

# Optimal response formatting for fixed-field data items

Ivonne Figueroa

U.S. Census Bureau

AAPOR 72<sup>st</sup> Annual Conference

May 20, 2017

New Orleans, Louisiana

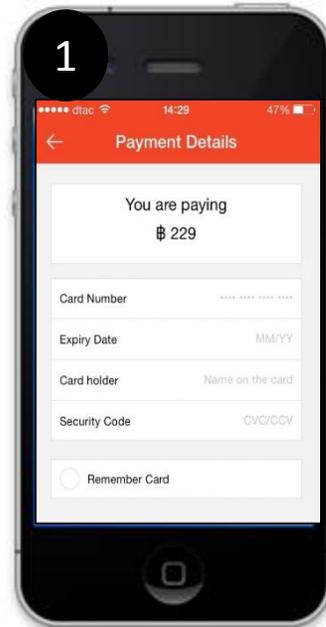
# Mobile phone usage

- 77% of Americans own smartphones
- <\$30K: Primary device for connection to the digital world
- Results in increased data entry on mobile phones.

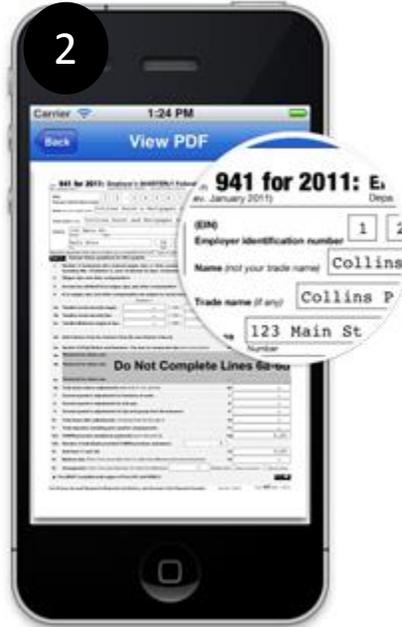
# Types of data entry

Census

Order food



File taxes



3



# What are fixed-field formats?

- Formats for entering numerical data with a fixed amount of digits.
  - Examples- phone #'s, date of birth (DOB), Login ID
- Manual versus automatic advancement?
- **Problem:** Users do not know how the fields operate, and what action is required to advance to the next field.
  - How is data quality and performance affected?



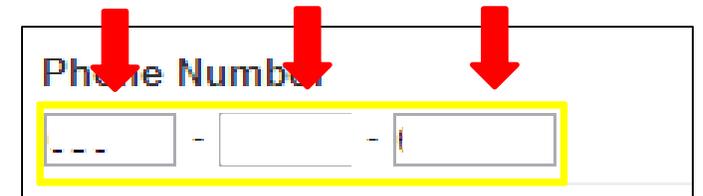
# Most common fixed-field formats

- Masking
  - Format embedded within the field once user activates field.
- Auto-tab
  - Separate fields available for each 'chunk' of string entry

The screenshot shows a form field for a phone number. The field is titled "Phone number" and contains a text input box with a vertical cursor and two hyphens. A tooltip box on the left contains the text: "1. Hyphens appear automatically when field is activated and cursor is in box." Below the input box, the placeholder text "XXX-XXX-XXXX" is visible. To the right of the input box, there is a red warning icon and the text: "Important: Please complete this required field".

# Most common fixed-field formats

- Masking
  - Format embedded within the field once user activates field.
- Auto-tab
  - Separate fields available for each 'chunk' of string entry



The diagram shows a form field labeled "Phone Number". The input area is divided into three segments by hyphens: "---" - " " - " ". Three red arrows point down to the top of each segment, indicating that the cursor automatically moves to the start of the next segment as the user types. A yellow border highlights the entire input area.

# Masking vs Auto-tab

	Masking
Pros	-Format provided ahead of time (for user to anticipate) -Can serve as a cue
	Easy entry

# Masking vs Auto-tab

	Masking	Auto-tabling
Pros	<ul style="list-style-type: none"><li>-Format provided ahead of time (for user to anticipate)</li><li>-Can serve as a cue</li></ul>	<ul style="list-style-type: none"><li>-Can advance users through long strings of numeric entries</li><li>-Number of text fields can serve as a cue</li></ul>
	Easy entry	Easy to adapt to (even if errors are initially made)

# Masking vs Auto-tab

	Masking	Auto-tabling
Pros	<ul style="list-style-type: none"><li>-Format provided ahead of time (for user to anticipate)</li><li>-Can serve as a cue</li></ul>	<ul style="list-style-type: none"><li>-Can advance users through long strings of numeric entries</li><li>-Number of text fields can serve as a cue</li></ul>
	Easy entry	Easy to adapt to (even if errors are initially made)
Cons	User performance may be slower because they still attempt to manually enter hyphens and special characters.	User isn't always expecting to be advanced automatically.

# Masking vs Auto-tab

- Both masking and auto-tabbing formats allow for ‘chunking.’
  - A strategy that enables recall of items (Miller, 1957).
  - Example: phone numbers

The image shows two side-by-side input fields for a phone number. The left field, titled "Phone Number", is divided into three separate boxes by hyphens, representing a masked format. The right field, titled "Phone number", is a single wide box with a blue border and a vertical cursor, representing an auto-tabbing format. Below the right field, the placeholder text "XXX-XXX-XXXX" is visible.

Easier to recall **3** ‘chunks’ of information rather than **10** individual items.

# Optimal response format: Masking or Auto-tab?

- Which format allows users to enter strings of fixed-field data:
  - More accurately? (Effectiveness)
  - Faster? (Efficiency)
  - Better satisfaction? (Subjective)

# Method

# Experimental design

- 57 participants (random assignment)
    - 228 total cases- 14 missing cases = 214
  - Mixed design
  - Independent variables: Formatting type, memory
  - Dependent variables: time, satisfaction rating, deviations
  - Four tasks (counterbalanced)
    - Phone number
    - Date of birth
    - Login
    - Credit card
- From memory
- Not from memory

