Time of Day and High Stakes Cognitive Assessments

Alessio Gaggero† and Denni Tommasi‡

†University of Granada
‡University of Bologna and IZA

September, 2021
Introduction

Motivation

- High-stake **cognitive assessment** to measure **ability** (e.g., SAT/GRE test, job interview, job performance evaluation, etc)
  - Initial conditions can have **lasting consequences** (e.g., Wachter, 2020)
Introduction

Motivation

- High-stake **cognitive assessment** to measure **ability** (e.g., SAT/GRE test, job interview, job performance evaluation, etc)
  - Initial conditions can have **lasting consequences** (e.g., Wachter, 2020)

- **Time-of-day**: When can **ability** be measured at best?
  - Cognitive functioning varies during the day: **circadian rhythm**
Introduction

Motivation

- High-stake **cognitive assessment** to measure **ability** (e.g., SAT/GRE test, job interview, job performance evaluation, etc)
  - Initial conditions can have **lasting consequences** (e.g., Wachter, 2020)

- **Time-of-day**: When can **ability** be measured at best?
  - Cognitive functioning varies during the day: **circadian rhythm**
    - **Key idea**: *it does not imply* that **time-of-day matters for performance**
    - **Behavioral responses (adaptation) to non-optimal time of day**
Introduction

Motivation

- High-stake **cognitive assessment** to measure **ability** (e.g., SAT/GRE test, job interview, job performance evaluation, etc)
  - Initial conditions can have **lasting consequences** (e.g., Wachter, 2020)
- **Time-of-day**: When can **ability** be measured at best?
  - Cognitive functioning varies during the day: **circadian rhythm**
    - **Key idea**: *it does not imply that time-of-day matters for performance*
    - **Behavioral responses* (adaptation) to non-optimal time of day*
  - Whether time-of-day **matters** remains a **testable hypothesis**
**Introduction**

**Motivation**

- High-stake **cognitive assessment** to measure **ability** (e.g., SAT/GRE test, job interview, job performance evaluation, etc)
  - Initial conditions can have **lasting consequences** (e.g., Wachter, 2020)

- **Time-of-day**: When can **ability** be measured at best?
  - Cognitive functioning varies during the day: **circadian rhythm**
    - **Key idea**: it does not imply that time-of-day matters for performance
    - **Behavioral responses** (adaptation) to non-optimal time of day
      - Whether time-of-day **matters** remains a **testable hypothesis**

- **Challenge**: often we can **choose** our schedules
Introduction

Motivation

- High-stake **cognitive assessment** to measure **ability** (e.g., SAT/GRE test, job interview, job performance evaluation, etc)
  - Initial conditions can have **lasting consequences** (e.g., Wachter, 2020)

- **Time-of-day**: When can **ability** be measured at best?
  - Cognitive functioning varies during the day: **circadian rhythm**
  - **Key idea**: *it does not imply* that time-of-day **matters for performance**
  - **Behavioral responses** *(adaptation)* to non-optimal time of day
    - Whether time-of-day **matters** remains a **testable hypothesis**

- **Challenge**: often we can **choose** our schedules

- University: students take multiple **exams** *(~ cognitive assessments)* and their **performance** is incredibly relevant:
Introduction

Motivation

- High-stake **cognitive assessment** to measure **ability** (e.g., SAT/GRE test, job interview, job performance evaluation, etc)
  - Initial conditions can have **lasting consequences** (e.g., Wachter, 2020)

**Time-of-day**: When can **ability** be measured at best?

- Cognitive functioning varies during the day: **circadian rhythm**
  - **Key idea**: it does not imply that time-of-day matters for **performance**
  - **Behavioral responses** (adaptation) to non-optimal time of day
    - Whether time-of-day **matters** remains a **testable hypothesis**

**Challenge**: often we can **choose** our schedules

- University: students take multiple **exams** (≈ cognitive assessments) and their **performance** is incredibly relevant:
  - We study the impact of **time-of-day** on **cognitive performance**
Introduction

Motivation

- High-stake cognitive assessment to measure ability (e.g., SAT/GRE test, job interview, job performance evaluation, etc)
  - Initial conditions can have lasting consequences (e.g., Wachter, 2020)

- **Time-of-day**: When can ability be measured at best?
  - Cognitive functioning varies during the day: circadian rhythm
    → **Key idea**: it does not imply that time-of-day matters for performance
    → **Behavioral responses** (adaptation) to non-optimal time of day
      - Whether time-of-day matters remains a testable hypothesis

