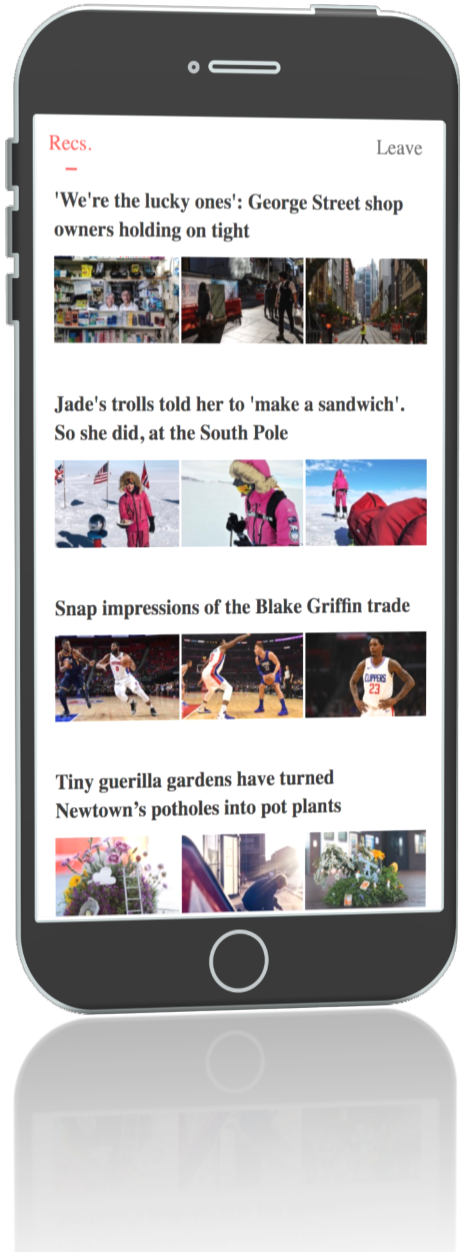



Tsinghua University

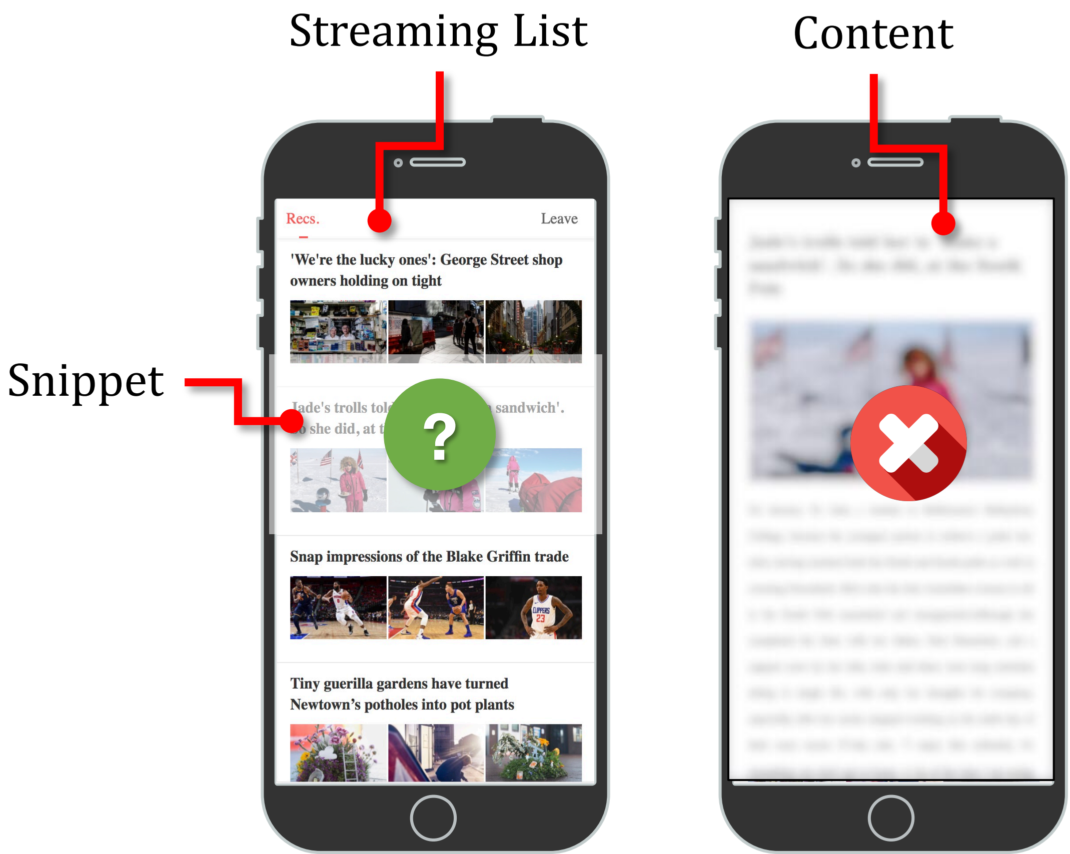
# **Quality Effects** on **User Preferences** and **Behaviors** in Mobile News Streaming

**Hongyu Lu**, Min Zhang, Weizhi Ma, Yunqiu Shao, Yiqun Liu, and Shaoping Ma

Tsinghua University

luhy16@mails.tsinghua.edu.cn

# Mobile News Streaming




**As Implicit feedback**

E.g. Train recommender system

**For evaluation**

E.g. Click-Through Rate, MRR..

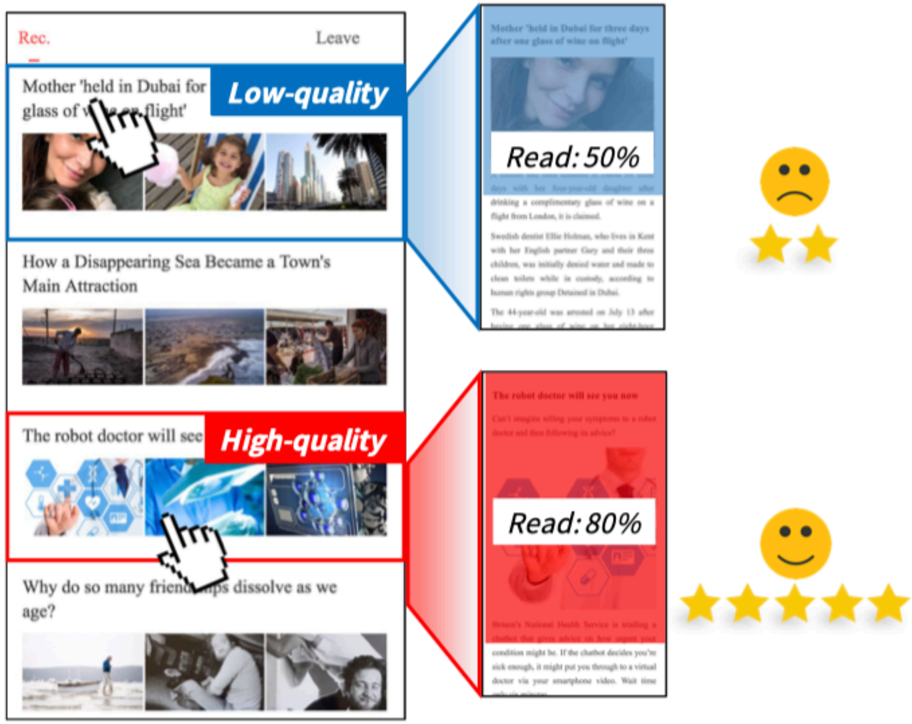
- ✦ Low-quality news exists
- ✦ When user read low-quality news?

  How they behave? (user behavior)

  How they experience? (user preference)

# Analysis Methodology

Compare user's behaviors and preferences when interacting with **low-quality** and **high-quality** news.

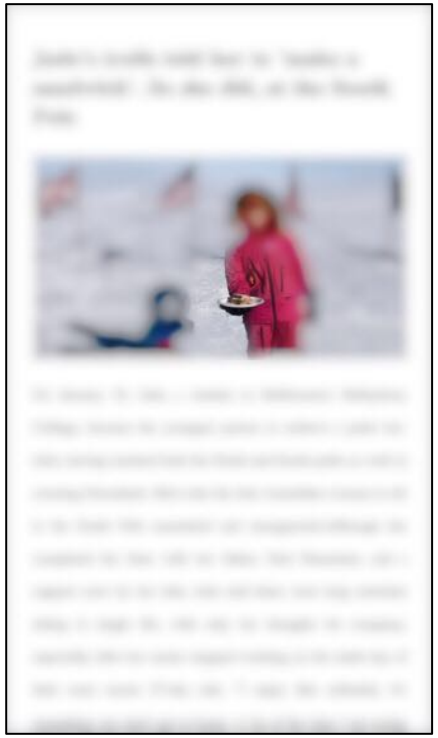



## We Need:

- ✦ **Control related aspects**
  - ✧ position, topic
- ✦ **Collect user various behaviors**
  - ✧ pre-click, post-click
- ✦ **Collect user experience**
  - ✧ preference, perceived quality