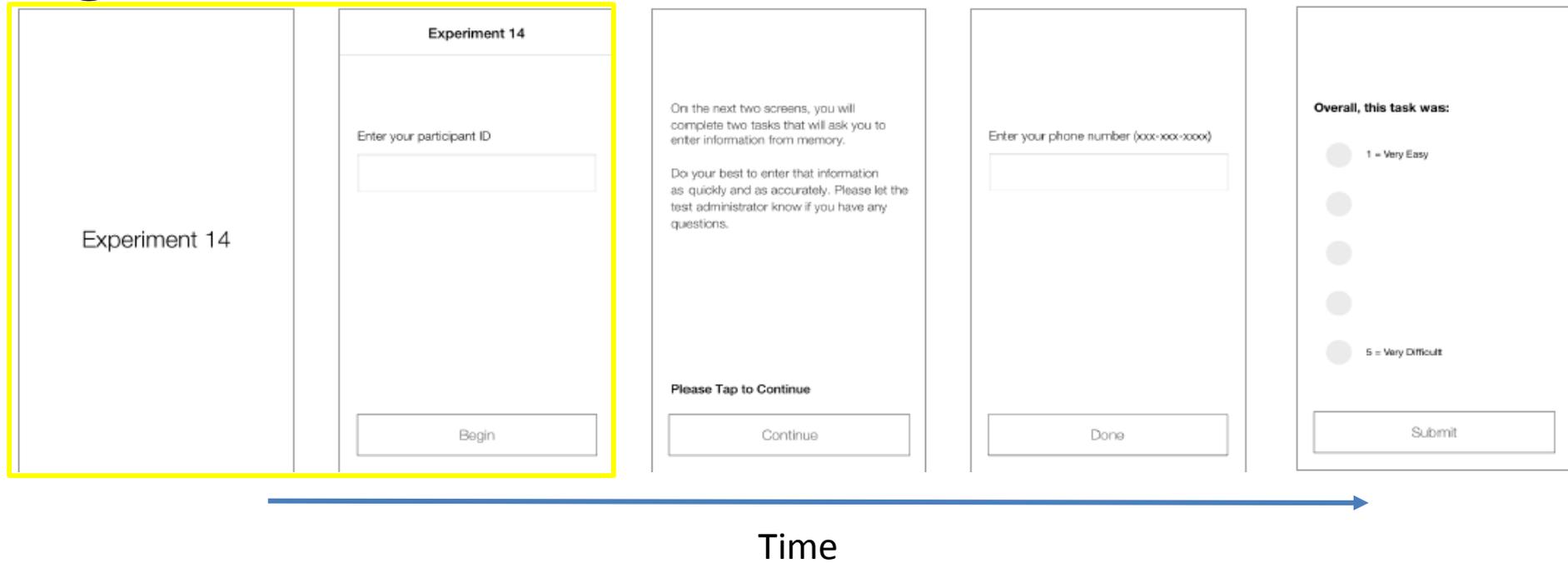
# Mixed design

Within subjects

		Memory		Not from memory	
		Phone #	DOB	Login	Credit Card
Between subjects	Formatting type				
	Auto-tab <i>n</i> = 110	-Time (secs) -Difficulty ratings -Deviations			
	Masking <i>n</i> = 104	-Time (secs) -Difficulty ratings -Deviations			

# Procedure

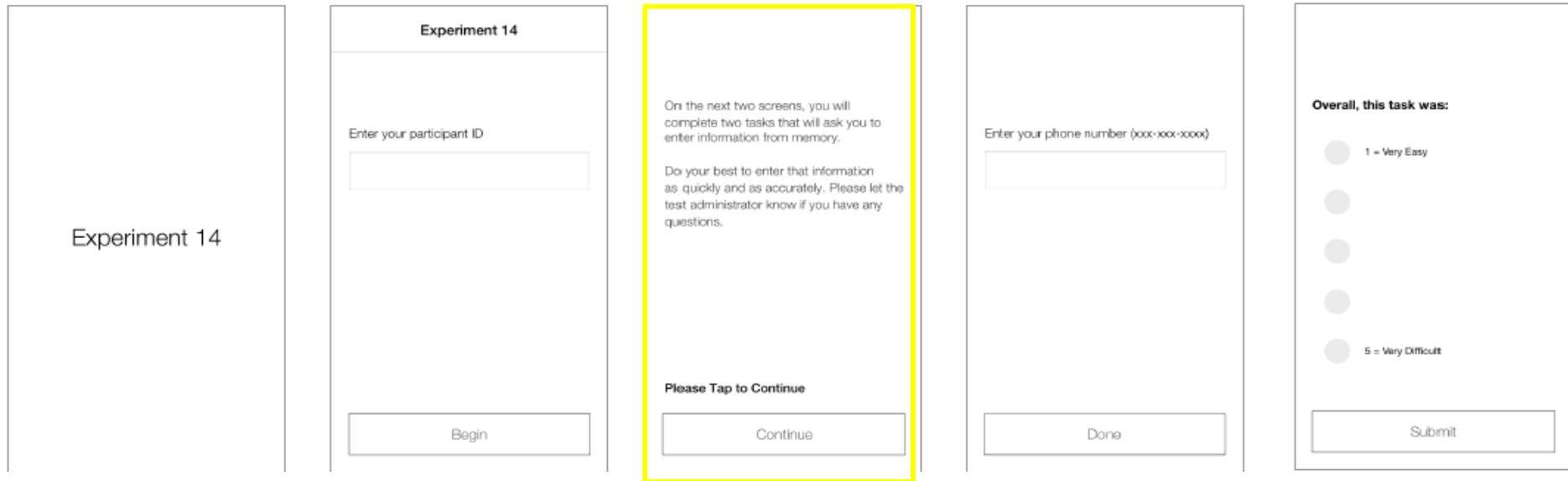
## 1 Test admin loads exp.



# Procedure

1 Test admin loads exp.

2 Instructions



“On the next two screens, you will complete two tasks that will ask you to enter information **[from memory/not from memory]**.

Do your best to enter that information as quickly and as accurately as possible. Please let the test administrator know if you have any questions.”

# Procedure

## 1 Test admin loads exp.

Experiment 14

Experiment 14

Enter your participant ID

Begin

## 2 Instructions

On the next two screens, you will complete two tasks that will ask you to enter information from memory.

Do your best to enter that information as quickly and as accurately. Please let the test administrator know if you have any questions.

Please Tap to Continue

Continue

## 3 Tasks and difficulty rating

Enter your phone number (xxx-xxx-xxxx)

Done

Overall, this task was:

1 = Very Easy



(Repeat X4)

Submit



Time

# Format properties- Hyphenating

Masking

Enter your date of birth (MM-DD-YYYY)

Next

Hyphens appear when user activates field.

Auto-tab

Enter your date of birth (MM-DD-YYYY)

-  -

Done

Hyphens are static.

# Format properties- Time tracking

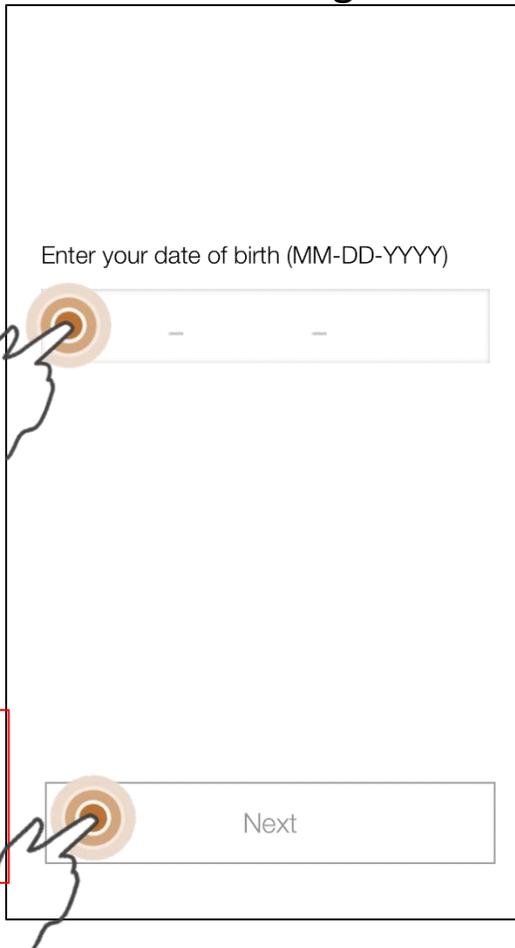
Masking

Time 1 begins when user activates field.