- **Challenge**: often we can choose our schedules

- University: students take multiple exams (~ cognitive assessments) and their performance is incredibly relevant:
  - We study the impact of time-of-day on cognitive performance
    → **Can efficiency gains be obtained by simple re-arrangements?**
In a nutshell

- Administrative **student-level panel** data from a large UK University:
  - **Complete information** on half million observations of University students
  - Exams performed at 9am, 1.30pm, 4.30pm
Introduction

In a nutshell

- Administrative **student-level panel** data from a large UK University:
  - **Complete information** on half million observations of University students
  - Exams performed at 9am, 1.30pm, 4.30pm

- **Identification:**
  - **Quasi-random assignment** of the day and time of exams
  - **Intuition:** students take different exams at different time, hence we randomize within students
Introduction

In a nutshell

- Administrative student-level panel data from a large UK University:
  - Complete information on half million observations of University students
  - Exams performed at 9am, 1.30pm, 4.30pm

- Identification:
  - Quasi-random assignment of the day and time of exams
  - Intuition: students take different exams at different time, hence we randomize within students

- Theory:
  - Trade-off between effort and rewards
Introduction

In a nutshell

- Administrative student-level panel data from a large UK University:
  - Complete information on half million observations of University students
  - Exams performed at 9am, 1.30pm, 4.30pm

Identification:

- Quasi-random assignment of the day and time of exams
- Intuition: students take different exams at different time, hence we randomize within students

Theory:

- Trade-off between effort and rewards
  - As the stakes of a cognitive assessment increase, the association between time and performance is less likely driven by low effort
  - Implication: The biological component may affect individual performances even in a high stakes environment
**Empirics**

*Main specification*

- We estimate:

\[ Y_{i,e,t} = \alpha(i) + T_{e,t}\beta + X_{i,e,t}\gamma + \nu_{i,e,t} \]

where:

- \( Y_{i,e,t} \) is the standardized mark achieved by student \( i \), in exam \( e \), in year \( t \)
- \( \alpha(i) \): student fixed effects (FE)
- \( T_{e,t} \): time of day variables
- \( X_{i,e,t} \): students’ and exams’ characteristics
- \( \nu_{i,e,t} \): unobservable shocks to students' exam mark
Results

Effects of Time of Day on Students’ Performance: Primary Results

<table>
<thead>
<tr>
<th></th>
<th>Primary Results</th>
<th>Robustness Checks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>$1(3.30pm = 1)$</td>
<td>0.046***</td>
<td>0.050***</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>$1(4.30pm = 1)$</td>
<td>0.050***</td>
<td>0.042**</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.017)</td>
</tr>
</tbody>
</table>

Test of equality of Early and Late Afternoon exams in linear regressions above:

| Pr > F | 0.815 | 0.623 | 0.036 | 0.032 | 0.313 | 0.021 | 0.033 |

Conditions: ✓ ✓ ✓ ✓ ✓ ✓ ✓
Student FE: ✓ ✓ ✓ ✓ ✓ ✓ ✓
Covariates: ✓ ✓ ✓ ✓ ✓ ✓ ✓
Duration FE: ✓ ✓ ✓ ✓ ✓ ✓ ✓
Exam FE: ✓ ✓ ✓ ✓ ✓ ✓ ✓
Day FE: ✓ ✓ ✓ ✓ ✓ ✓ ✓
Room FE: ✓ ✓ ✓ ✓ ✓ ✓ ✓

Clusters: 7665 7665 7626 6475 7373 7578 7626 7613
Observations: 503359 503359 500959 432185 312103 500906 500959 500920
Adjusted $R^2$: 0.000 0.022 0.462 0.471 0.452 0.531 0.466 0.473

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Observations are at the student-exam-year level. Standard errors are clustered by exam-year.

