Digital Literacy and Online Political Behavior

Age, Skills, and the Second Digital Divide

Andrew M. Guess Kevin Munger

Princeton University

Penn State University

June 10, 2025

Outline

The Motivation: Zuckerberg's Senate Testimony

- April 2018: Mark Zuckerberg testifies before U.S. Senate
- Senators showed confusion about basic tech concepts
- "Senator, we run ads" - explaining Facebook's business model
- Senate committee median age: nearly 80
- Vivid demonstration of generational gaps in technological savvy

Evidence of Age-Based Digital Divides

Facebook's own research (2010):

- Digital voting experiment with "I Voted" stickers
- Effect size for 50+ years old vs. 18-24 years old:
  - Nearly 4x larger for self-reported voting
  - Nearly 8x larger for information seeking

2016 Election Misinformation Studies:

- Fake news consumption was small overall
- But very unequally distributed
- Much higher among older internet users

Research Questions

1. How should we conceptualize and measure **digital literacy**?
2. How does digital literacy vary across different populations?
3. What are the implications for online political behavior research?
4. How does sample selection bias affect our understanding of digital media effects?

Digital Literacy Framework

Our Definition:

Digital literacy = online information discernment combined with the basic digital skills necessary to attain it

Key Components:

- Basic technological fluency
- Information evaluation abilities
- Skills for effective use of online tools
- Awareness of digital threats and privacy

Historical Context: Media Modality Changes

Broadcast Era (Radio/TV):

- Easy to consume - universal skills
- Homogeneous audience experiences
- Limited content variety
- Theories focused on uniform effects

Internet Era:

- Requires specialized skills
- Highly heterogeneous experiences
- Unlimited content variety
- Heterogeneity should be the baseline expectation

Three Survey Instruments

1. **Internet Skills Scale (Hargittai):**
   - 21 questions on familiarity with tech terms
   - Examples: “app,” “hashtag,” “phishing,” “spyware”
   - Validated against behavioral assessments
2. **Power User Scale:**
   - 12 questions on technology interaction
   - “I make good use of most features available...”
   - Designed to identify high-skill users
3. **Low End Scale (Novel):**
   - Designed to identify low-skill users
   - Complements the Power User scale

Information Retrieval Tasks

Three Questions of Varying Difficulty:

1. Who is the Prime Minister of Croatia?
2. What is the capital city of Malawi?
3. What is the only U.S. National Park that begins with "T"?

Purpose:

- Validate survey measures against behavioral performance
- Test real-world information-seeking skills
- Explicitly allowed participants to search online

Five Different Samples

1. **Mechanical Turk (MTurk)** (N=503)
   - Expected to skew high-skill due to platform barriers
2. **Facebook Ads** (N=451)
   - Expected broader skill distribution
3. **High-skill targeted** (N=83)
   - Tech company employees
4. **Low-skill targeted** (N=18)
   - Computer skills class participants
5. **Lucid (National)** (N=2,146)
   - Quota-matched to U.S. demographics

Information Retrieval Performance

- High-skill sample performed best overall
- Low-skill sample performed worst
- MTurk sample outperformed Facebook sample
- Demonstrates substantial variation in basic online information-seeking abilities