Enter your date of birth (MM-DD-YYYY)

Time 2 ends when user taps 'next.'

Next



Auto-tab

Enter your date of birth (MM-DD-YYYY)

Done



# Format properties- Backspacing

## Masking

Enter your date of birth (MM-DD-YYYY)

**xx** - **12** - **xxxx**

Next

Respondent can backspace without manually tapping into field 2.

# Format properties- Backspacing

Auto-tab

Enter your date of birth (MM-DD-YYYY)

-  -

Done

Masking

Enter your date of birth (MM-DD-YYYY)

Next

Respondent can backspace without manually tapping into field 2.

# Number of optimal entries per task

Task	Help text	# of optimal entries
Phone number	(xxx - xxx- xxxx)	10
Date of birth	(mm - dd- yyyy)	8
Login	(xxx-xxxx)	7
Credit card	(xxxx- xxxx- xxxx- xxxx)	16

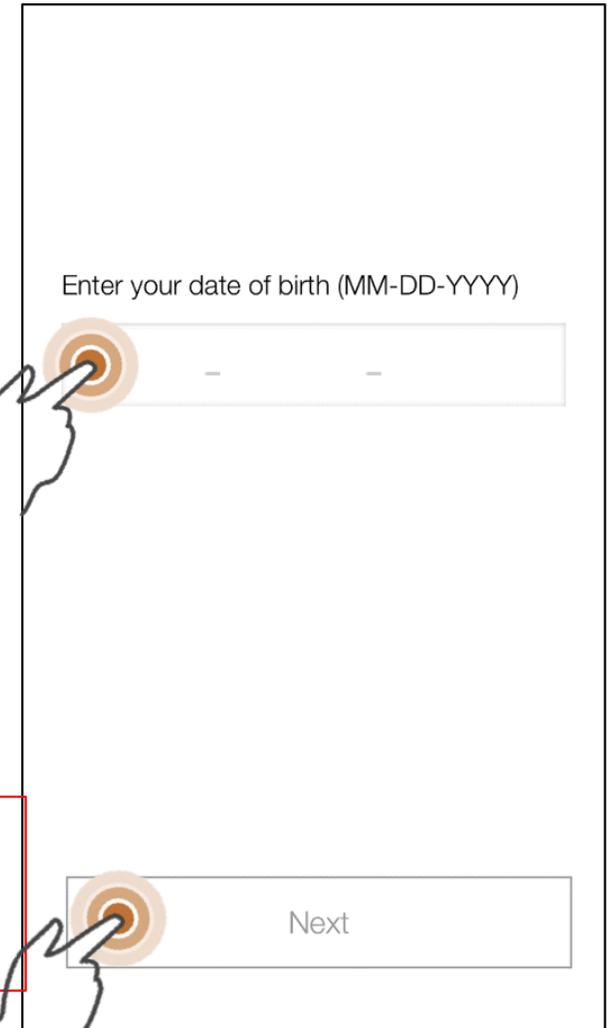
# Metrics of data quality and performance

## 1. Efficiency:

Time on task ( $t_{Norm}$ ) = Time2 – Time 1  
optimal # entries

Time 1  
begins  
when user  
activates  
field.

Time 2 ends  
when user  
taps 'next.'



Enter your date of birth (MM-DD-YYYY)

Next

The image shows a mobile form with two input fields. The top field is for a date of birth, with a placeholder 'Enter your date of birth (MM-DD-YYYY)' and a text input area containing dashes. The bottom field is a 'Next' button. A hand is shown tapping the top field, and another hand is shown tapping the 'Next' button. Red boxes with text annotations point to these actions: 'Time 1 begins when user activates field.' points to the top field, and 'Time 2 ends when user taps 'next.'" points to the 'Next' button.

# Metrics of data quality and performance

Efficiency- time on task

Effectiveness- deviations

- Extra entries made that deviate from the optimal entries.

Examples of deviations: Backspaces, Touch (taps) outside of field

Task	Help text	# of optimal entries
Phone number	(xxx - xxx- xxxx)	10
Date of birth	(mm - dd- yyyy)	8
Login	(xxx-xxxx)	7
Credit card	(xxxx- xxxx- xxxx- xxxx)	16