# News quality annotation

News from: *social, entertainment, technology, history, sports.*

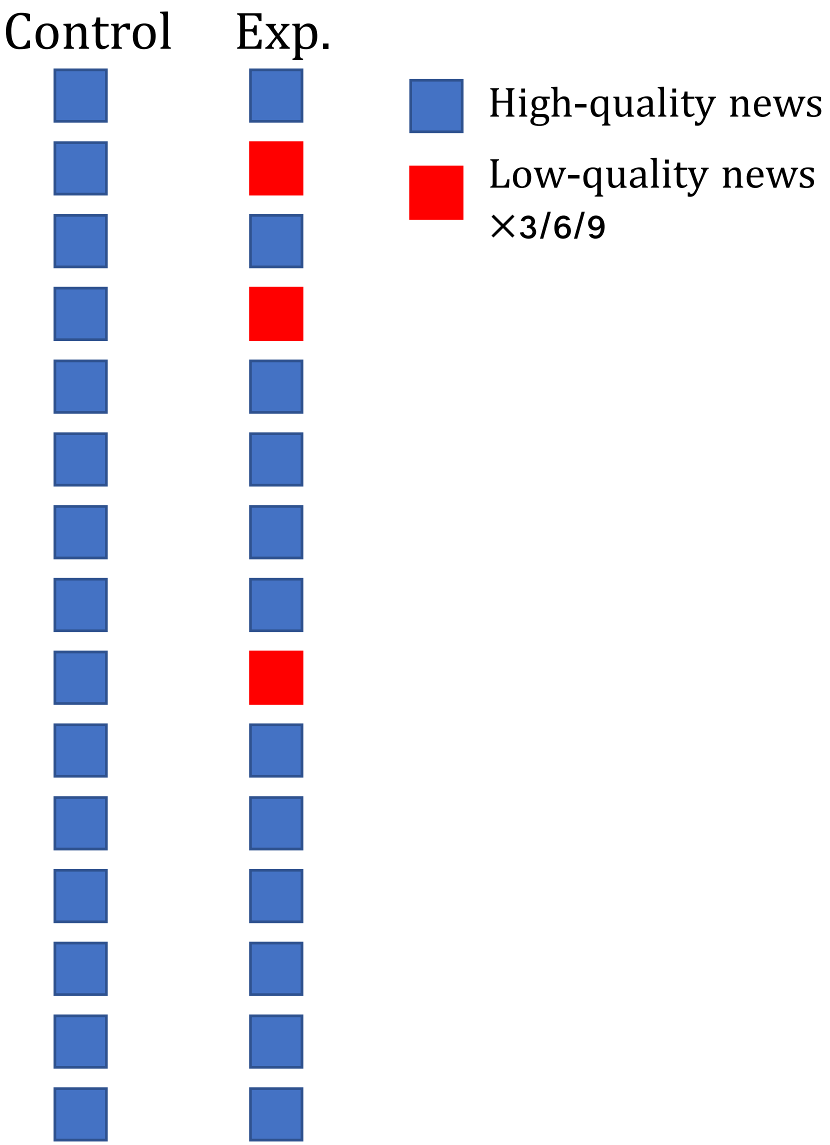


- ☐ Authenticity
- ☐ Value
- ☐ Expression
- ☐ Headline

**Expert Labelled Quality**

- ☐ Low-quality
- ☐ High-quality

# Conducting the experiment lists



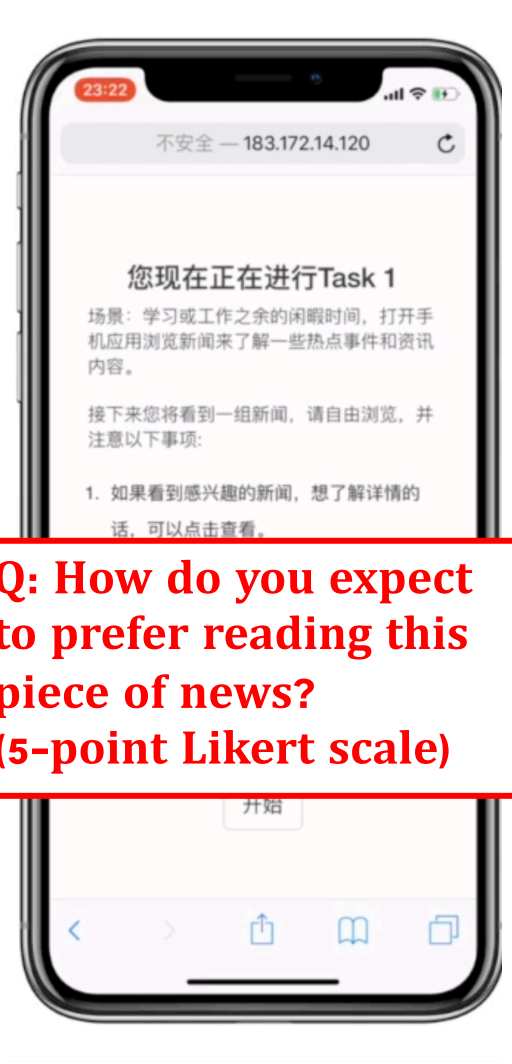

Randomly and with Latin Square Principles, we make sure:

Low-quality news vs. High-quality news:

1. **Same position distribution**
2. **Same topic distribution**

# Experiment Procedure

*Collecting user preferences in different phases*




**Q: How do you expect to prefer reading this piece of news?**
**(5-point Likert scale)**

Task Begin

List Browsing

Click

Before-Read Questionnaires

**Before-Read Preference**

# Experiment Procedure

*Collecting user preferences in different phases*


**Q: How do you like reading this piece of news? (5-point Likert scale)**

**What do you think of the content quality of this piece of news?**

**What do you think of the consistency between title and content of the news?**

Task Begin

List Browsing

Click

Before-Read Questionnaires

Read

End Read

After-Read Questionnaires

**Before-Read Preference**

**After-Read Preference**

**Perceived Content and Title Quality**

# Experiment Procedure

*Collecting user preferences in different phases*


**Q: How do you like reading this piece of news? (5-point Likert scale)**

Task Begin

List Browsing

Click

Before-Read Questionnaires

Read

End Read

After-Read Questionnaires

End Browsing

Post-Task Questionnaires

**Before-Read Preference**

**After-Read Preference**

**Perceived Content and Title Quality**

**Post-Task Preference**

# User Study Dataset

*15* News per Task

*4* Tasks per user

*32* Participants

*128* Tasks

*1,920* Impressions (*576* low-quality)

*631* Clicks (*209* low-quality)

