, to address two questions:

- Are there differences between various student subgroups in terms of test-taking effort?
- Are achievement gap estimates sensitive to test-taking effort?

Test effort differs substantially by student subgroup. Consistent with prior research, this study found that disengaged test taking was common, especially in reading, and became more frequent in higher grades. The frequency of rapid guessing behavior varied substantially by student subgroup.



Plots of the proportion of students rapidly-guessing on 10% or more of items (Y-axis) by subject and student subgroup show that test effort differs by gender and race. Each time period (X-axis) represents a fall, winter, or spring test administration, from time 1 (fall of grade 5) to time 15 (spring of grade 9). See paper for data for other student subgroups.

Test effort tended to be lower for most students of color relative to white students and for male students relative to female students, with rapid-guessing rates differing by more than .3 standard deviations in later grades. For example, by spring of grade 9, 25 percent of African American students, and about 20 percent of Latinx students, showed low test effort, compared to 15 percent of white students. Test effort differences in mathematics tended to be smaller than in reading, but were quite high for males versus females: roughly twice as many male students disengaged from math tests in high school compared to female students.

Because lower effort tends to bias scores downward, observed test scores likely understate achievement for students of color relative to white students, and for males relative to females.

### Rapid guessing did not change most interpretations of achievement gaps in a meaningful way

While rapid guessing rates differed across student subgroups, in most cases, interpretation of gap estimates did not change drastically when rapid guessing was taken into account by comparing students who rapidly-guessed on the same proportion of items across the test. In mathematics, there were no statistically significant changes in achievement gaps for any racial subgroups based on the differences in test effort. For male-female gaps, the math achievement gap changed by .05 standard deviation units, on average, when rapid guessing was considered, with a minimum change of .02 standard deviations in fifth grade and .09 standard deviations in ninth grade.

Reading achievement gap estimates for racial minorities were more sensitive to test effort, but the magnitude of the shifts also was generally modest, especially in elementary school. For Latinx students, estimates differed by .06 standard deviation units, on average, when rapid guessing was considered, with a range of .03 standard deviation units in fifth grade to .08 standard deviation units in ninth grade. For African American students, estimates differed by a mean of .09 standard deviation units with a range of -.05 standard deviation units in fifth grade to -.14 standard deviation units in ninth grade. Meanwhile, male-female gaps tended to shrink when response time effort was considered, with a mean change of .09 standard deviation units.

## In some cases, achievement gap estimates changed substantially when test effort was taken into account

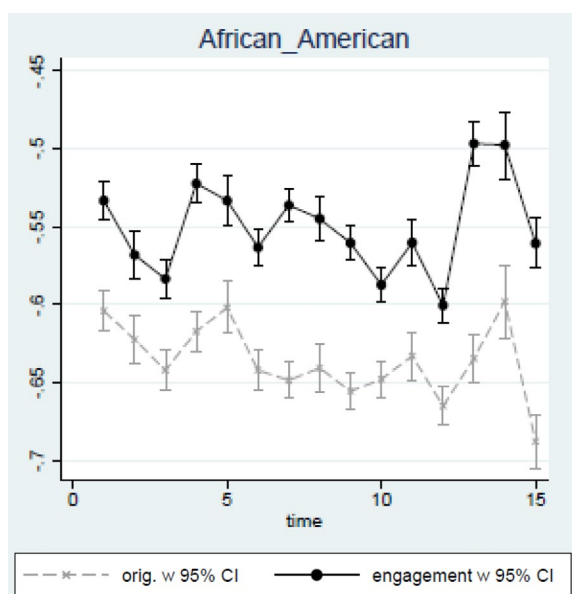
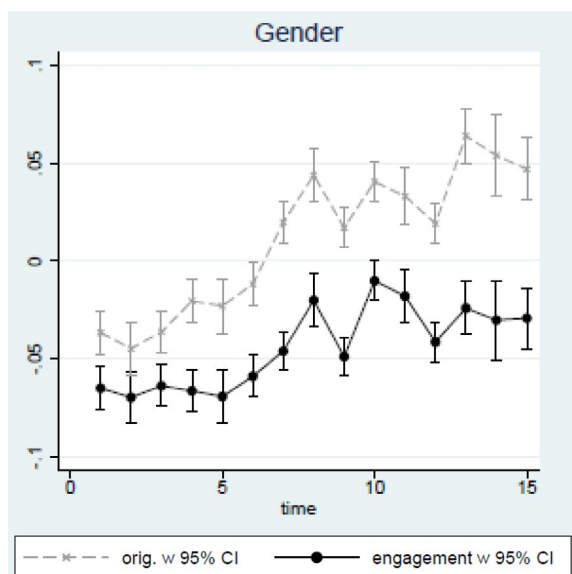
In most cases, shifts in achievement gaps were fairly modest when adjusted for test effort, but in three cases, resulting changes appeared meaningful.

First, male-female gaps appeared to be relatively sensitive to test effort. The gender gap in mathematics increased by around 200 percent in upper grades when rapid guessing was considered. This shift was large enough to reverse the direction of the mathematics gap in later grades, shifting from showing females exhibiting slightly higher average achievement when test effort was not taken into account, to favoring males by more than .05 standard deviations when effort was considered.

Second, the gap in reading achievement between African American students and white students differed substantially when test effort was considered, reaching more than 0.1 standard deviation in later grades.

Third, since rates of rapid guessing increased at different rates for different subgroups as students progressed through school, changes in gaps across grades also appeared somewhat sensitive, especially for male-female gaps.