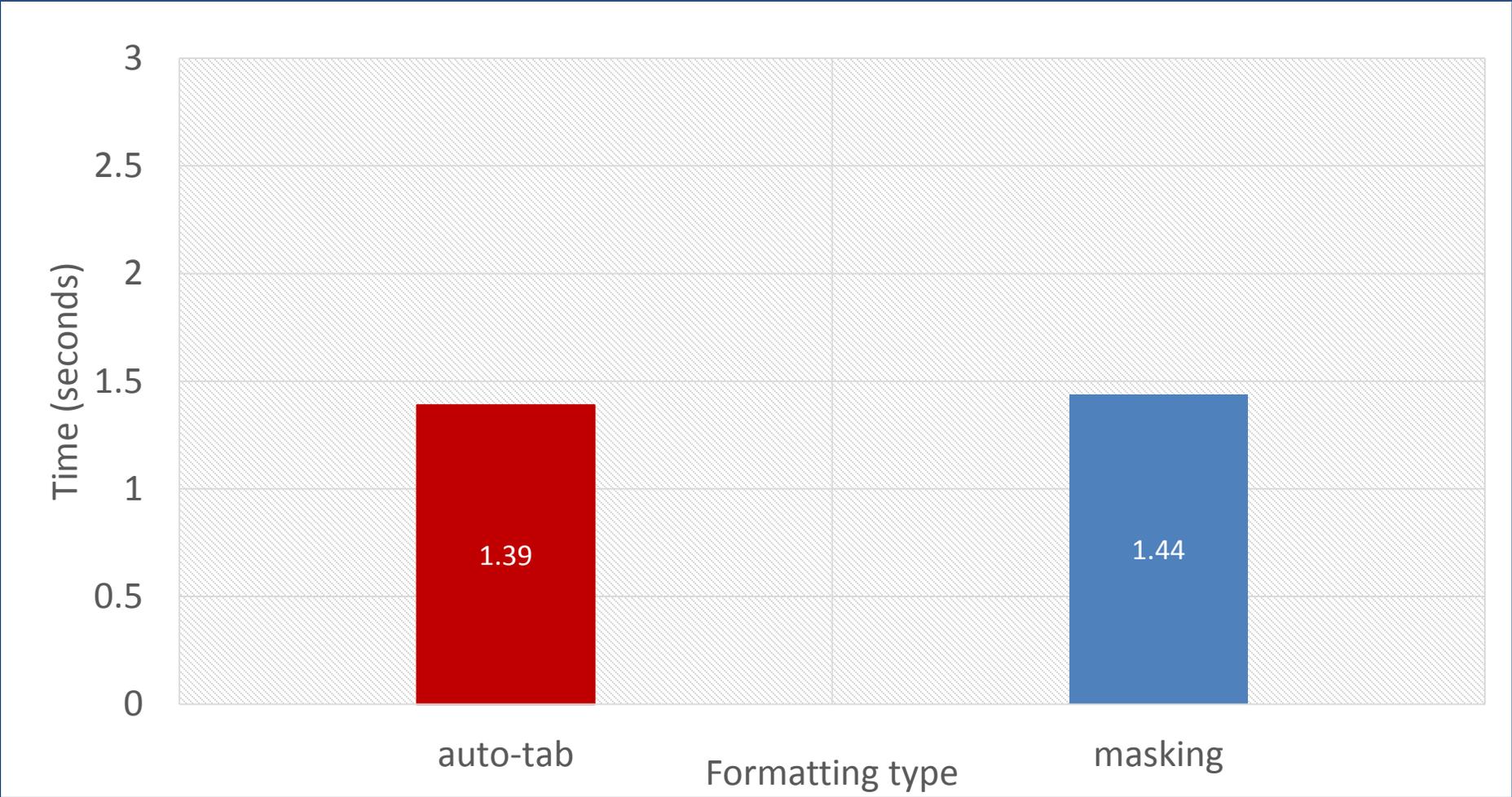
$$\text{Deviations } (dev_{norm}) = \frac{\text{total \# of entries} - \text{optimal}}{\text{optimal}}$$

# Metrics of data quality and performance

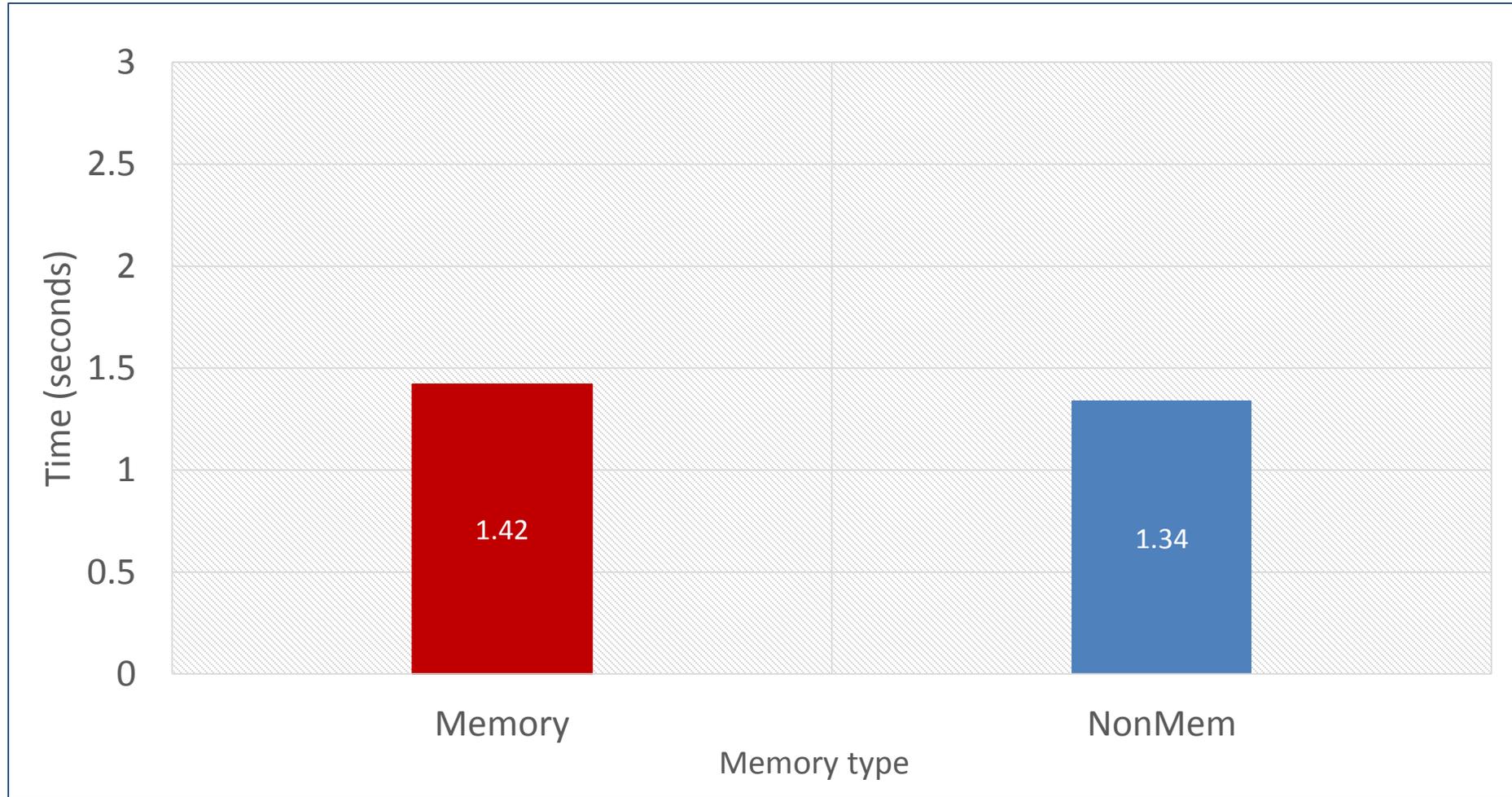
1. Efficiency- time on task
2. Effectiveness- deviations
3. Subjective- difficulty ratings
  - Rating the ease of entering data on a scale of 1 – 5, with ‘1’ being *very easy* and ‘5’ being *very difficult*.