# Research Questions

Focus on Three Concepts…

**Quality**

Content / title quality

- Expert labelled
- User perceived

**Preference**

- Multiple phases
- Topic interests

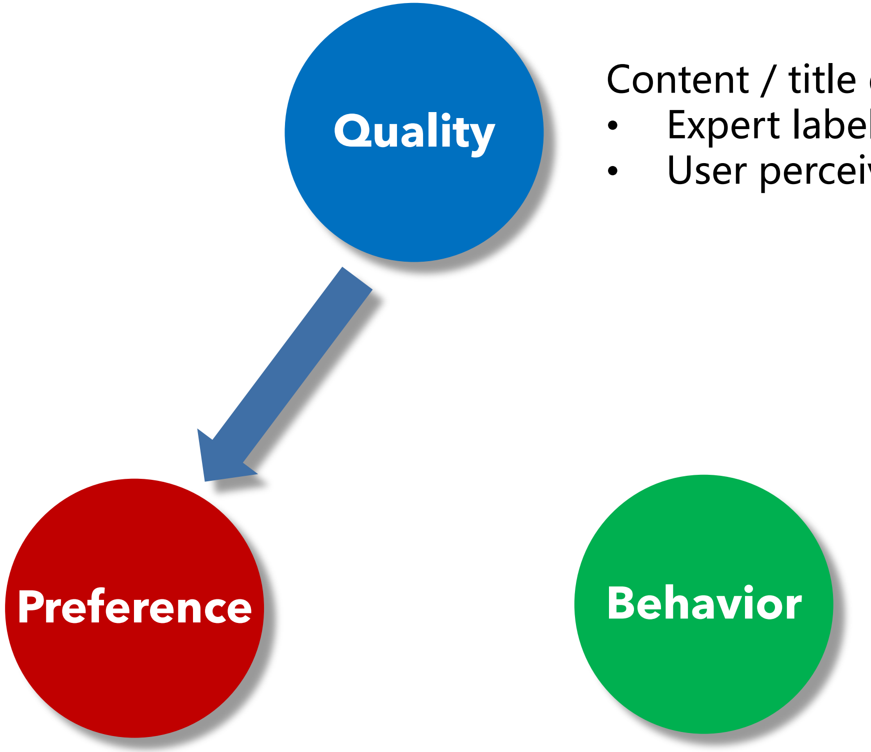
**Behavior**

- Click behavior
- Reading behavior

# Research Questions

**RQ1** Does quality affect user preferences? If yes, how?

Content / title quality
- Expert labelled
- User perceived

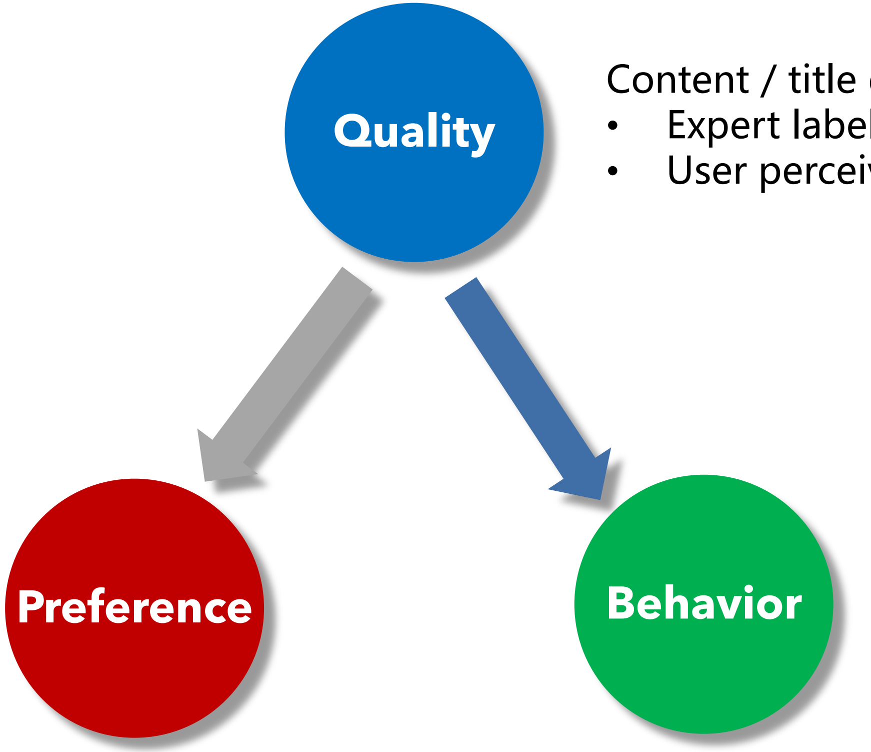



- Multiple phases
- Topic interests

- Click behavior
- Reading behavior

# Research Questions

**RQ2** Does quality affect user behaviors during the browsing and reading process? If yes, how?



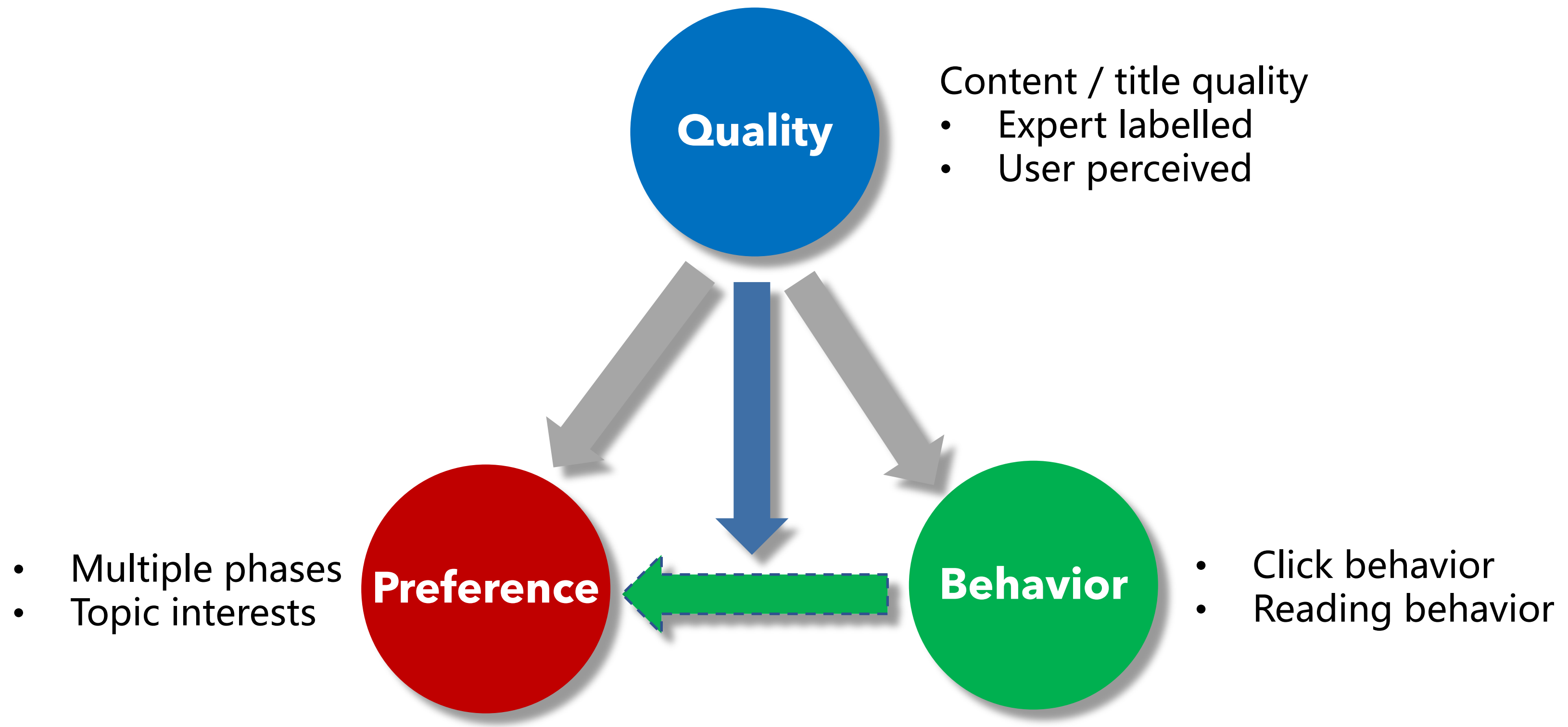

Content / title quality
- Expert labelled
- User perceived

- Multiple phases
- Topic interests

- Click behavior
- Reading behavior

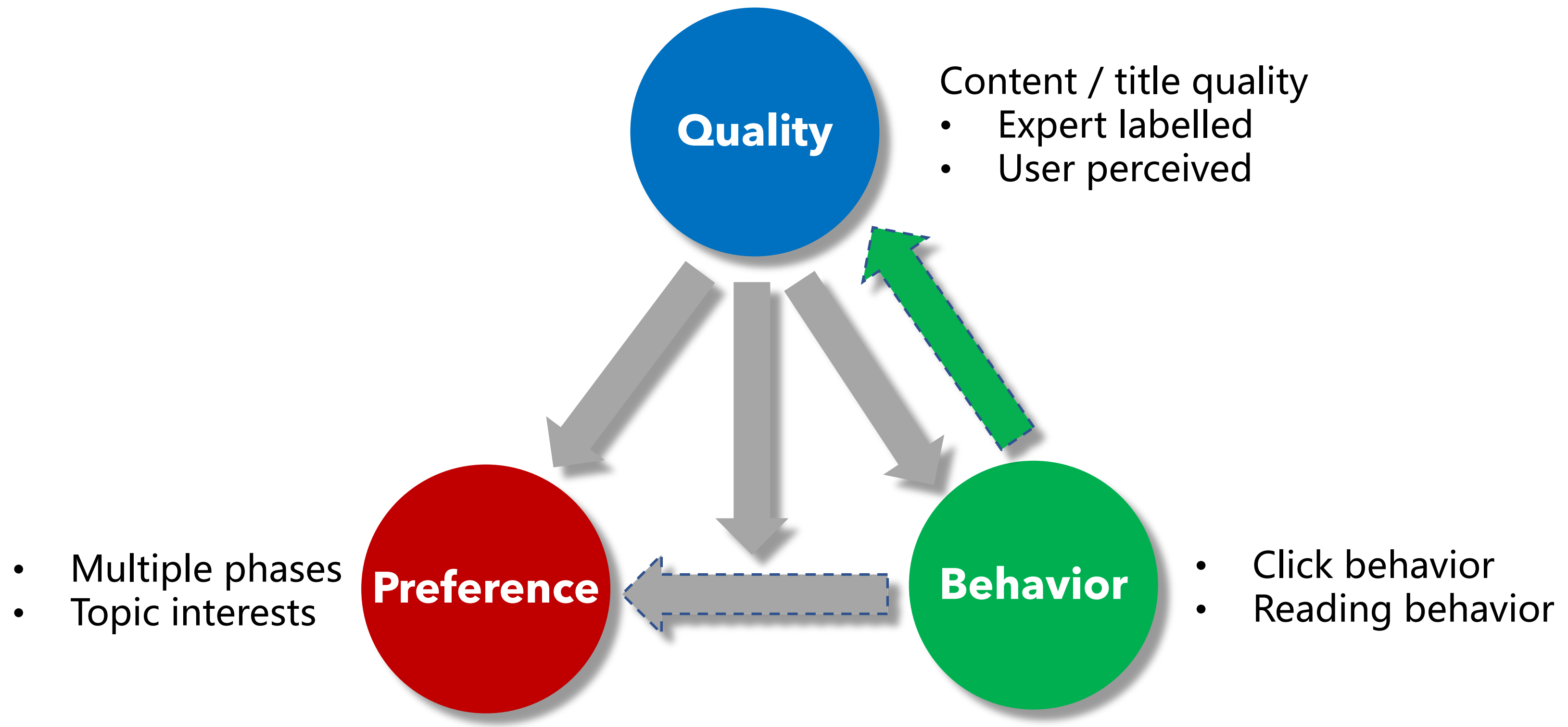
# Research Questions

**RQ3** Can incorporating quality help build implicit feedback?



# Research Questions

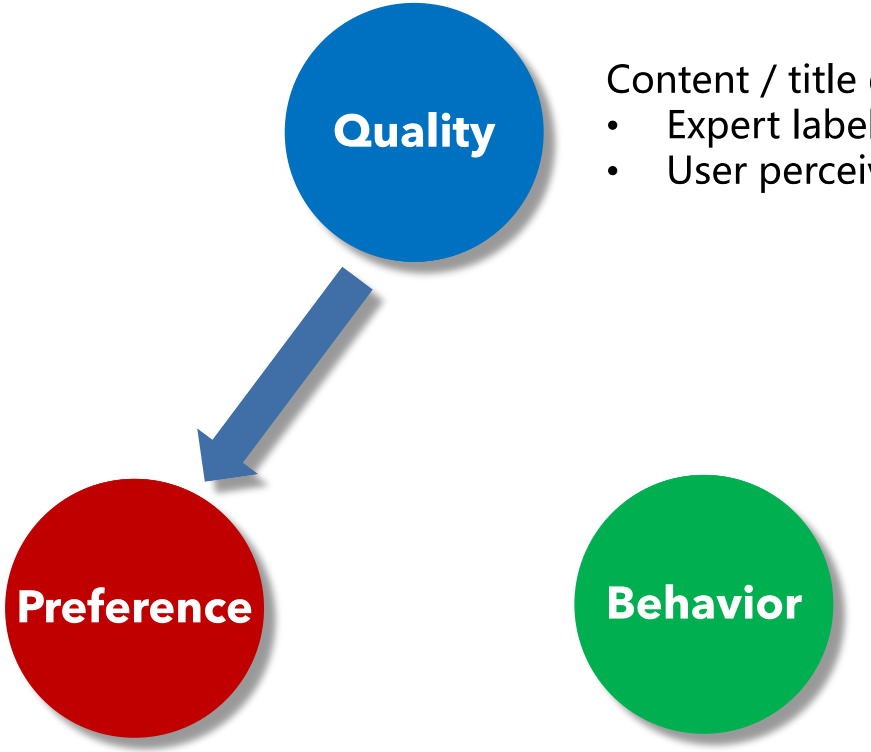
**RQ4** Can we identify quality based on user behavior signals?



# Research Questions

**RQ1** Does quality affect user preferences? If yes, how?

Content / title quality
- Expert labelled
- User perceived




- Multiple phases
- Topic interests

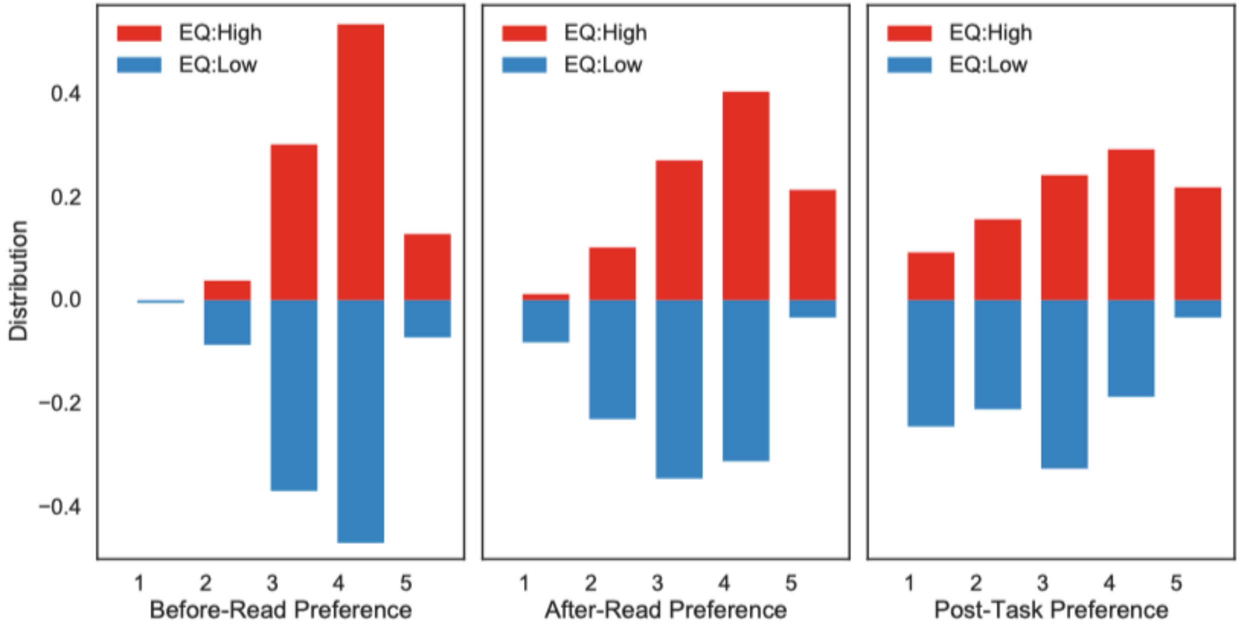
- Click behavior
- Reading behavior

# RQ1: Quality & Preference

- When news quality is low, how is the distribution of user preference in three phases?

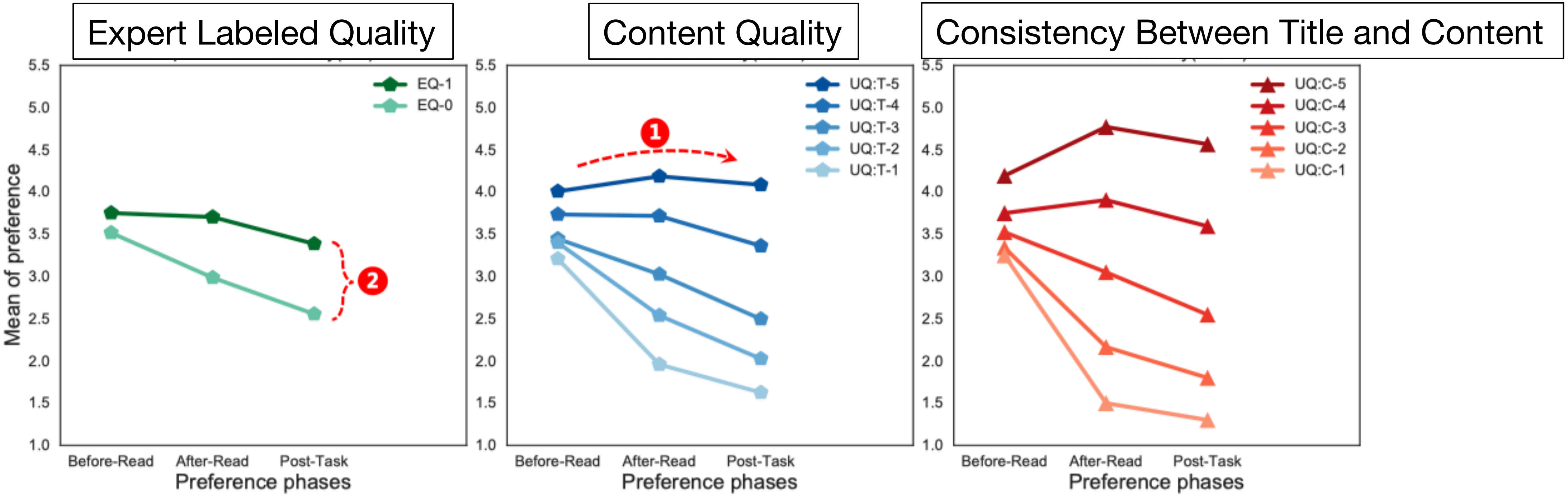
P<0.01, d=0.32 P<0.01, d=0.74 P<0.01, d=0.69

Degrees of the differences vary in three phases




Finding#1: User preference varies when news quality is different.

# RQ1.1 Quality vs. Preference: In different interaction phases




Finding#1: User preference for low-quality news continually drops

Finding#2: quality effect: before-read phase < after-read & post-task phase

Finding#3: quality effect: of user perceived qualities > of expert labeled qualities