→ Students perform better early in the afternoon
→ Results are robust to several robustness checks (even more in the paper)
→ Inverse-U shape relationship between time of day and performance
Heterogeneous Analysis

*Based on the circadian rhythm literature*

- **Continuum of chronotypes** in population, from very early to very late, which can be assessed with questionnaires ([Roenneberg et al, 2003](#))

  - The clock is affected by day length / sunlight exposure ([Kantermann et al, 2007](#))
  - The problem with chronotypes and performance is not only sleep deprivation, but also being tested (or performing a task) at a non-optimal time-of-day ([Zerbini et al, 2017](#))
  - The clock is not "fixed" but it can vary, among other things, with age: chronotype of young adults is late ([Zerbini and Merrow, 2017](#))

→ All these (correlation) studies are consistent with our causal evidence
Heterogeneous Analysis

Based on the circadian rhythm literature

- **Continuum of chronotypes** in population, from very early to very late, which can be assessed with questionnaires (Roenneberg et al, 2003)

- The clock is affected by day length / sunlight exposure (season) (Kantermann et al, 2007)

---

Intro Conclusion
Time of day and Performance: By Sunlight

Graphical illustration

Figure: Time-of-Day and Performances
Heterogeneous Analysis

*Based on the circadian rhythm literature*

- **Continuum of chronotypes** in population, from very early to very late, which can be assessed with questionnaires (*Roenneberg et al, 2003*).

- The clock is affected by day length / sunlight exposure (season) (*Kantermann et al, 2007*).
Heterogeneous Analysis

*Based on the circadian rhythm literature*

- **Continuum of chronotypes** in population, from very early to very late, which can be assessed with questionnaires ([Roenneberg et al., 2003](#))

- The clock is affected by day length / sunlight exposure (season) ([Kanterman et al., 2007](#))

- The problem with chronotypes and performance is not only sleep deprivation, but also being **tested** (or performing a task) at a **non-optimal time-of-day** ([Zerbini et al., 2017](#))
Time of day and Performance: By Task

Graphical illustration

Figure: Time-of-Day and Performances

Panel A: Fall

Panel B: Spring
Heterogeneous Analysis

*Based on the circadian rhythm literature*

- **Continuum of chronotypes** in population, from very early to very late, which can be assessed with questionnaires ([Roenneberg et al., 2003](#))

- The clock is affected by day length / sunlight exposure (season) ([Kantermann et al., 2007](#))

- The problem with chronotypes and performance is not only sleep deprivation, but also being **tested** (or performing a task) at a **non-optimal time-of-day** ([Zerbini et al., 2017](#))
Heterogeneous Analysis

Based on the circadian rhythm literature

- **Continuum of chronotypes** in population, from very early to very late, which can be assessed with questionnaires (Roenneberg et al, 2003)

- The clock is affected by day length / sunlight exposure (season) (Kantermann et al, 2007)

- The problem with chronotypes and performance is not only sleep deprivation, but also being tested (or performing a task) at a non-optimal time-of-day (Zerbini et al, 2017)

- The clock is not “fixed” but it can vary, among other things, with age: chronotype of young adults is late (Zerbini and Merrow, 2017)
Time of day and Performance: By Age

Graphical illustration

Figure: Time-of-Day and Performance

Panel A: Fall

Panel B: Spring
Heterogeneous Analysis

*Based on the circadian rhythm literature*

- **Continuum of chronotypes** in population, from very early to very late, which can be assessed with questionnaires ([Roenneberg et al, 2003](#)).

- The clock is affected by day length / sunlight exposure (season) ([Kantermann et al, 2007](#)).

- The problem with chronotypes and performance is not only sleep deprivation, but also being **tested** (or performing a task) at a **non-optimal time-of-day** ([Zerbini et al, 2017](#)).

- The clock is not “fixed” but it can vary, among other things, with age: **chronotype** of young adults is **late** ([Zerbini and Merrow, 2017](#)).
**Heterogeneous Analysis**

*Based on the circadian rhythm literature*

- **Continuum of chronotypes** in population, from very early to very late, which can be assessed with questionnaires ([Roenneberg et al, 2003](#))

- The clock is affected by day length / sunlight exposure (season) ([Kantermann et al, 2007](#))

- The problem with chronotypes and performance is not only sleep deprivation, but also being **tested** (or performing a task) at a **non-optimal time-of-day** ([Zerbini et al, 2017](#))

- The clock is not “fixed” but it can vary, among other things, with age: **chronotype** of young adults is **late** ([Zerbini and Merrow, 2017](#))

→ **All these (correlation) studies are consistent with our causal evidence**
Conclusion
**Conclusion**

1. **Time of day matters** for your performance!

2. **Time of day matters** differentially (season, task, age)

3. Policy implications: **Efficiency gains** in education and elsewhere
Thank you!

Email: denni.tommasi@unibo.it

Twitter: @DenniTommasi