# Results

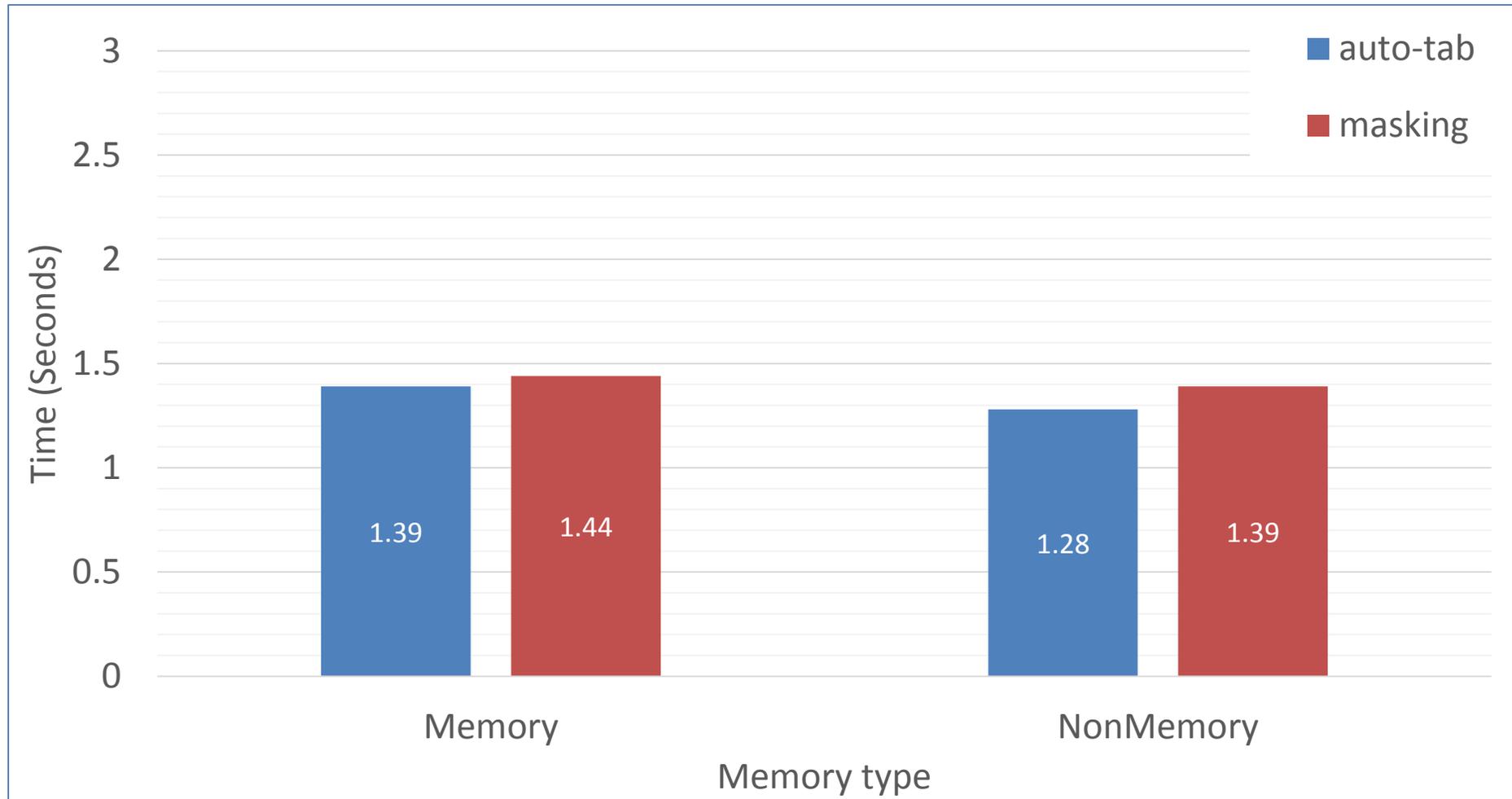
# Mean time by formatting type



# Mean time by memory

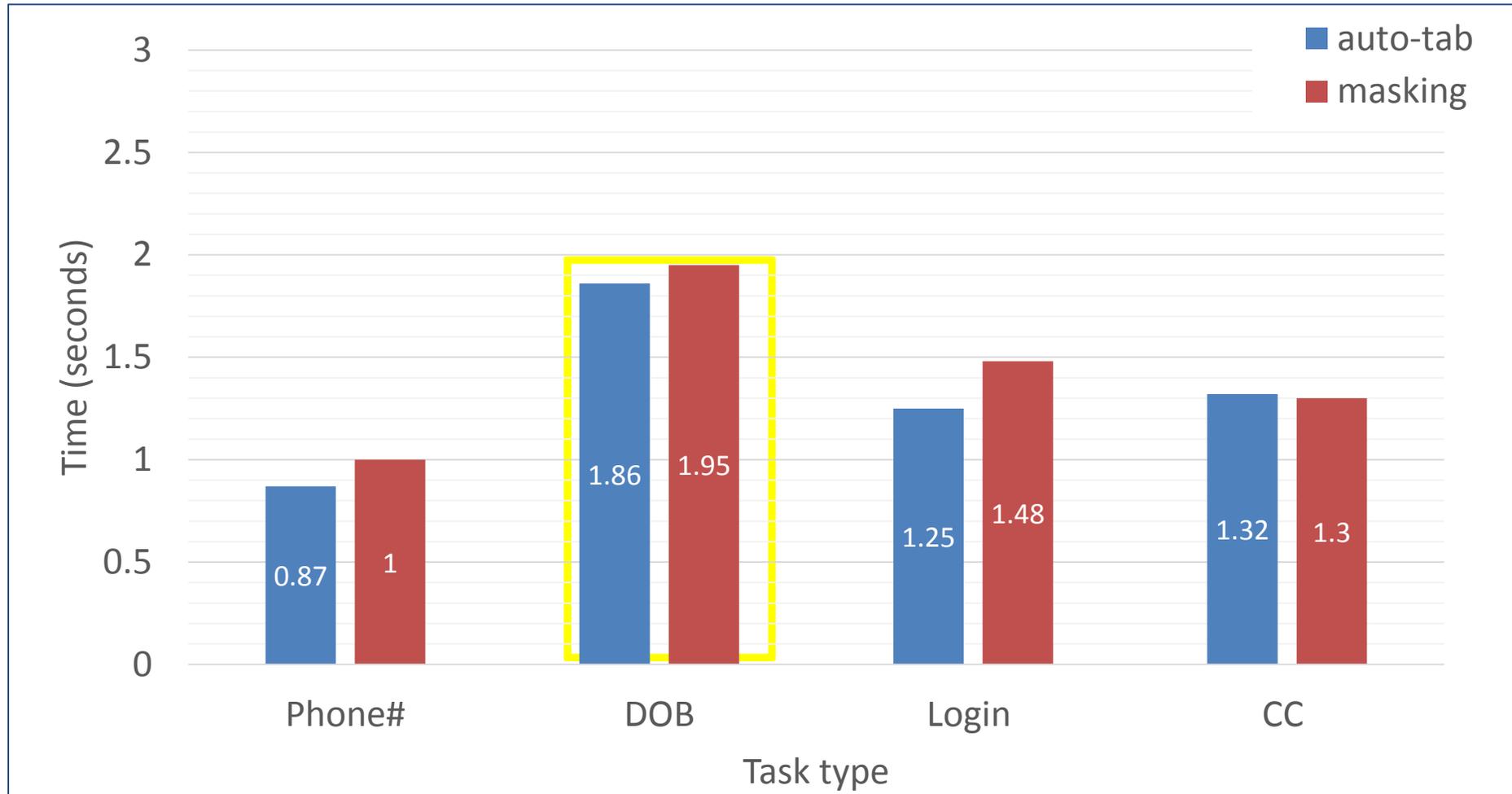


# Mean time by formatting and memory



There are no differences in time by formatting type or memory type.