# RQ1.2 Quality vs. Preference: with Different topic interest

**Table 2: Quality effects, measured by the difference (Δ, Cohen's *d*) of user preferences between low-quality and high-quality news, when user has different topic interests (TI).**

| | Before-Read Preference | | | | After-Read Preference | | | | Post-Task Preference | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | EQ:Low | EQ:High | Δ | *d* | EQ:Low | EQ:High | Δ | *d* | EQ:Low | EQ:High | Δ | *d* |
| TI=Min | 3.610 | 3.689 | +0.079 | 0.114 | 3.170 | 3.597 | +0.427 | 0.447 | 2.627 | 3.176 | +0.549 | 0.454 |
| TI=Mid | 3.465 | 3.630 | +0.166 | 0.223 | 2.831 | 3.674 | **+0.843** | **0.870** | 2.563 | 3.442 | +0.879 | 0.751 |
| TI=Max | 3.494 | 3.897 | **+0.403** | **0.532** | 2.987 | 3.806 | +0.819 | 0.838 | 2.494 | 3.491 | **+0.997** | **0.801** |

Finding#1: When user has higher topic interest, the quality effect is larger.

Finding#2: if the quality is low, lower topic interest leads to higher preference

(User has high quality sensitiveness (low tolerance) for the news of his/her interested topics.)

# RQ1.2 Quality vs. Preference: with Different topic interest

Table 2: Quality effects, measured by the difference (Δ, Cohen's *d*) of user preferences between low-quality and high-quality news, when user has different topic interests (TI).

| | Before-Read Preference | | | | After-Read Preference | | | | Post-Task Preference | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | EQ:Low | EQ:High | Δ | *d* | EQ:Low | EQ:High | Δ | *d* | EQ:Low | EQ:High | Δ | *d* |
| TI=Min | 3.610 | 3.689 | +0.079 | 0.114 | 3.170 | 3.597 | +0.427 | 0.447 | 2.627 | 3.176 | +0.549 | 0.454 |
| TI=Mid | 3.465 | 3.630 | +0.166 | 0.223 | 2.831 | 3.674 | **+0.843** | **0.870** | 2.563 | 3.442 | +0.879 | 0.751 |
| TI=Max | 3.494 | 3.897 | **+0.403** | **0.532** | 2.987 | 3.806 | +0.819 | 0.838 | [illegible] | [illegible] | **+0.997** | **0.801** |

Finding#1: When user has higher topic interest, the quality effect is larger.