Figure: Information Retrieval Accuracy Across Four Samples

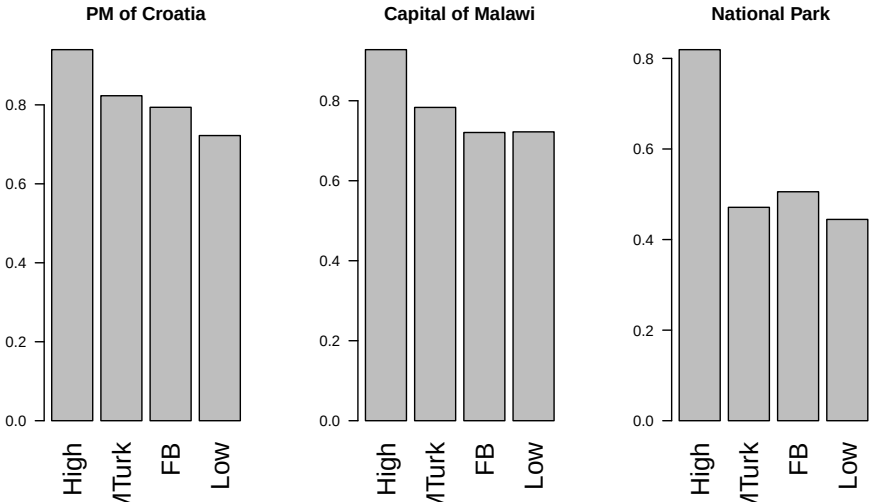
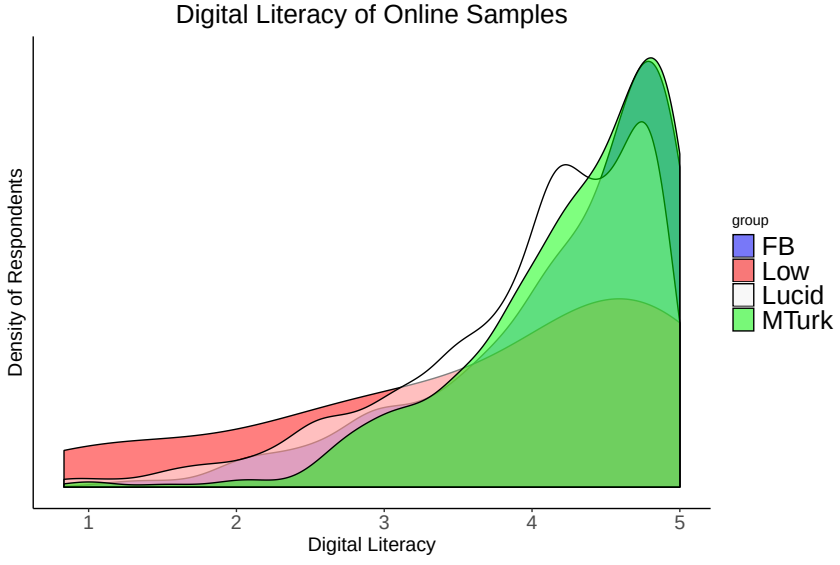
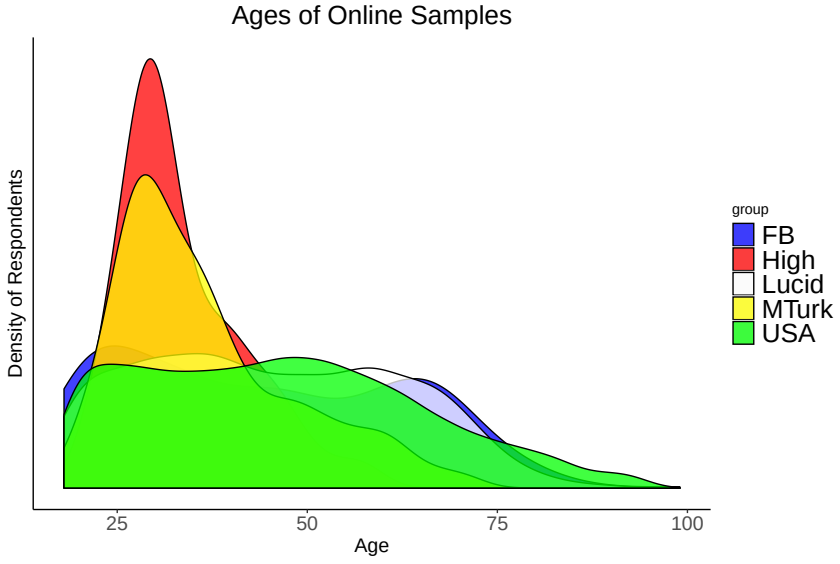


Figure: *


Digital Literacy of Online Samples
Density of Respondents
1
2
3
4
5
Digital Literacy
group
FB
Low
Lucid
MTurk


Ages of Online Samples
Density of Respondents
25
50
75
100
Age
group
FB
High
Lucid
MTurk
USA

Sample Composition Differences

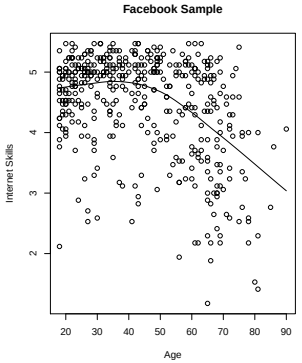
Age Distributions:

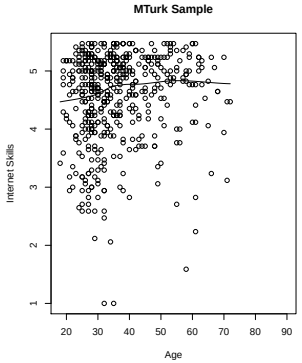
- MTurk & High-skill: Skewed young
- Facebook: Skewed slightly older
- Lucid: Matches Census demographics

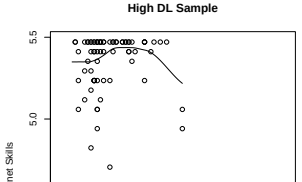
Digital Literacy Distributions:

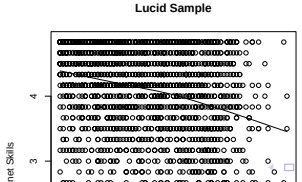
- MTurk: Hard floor at 2.5/5 on skills scale
- Facebook: More normal distribution
- Power User scale: MTurk identical to tech workers!