# Mean time on task

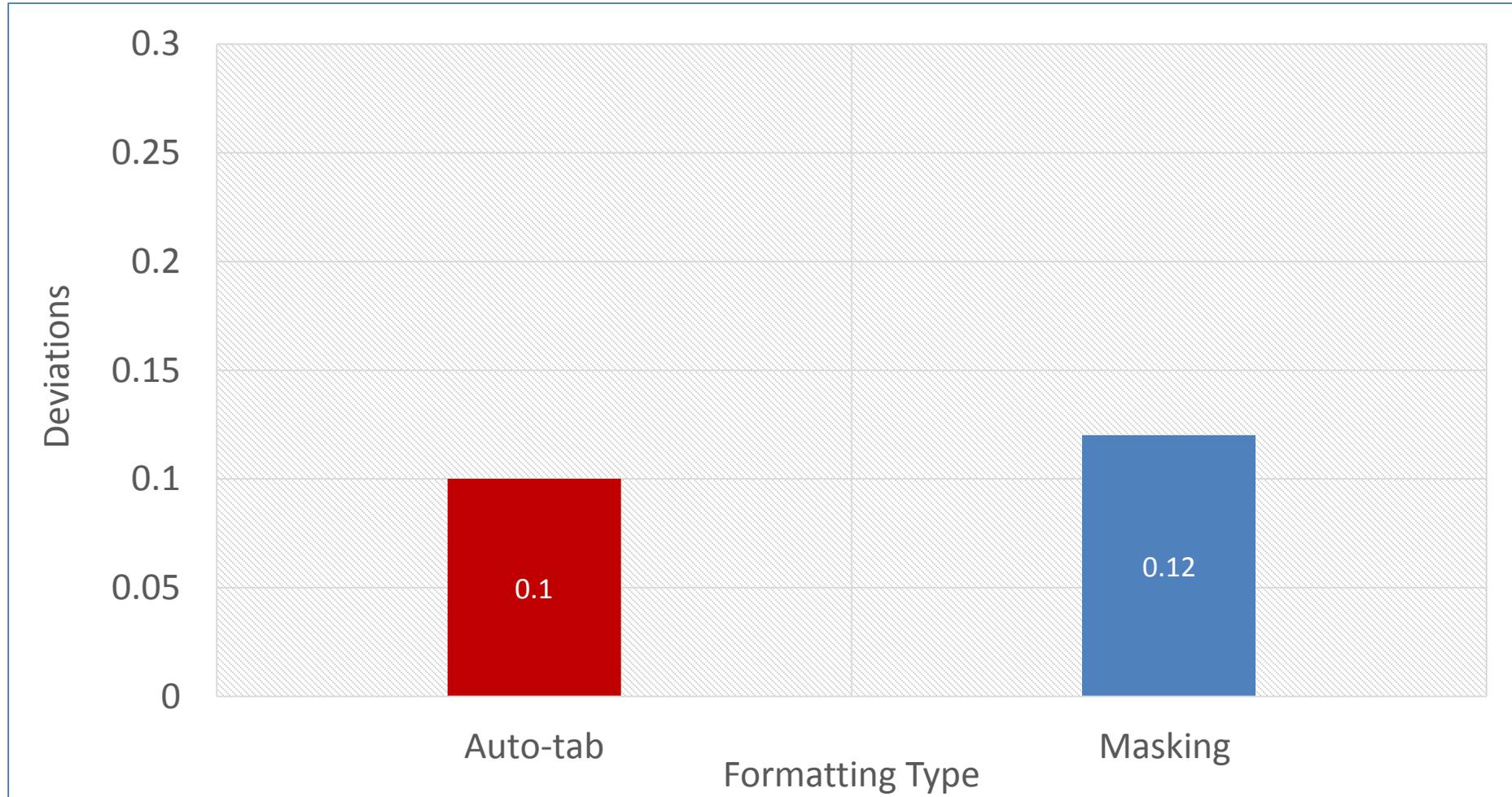


Entering data for DOB takes longer than other tasks.

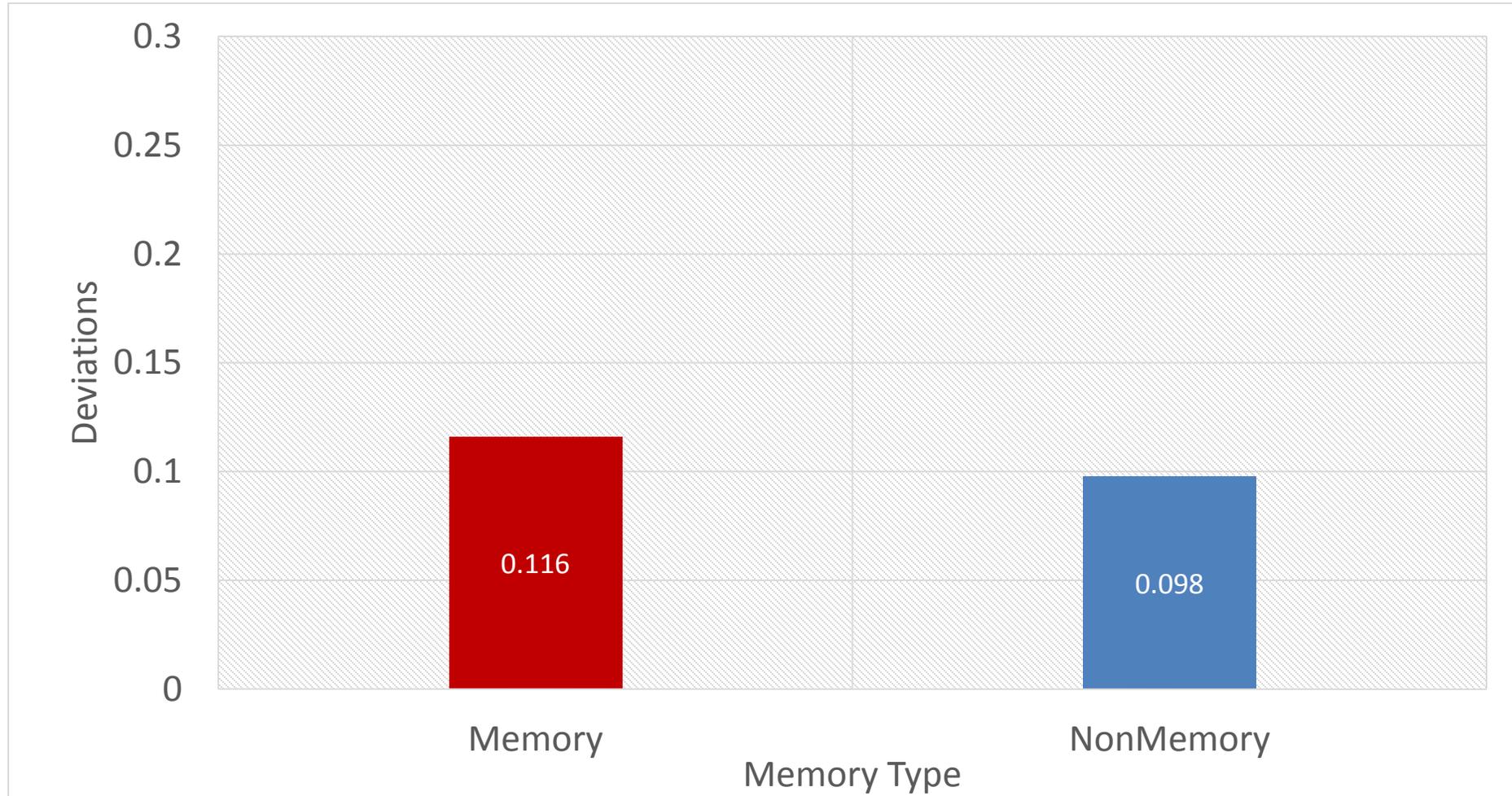
## Results: Time on task (efficiency)