Finding#2: if the quality is low, lower topic interest leads to higher preference

(User has high quality sensitiveness (low tolerance) for the news of his/her interested topics.)

## Question

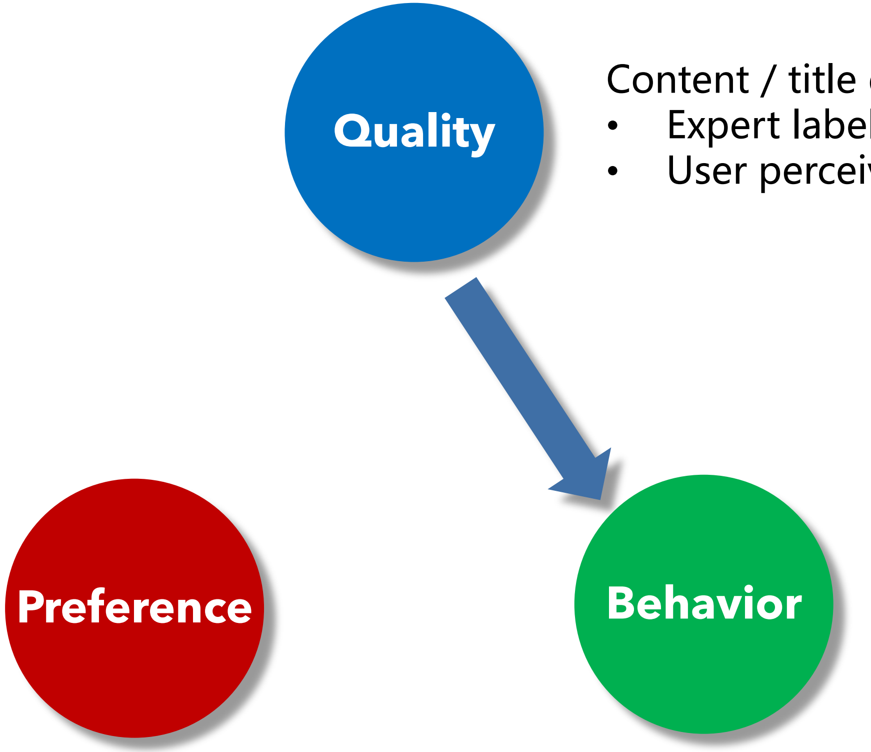
- Does quality affect user preferences? If yes, how?

## Observation

- Yes, lower quality leads to lower preference.
  - Especially in two phases of after reading
  - Especially when user has higher topic interest

# Research Questions

**RQ2** Does quality affect user behaviors during the browsing and reading process? If yes, how?




Content / title quality
- Expert labelled
- User perceived

- Multiple phases
- Topic interests

- Click behavior
- Reading behavior

# RQ2.1 Quality vs. Click behaviors

① Conditional probability

- P(click|EQ=1) = 0.3140
- P(click|EQ=0) = 0.3628

② Add position (top-k)

③ Large scale log analysis

(sampled from multiple days' log data, 1.5K impressions per news on avg.)

- High-quality news CTR (0.0835)
- Low-quality news CTR (0.1539)

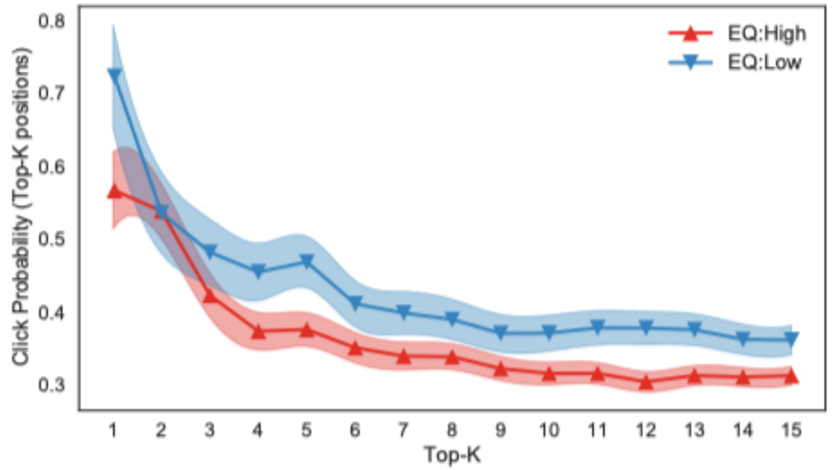



**Figure 5: Click Probability of the news *up to position k* conditioned by the news quality. The low-quality news attracts more clicks.**

Finding: Low-quality news has higher click probability

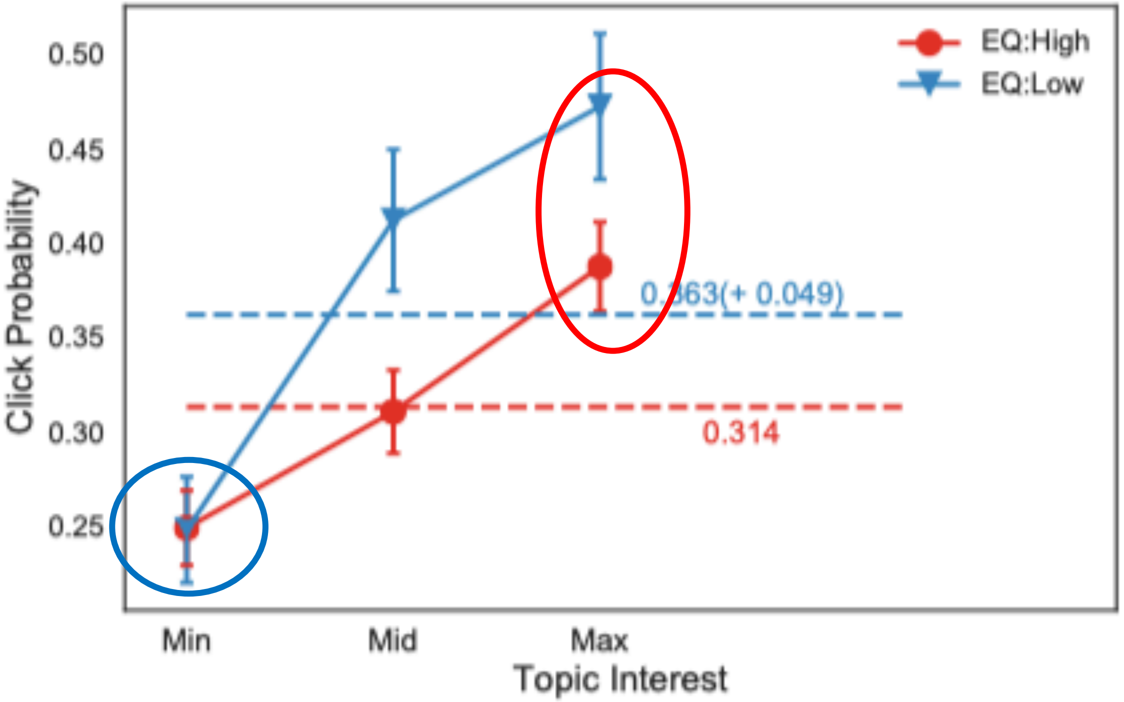
# Why low quality news receive more clicks?

Supplementary annotation for the persuasion of the title.

- Title persuasion: The extent that user is seduced to click the news (4-scales)
- 3 different assessors per news (Fleiss' k=0.4259)

Finding: Generally low-quality news has higher persuasion than high-quality news. (2.16 vs. 1.61)

# Topic Interests, Quality vs. Click




Finding: When topic interest is high, the difference of click probability is big

When topic interest is low, the difference of click probability is small

# Contextual effect of Quality vs. Click

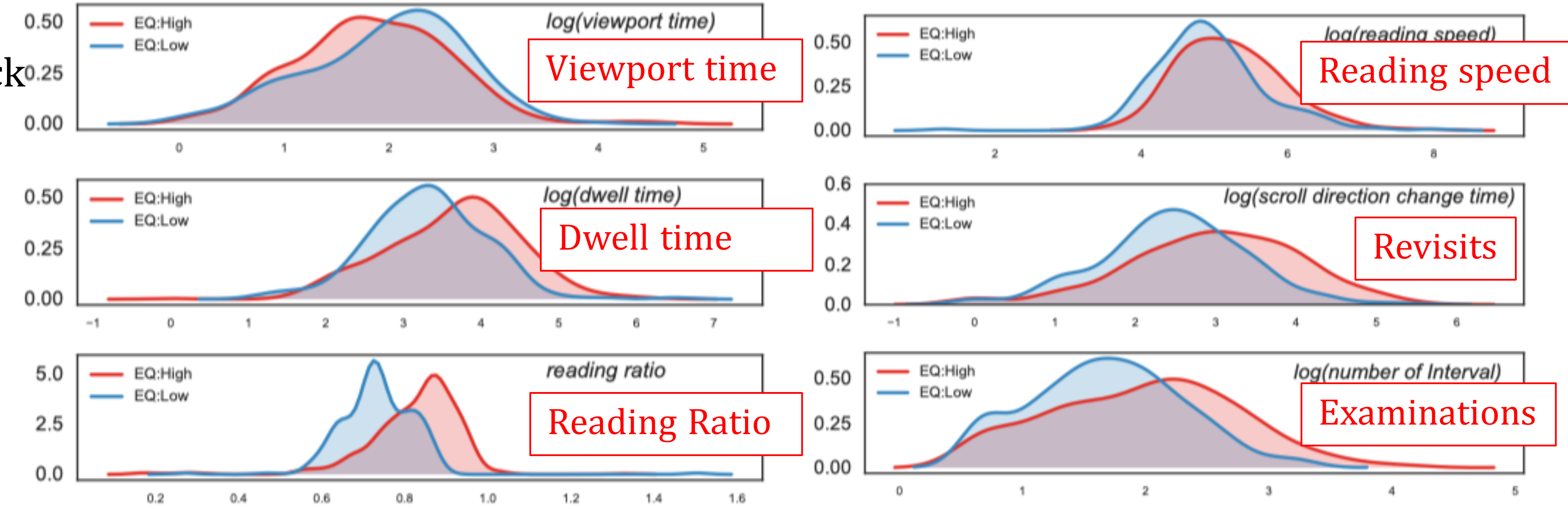
- Whether the quality of last displayed news ($lEQ$) affects the click probability of current news ($cEQ$) ?