Age and DL

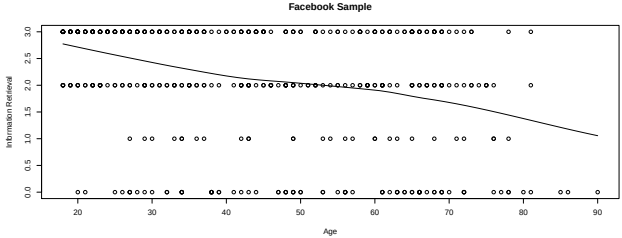

Facebook Sample
Internet Skills
Age
20 30 40 50 60 70 80 90
2 3 4 5

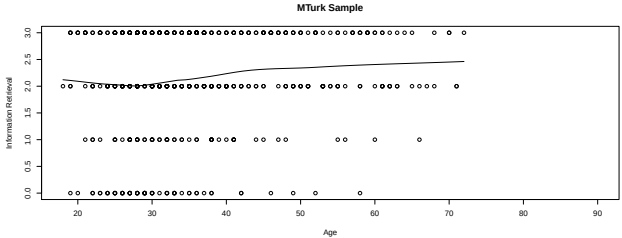

MTurk Sample
Internet Skills
Age
20 30 40 50 60 70 80 90
1 2 3 4 5

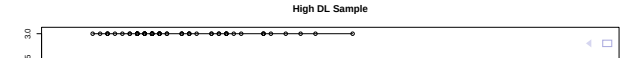

High DL Sample
5.0 5.5


Lucid Sample
3 4

Age and Information Retrieval







Critical Finding: The MTurk Problem

Age and Digital Literacy Correlations:

- **Facebook & Lucid samples:** Strong negative correlation between age and digital skills
- **MTurk sample:** NO correlation between age and digital skills
- **Implication:** MTurk structurally excludes low-skill users

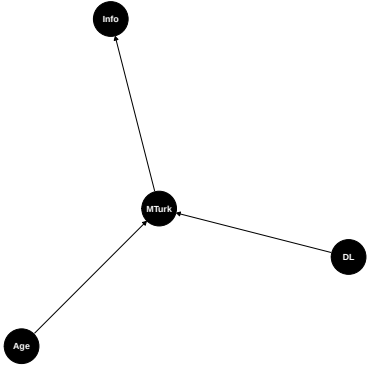
The Selection Bias:

- Only 1.4% of MTurk users below digital literacy threshold
- Compare to 16.7% of low-skill sample, 5.1% of Facebook sample
- "Conditioning on a collider" - selecting on the dependent variable

	Facebook Sample				*MTurk Sample*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Skills	0.548*	0.492*	0.484*	0.441*	0.462*	0.360*	0.439*	0.396*
	(0.073)	(0.080)	(0.087)	(0.086)	(0.068)	(0.079)	(0.083)	(0.085)
Low		−0.118*	−0.113	−0.081		−0.133**	−0.151*	−0.155*
		(0.072)	(0.075)	(0.075)		(0.053)	(0.054)	(0.053)
Power			0.008	−0.037			−0.092*	−0.072*
			(0.033)	(0.034)			(0.035)	(0.035)
Age				−0.013*				0.009*
				(0.003)				(0.004)
Cons.	0.033	0.498	0.483	1.349*	0.377	1.016*	1.219*	0.940*
	(0.267)	(0.388)	(0.393)	(0.431)	(0.253)	(0.359)	(0.365)	(0.379)

The Collider Problem

Directed Acyclic Graph



Result: Within MTurk sample, age-digital literacy relationship is broken
Analogy: Like studying height-performance relationship only among NBA players

Recommendations for Researchers

Sample Selection:

- Avoid MTurk for studies of digital media effects
- Consider theoretical relevance of digital literacy
- Use samples with sufficient variation in skills

Measurement:

- No universal formula for digital literacy
- Choose measures based on theoretical arguments
- Consider multiple dimensions of the concept

Theory:

- Make theoretically informed sampling decisions
- Consider effect heterogeneity as baseline expectation

Methodological Contributions

For Political Science:

- Framework for incorporating digital literacy
- Validated survey instruments
- Documentation of selection bias in common samples

For Online Research:

- "Fit for purpose" sampling approach
- Importance of theoretical justification
- Moving beyond demographic matching

Key Takeaways

1. **Digital literacy matters** for online political behavior
2. **Substantial variation exists** in the population
3. **Sample selection bias** is a serious methodological concern
4. **Theoretical reasoning** should guide both sampling and measurement decisions
5. **Heterogeneity** should be the baseline expectation for digital media effects