- No differences in time between auto-tab or masking.
  - Neither format affects speed of performance
- No differences in time between entering data from memory versus not from memory.
- DOB entries take longer than other tasks.
  - Possibly from respondent's prior experience with differing date of birth formats. (e.g., mm/dd/yyyy; m/d/yy)

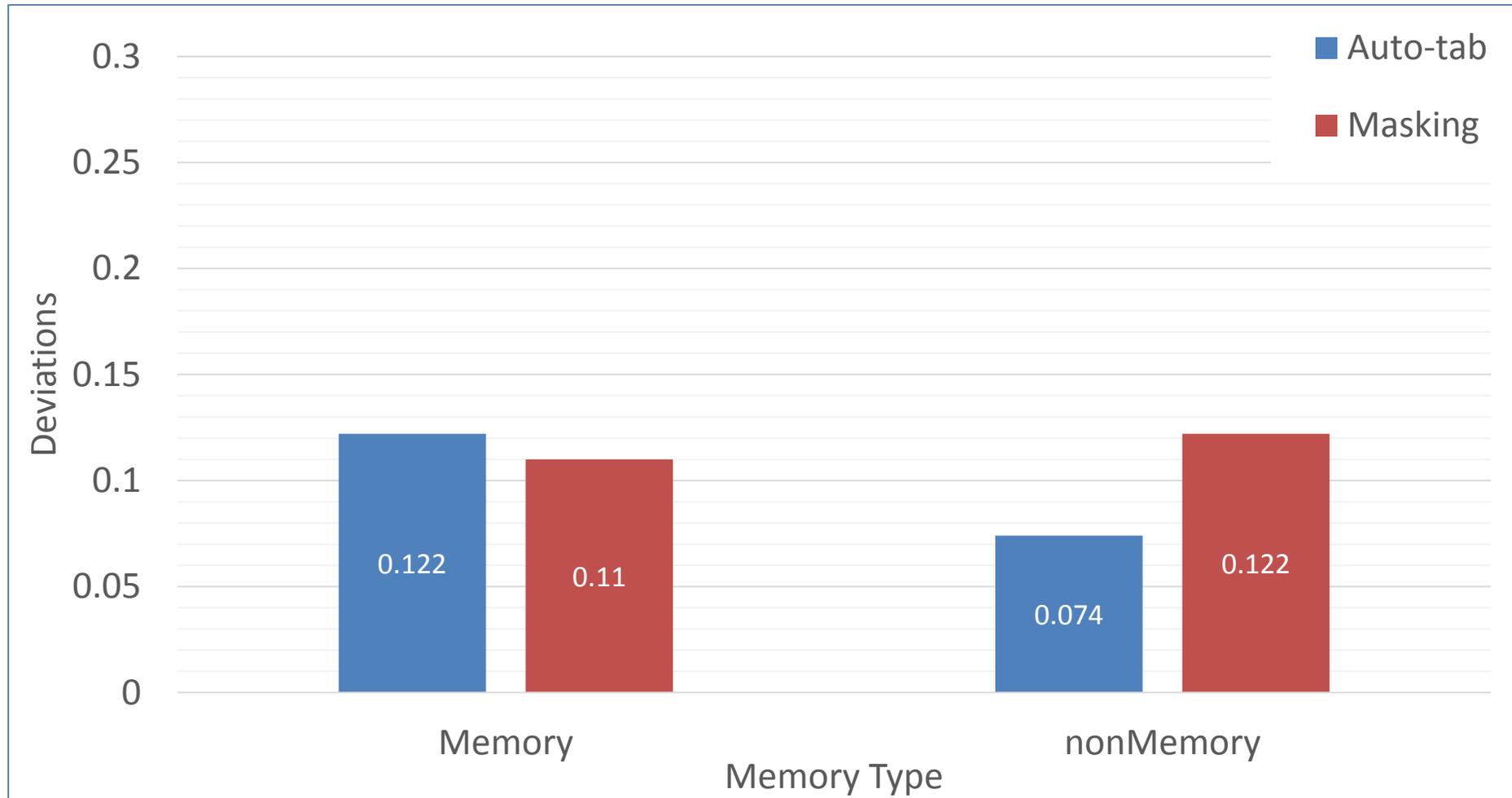
# Deviations by formatting type



# Deviations by memory

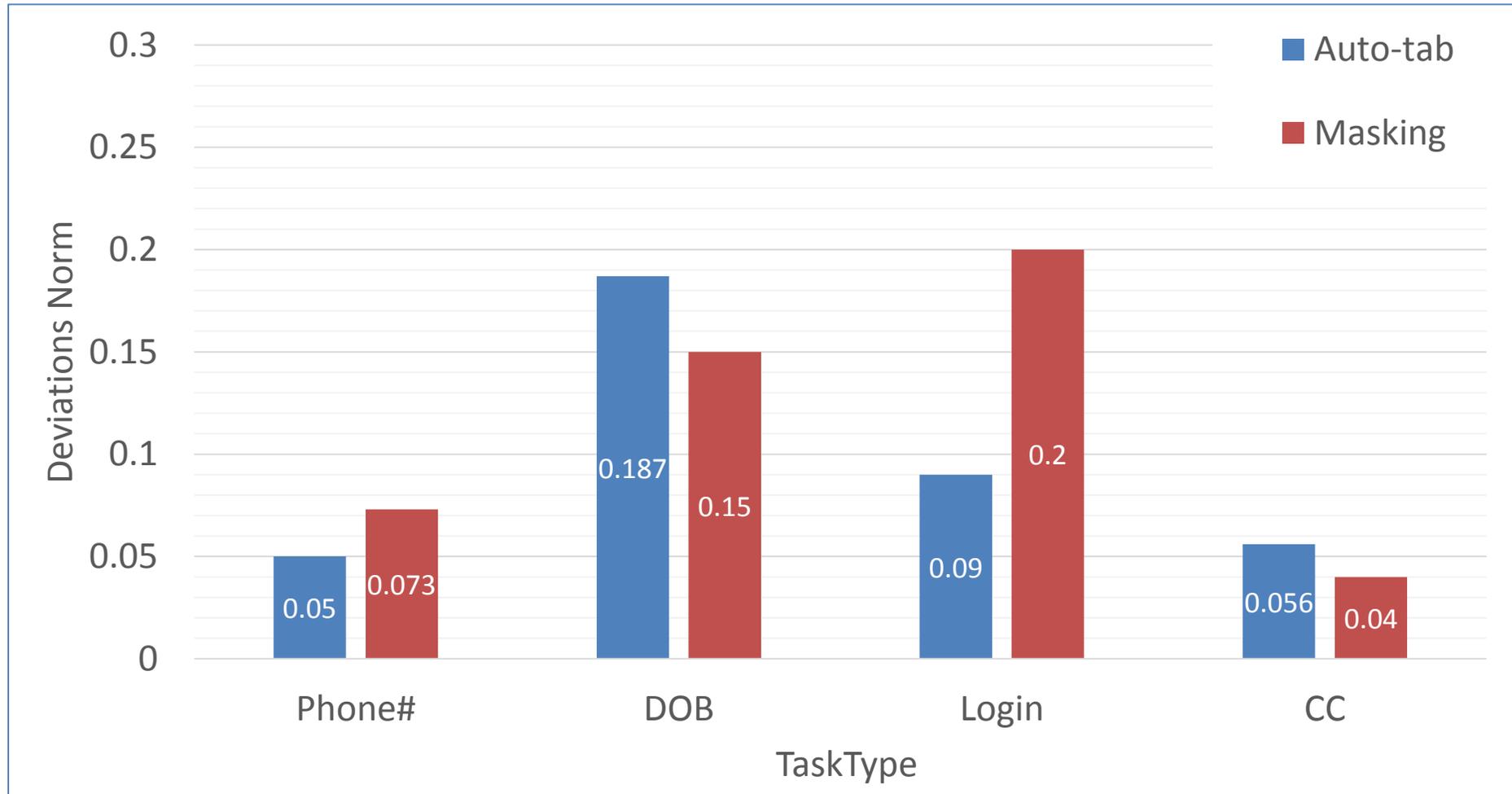


# Deviations by formatting and memory



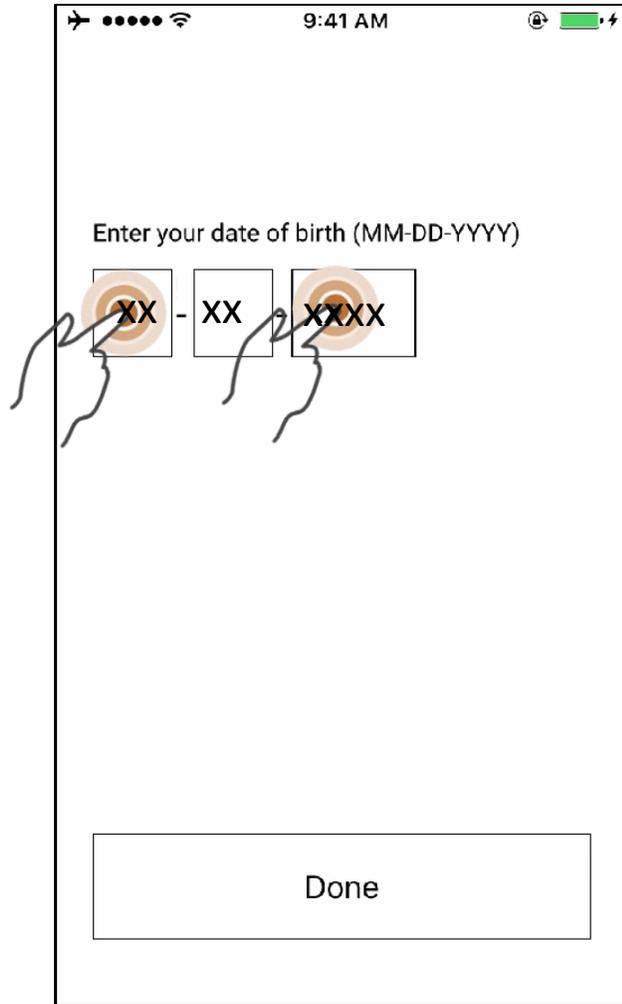
There are no differences in deviations by formatting type or memory type.

# Deviations by task



Formatting has differential effects on different types of tasks.

# Task analysis for entering date of birth



## Steps

1. Taps into field 1
2. Enters digits for month
3. Automatically advanced to field 2
4. Enters digits for day
5. Automatically advanced to field 3
6. Enters 2 digits for year
7. Taps into field 3 and deletes two digits for year
8. Automatically sent back to field 2
9. Enters digit and automatically sent back to field 3
10. Enters for digits for year.

# Task analysis for entering date of birth

9:41 AM

Enter your date of birth (MM-DD-YYYY)

XX - XX - XXXX

Done

- Participants prior expectations of date of birth formats may affect their response times deviations from optimal entries.

# Results: Number of deviations

- Combining data from all four tasks, no statistically significant differences in time between auto-tab or masking were found, nor in deviations from optimal entry.
- Preliminary results suggest that there are differences in deviations between auto-tabbing and masking for different tasks (e.g., DOB vs Login).

# Difficulty ratings by condition

There are no differences in satisfaction ratings between masking or auto-tabbing formats.

		Auto-tab	Masking
Difficulty rating	1-very easy	93	84
	2	10	15
	3	7	4
	4	0	1
	5-very difficult	0	0

$$X^2(3) = 3.11, p > .05$$

# Conclusions

- There are no differences in time, difficulty ratings, or deviations between masking or auto-tabbing conditions when data from all four tasks pooled together.
- There appear differences in formatting effects on different data entry tasks. Further research at task level is warranted.

Thank you!  
Questions?

# Optimal no. of entries versus time

Task	# of optimal entries	Avg. # of deviations	Time (secs)
Phone number	10	.06	.49
Date of birth	8	.16	1.91
Login	7	.15	1.37
Credit card	16	.05	1.31

DOB and Login take more time for data entry and have more average deviations than other tasks.