| | $lEQ = low$ | | $lEQ = high$ | |
|---|---|---|---|---|
| $P(Click\|\ lEQ)$ | 0.3507 | | 0.2898 | |
| | $cEQ = 0$ | $cEQ = 1$ | $cEQ = 0$ | $cEQ = 1$ |
| $P(Click\|\ lEQ, cEQ)$ | 0.4000 | 0.3108 | 0.2838 | 0.2917 |

Finding: If the quality of last news is low, user will have higher probability to click current news.

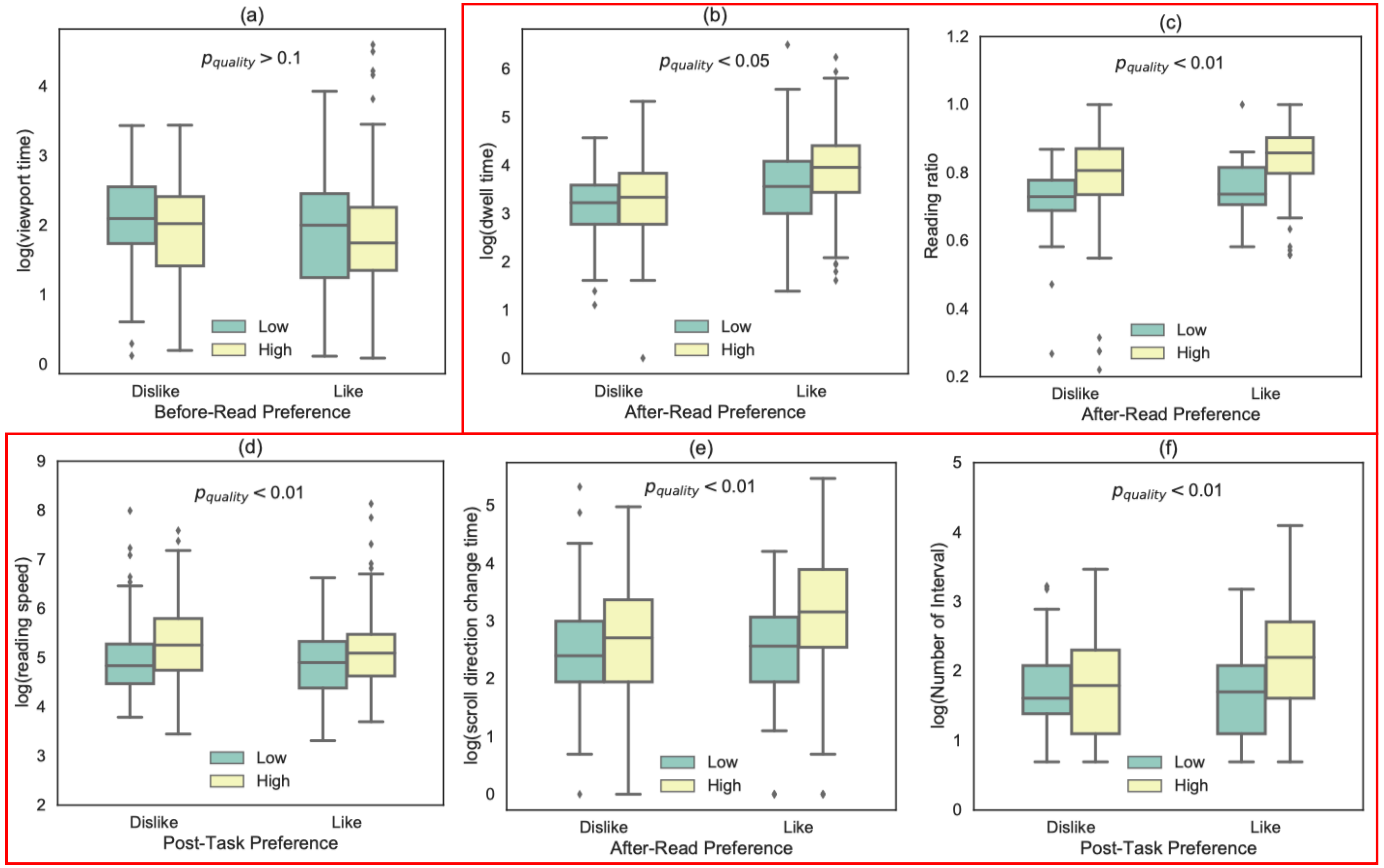
# RQ2.2 Quality vs. Reading behaviors

- When users read low-quality news, they will:
  - Spend more time before click
  - Spend less time reading
  - Leave earlier
  - Read slower
  - Have fewer revisits
  - Have fewer careful examinations



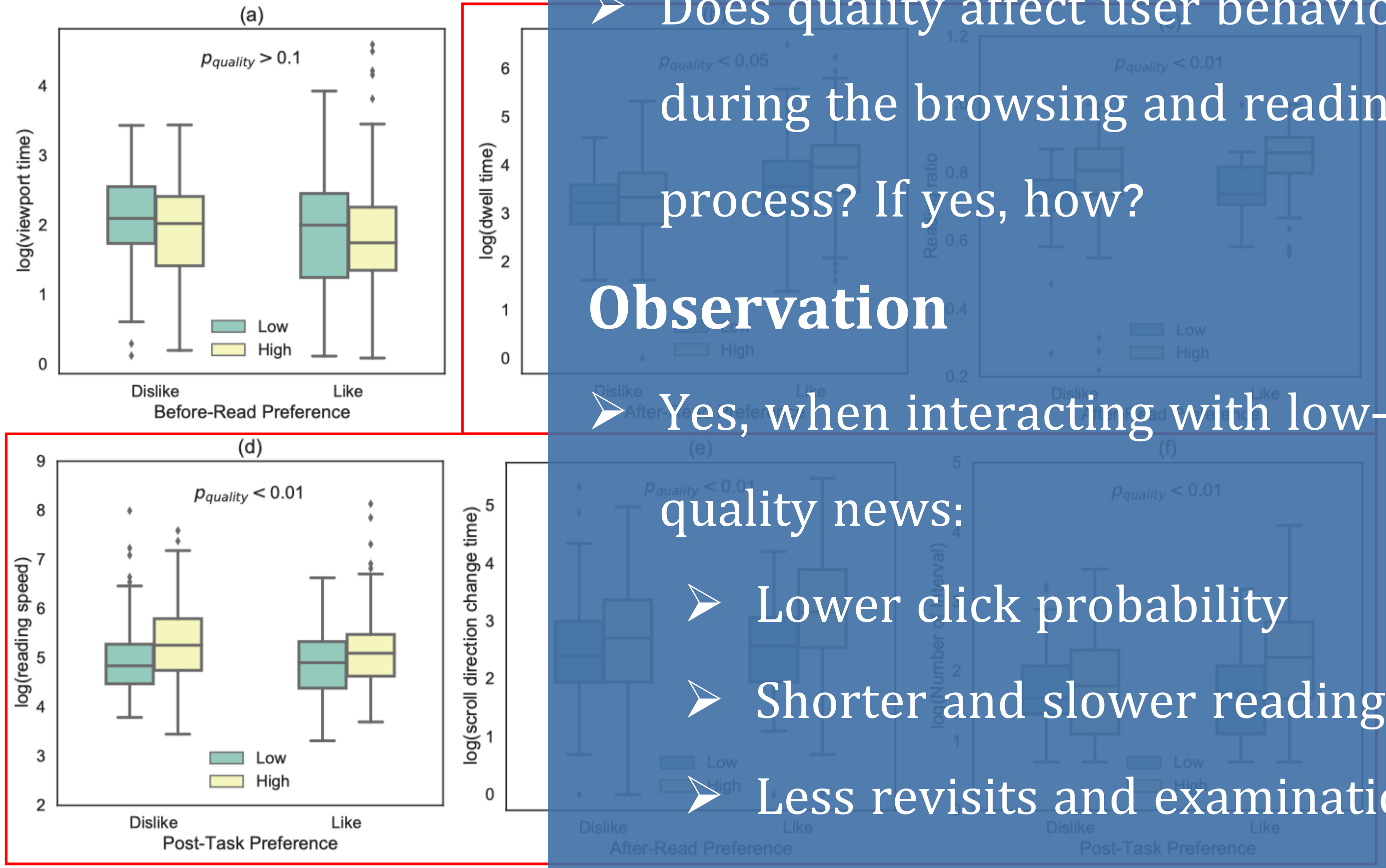
# Control user preference to study Quality vs. Reading behaviors