The specific shifts in gap estimates this study found may be less important than the new lens it brings to understanding these gaps: that the gaps are not measures of students' intelligence or ability, but of performance. Observed scores, then, do not reflect only what students have learned in school, but are impacted by factors including students' motivation to perform on the test and students' experiences in school, in their families, and communities.



Achievement gaps accounting for test effort (black line) and unadjusted. **Top:** math achievement gap for female students; **bottom:** reading achievement gap between African American and white students. Gaps are shown in standard deviation units; axes scaled differently so that gaps can be examined more closely. See paper for data for other student subgroups. For male-female gaps, negative gap estimates (Y-axis) favor male students. For race-based gaps, negative estimates favor white students.

<sup>i</sup> Reardon, S. F., Robinson, J. P., & Weathers, E. S. (2008). Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps. *Handbook of Research in Education Finance and Policy*, 497-516.

<sup>ii</sup> Ma, X. (2008). Within-school gender gaps in reading, mathematics, and science literacy. *Comparative Education Review*, 52(3), 437-460.

<sup>iii</sup> Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, 42(4), 223-235.

<sup>iv</sup> Briesch, A. M., & Chafouleas, S. M. (2009). Review and analysis of literature on self-management interventions to promote appropriate classroom behaviors (1988-2008). *School Psychology Quarterly*, 24(2), 106.

## RECOMMENDATIONS

### **Consider the potential impact of differential test effort when interpreting achievement gaps.**

This study showed that test effort varied substantially across student gender and racial subgroups and that these differences in test effort can change some achievement gap estimates in meaningful ways.

Research often highlights that, while achievement gaps in math favoring males have existed for decades, these gaps appear to be narrowing. Some credit this narrowing to policies designed to support women in science- and math-related careers. This study found differences in test effort that may complicate this interpretation. Over the past decade, there has been much celebration about women “catching up” in mathematics. This study suggests that, when we compare apples to apples (those females and males who show effort on assessments), the gaps do not appear to be closing as students move through school. If the math achievement gap favoring males has narrowed over the past decade in part due to differential test effort rather than content mastery, then educational leaders trying to close gaps may not have a clear picture of the true size of gaps or of whether programs and policies designed to close gaps are working.

Other studies have used gaps to examine how practices and policies impacted disparities in black-white achievement. If gaps estimates are biased by differential rapid guessing, they are less likely to be useful as instruments of evaluation, and may even lead to ineffective programs being deemed effective or vice-versa.

While differential test effort did not have a large impact on most gaps estimates in this study, it did in some cases. Educators and policymakers should be aware of this, work to increase student engagement on tests, and consider using statistical approaches to identify and rescore tests to adjust for the bias rapid guessing can cause when it does occur.

This research also demonstrates the value of computer based assessments that measure and report test engagement, so educators in districts and states can better understand and address student test disengagement, and the impact it may have on achievement gaps, in their own contexts.

### **Broad conversation and additional research are needed to understand how test effort, academic effort, and social-emotional learning are related.**

The differences seen in rapid guessing rates in this study may make educational leaders wonder why differences in test effort fall down racial and gender lines? Why are some groups of students more motivated to take time and effort on tests to demonstrate what they know and can do academically? Other research has shown that academic motivation and self-management, which measures how focused students are on academic tasks, also decrease as students move through middle school and high school. Investigating why disengagement is so pervasive in some subgroups may shed light on student attitudes, beliefs, and aspirations about schooling that often drive academic disengagement. If test motivation is related to broader engagement in school, then addressing disengagement could help reduce gaps. If additional research finds that rapid guessing is related to social-emotional learning constructs like self-management, and to broader disengagement from school, then RTE may provide a useful proxy for these constructs, and so could inform interventions designed to close gaps.

#### **This brief describes research documented in:**

Soland, James. (2018). Are Achievement Gap Estimates Biased by Differential Student Test Effort? Putting an Important Policy Metric to the Test. *Teachers College Record*, 120(12),1-26. <http://www.tcrecord.org.22443>

#### **Suggested citation:**

Soland, James. (2019). Are Achievement Gap Estimates Biased by Differential Student Test Effort? (The Collaborative for Student Growth at NWEA Research Brief).

## ABOUT THE AUTHOR

Dr. Jim Soland is a Senior Research Scientist at the Collaborative for Student Growth at NWEA, and is an Adjunct Professor of Quantitative Methodology in the College of Education at Oregon State University. His research focuses on assessment and evaluation policy and practice, with particular emphasis on measuring social emotional learning, test engagement, and estimating teacher and school effectiveness. Soland completed a PhD in Educational Psychology at Stanford University with a concentration in measurement and policy.



## ABOUT THE COLLABORATIVE FOR STUDENT GROWTH

The Collaborative for Student Growth at NWEA® is devoted to transforming education research through advancements in assessment, growth measurement, and the availability of longitudinal data. The work of our researchers spans a range of educational measurement and policy issues including achievement gaps, assessment engagement, social-emotional learning, and innovations in how we measure student learning. Core to our mission is partnering with researchers from universities, think tanks, grant-funding agencies, and other stakeholders to expand the insights drawn from our student growth database—one of the most extensive in the world.



NWEA is a not-for-profit organization that supports students and educators worldwide by providing assessment solutions, insightful reports, professional learning offerings, and research services. Visit [NWEA.org](https://www.nwea.org) to find out how NWEA can partner with you to help all kids learn.

© 2019 NWEA. NWEA and MAP are registered trademarks, and MAP Growth is a trademark of NWEA in the US and in other countries.

JUN19 | KAP3850