- Quality has significant effects on most of user behaviors, which is independent to the preference effect on behaviors
  - Dwell time
  - Reading ratio
  - Reading speed
  - Revisits
  - Examinations



# Control user preference to study Quality vs. Reading behaviors

- Quality has significant effects on most of user behaviors, which is independent to the preference effect on behaviors
  - Dwell time
  - Reading ratio
  - Reading speed
  - Revisits
  - Examinations




## Question

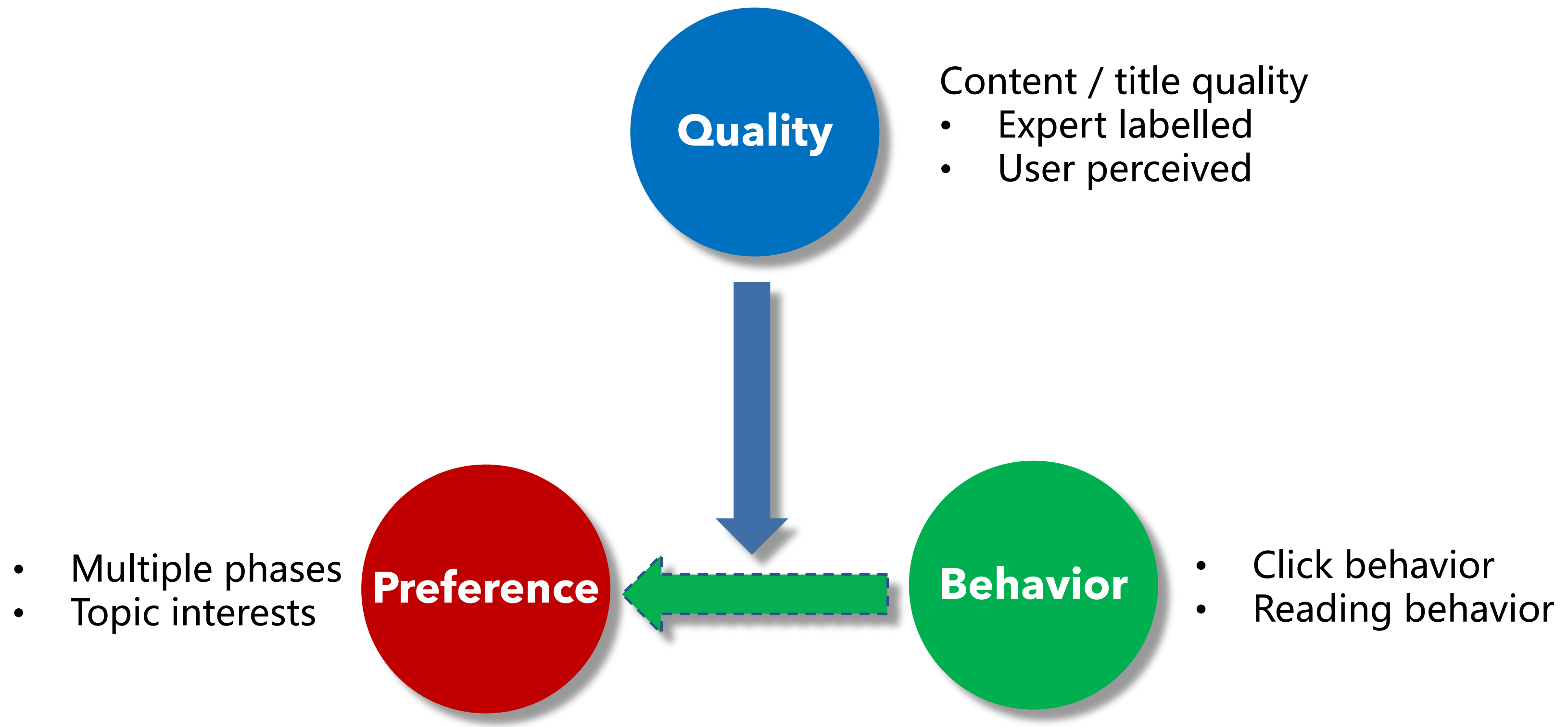
- ➢ Does quality affect user behaviors during the browsing and reading process? If yes, how?

## Observation

- ➢ Yes, when interacting with low-quality news:
  - ➢ Lower click probability
  - ➢ Shorter and slower reading
  - ➢ Less revisits and examinations.

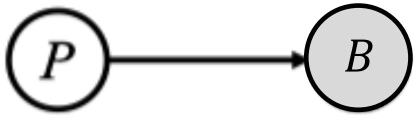
# Research Questions

**RQ3** Can incorporating quality help build implicit feedback?



# RQ3. Preference-Behavior-Quality model

Traditional implicit feedback
e.g. click, sat-click




(a) PB-model

$$P(P=1|B) = \frac{P(B|P=1)P(P=1)}{\sum_{i\in\{0,1\}} P(B|P=i)P(P=i)}$$

# Results

- Estimating *whether a user likes a clicked news*. (<=3: dislike; >3: like)
- Ground truth: *Post-Task Preference*
- Evaluation metric: *AUC*

| Behavior metric | AUC(PB) | AUC(PBQ) | *p* | cohens' *d* |
|---|---|---|---|---|
| viewport time | 0.5775 | **0.6249** | ** | 1.25 |
| dwell time | 0.6225[1] | **0.6526** | ** | 0.88 |
| reading ratio | 0.6382 | **0.6486** | | 0.23 |
| reading speed | 0.4490 | **0.6142** | ** | 3.32 |
| direction change times | 0.5904 | **0.6477** | ** | 1.17 |
| number of interval | 0.6111 | **0.6709** | ** | 1.33 |

[1] Sat-click, the widely used implicit feedback, can be interpreted as dwell time-based PB-model.

**PBQ-model** outperforms the PB model when using all the behavior signals

# Results

- Estimating *whether a user likes a clicked news*. (<=3: dislike; >3: like)
- Ground truth: *Post-Task Preference*
- Evaluation metric: *AUC*

| Behavior metric | AUC(PB) | AUC(PBQ) | $p$ | cohens' $d$ |
|---|---|---|---|---|
| viewport time | 0.5775 | **0.6249** | ** | 1.25 |
| dwell time | 0.6225[1] | **0.6526** | ** | 0.88 |
| reading ratio | 0.6382 | **0.6486** | | 0.23 |
| reading speed | 0.4490 | **0.6142** | ** | 3.32 |
| direction change times | 0.5904 | **0.6477** | ** | 1.17 |
| number of interval | 0.6111 | **0.6709** | ** | 1.33 |

[1] Sat-click, the widely used implicit feedback, can be interpreted as dwell time-based PB-model.

## Question

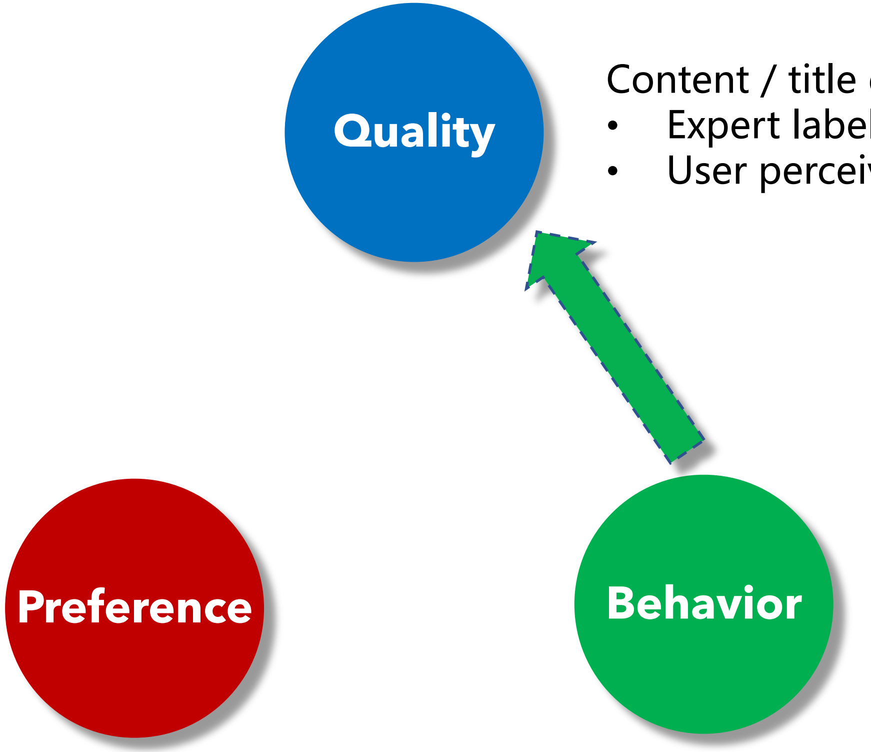
➢ Can incorporating quality help build implicit feedback?

## Observation

➢ Yes, significantly.

# Research Questions

**RQ4** Can we identify quality based on user behavior signals?




Content / title quality
- Expert labelled
- User perceived

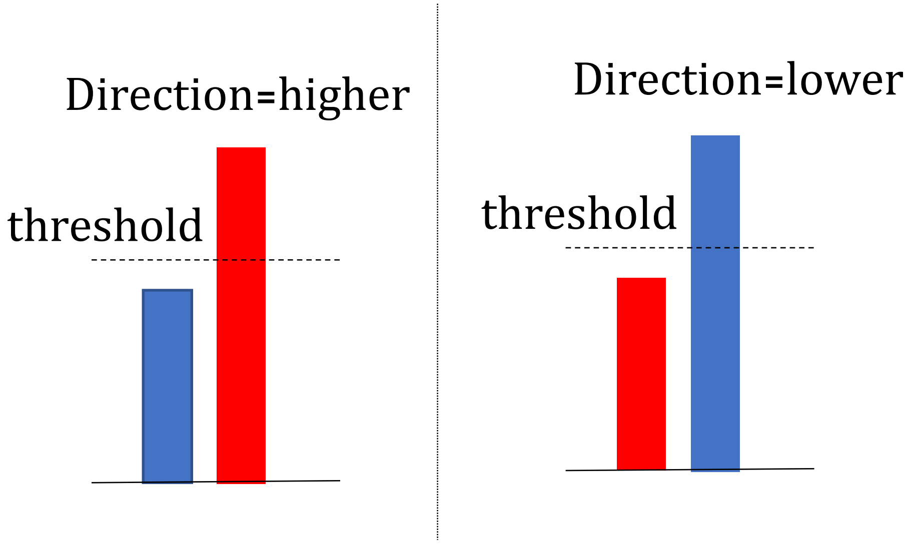
- Multiple phases
- Topic interests

- Click behavior
- Reading behavior

# RQ4. Can we identify the news quality based on user behavior?

- Point-wise distinguishing Ability

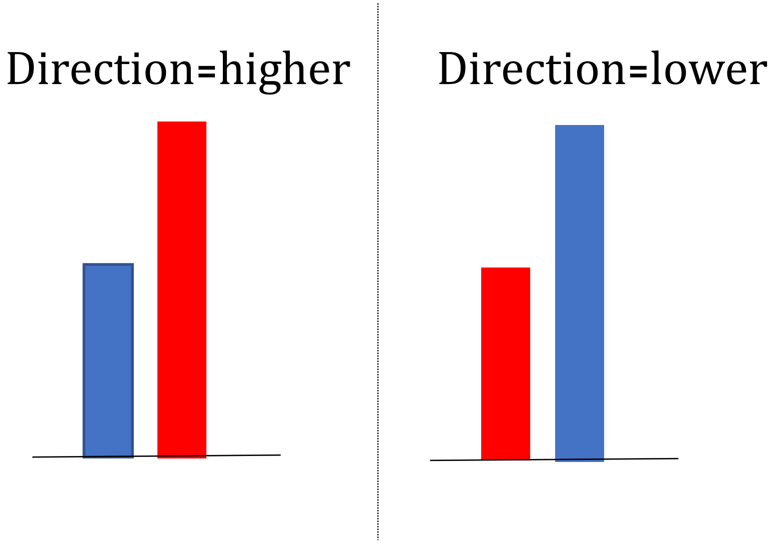
$D_{\text{point}}(b)$




$$D_{point}(b) = \max_{\alpha, t_b} \frac{\sum_{i \in C} I[\hat{q}_{\alpha, t_b}(i) = q_i]}{n(C)}$$

- Pair-wise distinguishing Ability

$D_{pair}(b)$




$$D_{pair}(b) = \max_{\alpha \in \{-1, 1\}} \frac{\sum_{<n_i, n_j> \in S} r_\alpha(n_i, n_j)}{n(S)}$$

# Results

Distinguishing Ability for both
Expert Labelled Quality and User Perceived Quality

Reading ratio has the highest ability to distinguish expert labelled quality with threshold $t_b$ = 0.74

Whether user read more than 74% of the news content can be used as an indicator for the high quality news.

| | Expert Quality | | | |
|---|---|---|---|---|
| | $D_{point}$ | $\alpha$ | $D_{pair}$ | $\alpha$ |
| viewport time | 0.6703 | - | 0.5850 | - |
| dwell time | 0.6751 | + | 0.6650 | + |
| reading ratio | **0.7084** | + | **0.8010** | + |
| reading speed | 0.6799 | + | 0.6210 | + |
| direction change times | 0.6688 | + | 0.6590 | + |
| number of interval | 0.6719 | + | 0.5174 | + |

+ Positive relative relation. - negative relative relati

# Results

Distinguishing Ability for both
Expert Labelled Quality and User Perceived Quality

Reading ratio has the highest ability to distinguish expert labelled quality with threshold $t_b$ = 0.74

Whether user read more than 74% of the news content can be used as an indicator for the high quality news.

| | Expert Quality | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $D_{point}$ | $\alpha$ | $D_{pair}$ | $\alpha$ | $D_{pair}$ | $\alpha$ | $D_{pair}$ | $\alpha$ |
| viewport time | 0.6703 | - | 0.5850 | - | [illegible] | [illegible] | [illegible] | [illegible] |
| dwell time | 0.6751 | + | 0.6650 | + | 0.6940 | + | 0.6350 | + |
| reading ratio | **0.7084** | + | **0.8010** | + | [illegible] | + | [illegible] | + |
| reading speed | 0.6799 | + | 0.6210 | + | 0.5240 | - | 0.5300 | - |
| direction change times | 0.6688 | + | 0.6590 | + | [illegible] | + | [illegible] | + |
| number of interval | 0.6719 | + | 0.5174 | + | 0.5547 | + | 0.5106 | + |

+ Positive relative relation. - negative relative relation.

Dwell time and Reading ratio has the highest ability to distinguish the User Perceived Quality

## Question

- Can we identify quality based on user behavior signals?

## Observation

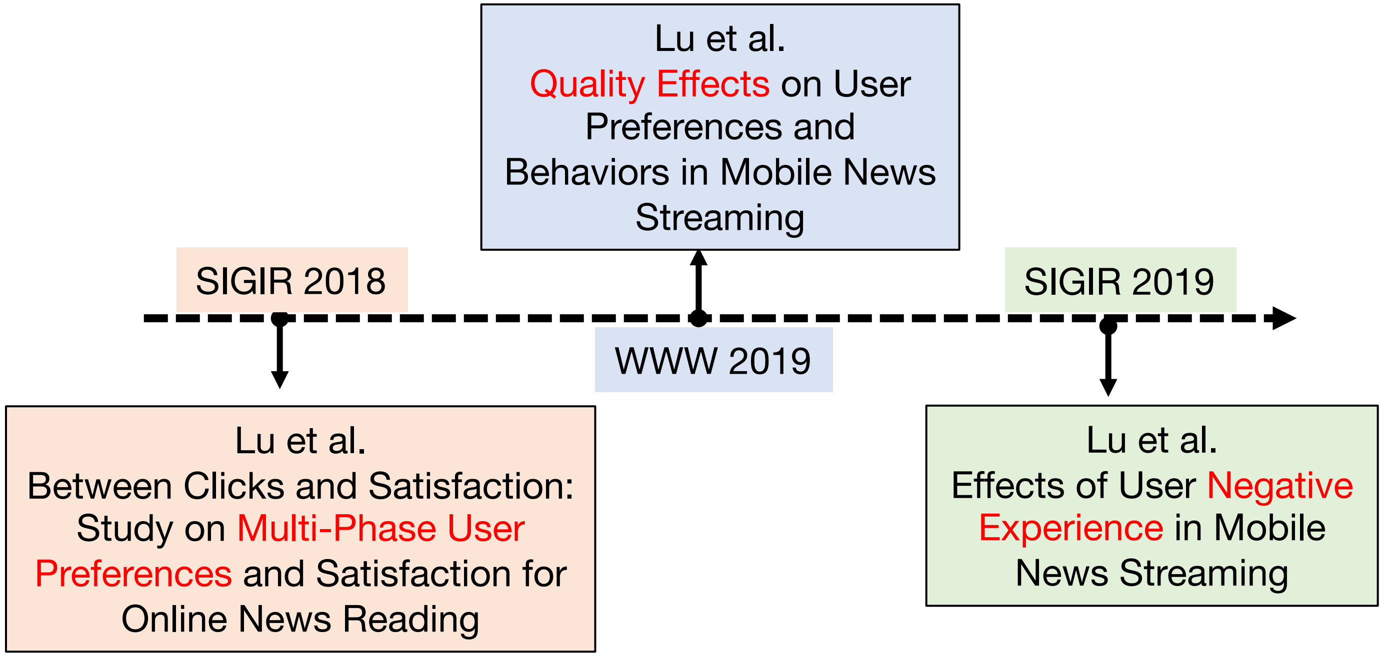
- Yes, especially using *reading ratio* and *dwell time*.

# Takeaways


- ✦ Low quality leads to low preference
  - ✦ related with interaction phases and topic interest
- ✦ Quality significantly affects user behaviors
  - ✦ Low-quality news attracts more clicks especially when the user has higher interest in news topic.
  - ✦ Read less and slowly, with fewer revisits and fewer examinations.
- ✦ Quality helps building implicit feedback (PBQ-model)
- ✦ User behaviors, especially reading ratio and dwell time, can be used to identify quality (Future: multiply behaviors & content)

# Thanks

**THANKS FOR YOUR ATTENTION, ANY QUESTIONS?**
**My email: luhy16@mails.tsinghua.edu.cn**




1. User Behaviors:
   - Stop
   - Return
   - Gaze
2. User Experience:
   - Intent
   - Context
3. Application:
   - Multi-behavior
   - Evaluation

**Papers and Data can be found: luhongyu.github.io & www.thuir.cn/group/~mzhang/**