Allocation Schemes in Analytic Evaluation: Applicant-Centric Holistic or Attribute-Centric Segmented?

Jingyan Wang, Carmel Baharav, Nihar B. Shah, Anita Williams Woolley, R. Ravi
## Allocation schemes

| Q1 | Q2 | Q3 | ...... |
|-----|-----|-----|--------|
| Alice | | | |
| Bob | | | |
| Carol | | | |
| ...... | | | |

| Q1 | Q2 | Q3 | ...... |
|-----|-----|-----|--------|
| Alice | | | |
| Bob | | | |
| Carol | | | |
| ...... | | | |
Allocation schemes

Allocation schemes in more complicated applications?

- **Large-scale**: distribute task to many reviewers
- **Separable**: evaluate individual attributes

Admissions

Hiring
Scope: **analytic** evaluation

**Analytic:** evaluate a pre-defined set of attributes

**Non-analytic:** not require to evaluate individual attributes

- overly rely on people’s general impression
- inconsistency and inaccuracy compared to analytic  
  
  [Jönsson, Balan, and Hartell 2021]

| Score | Grades | Essays | Letters | ...... |
|-------|--------|--------|---------|--------|
| App 1 | App 1  |        |         |        |
| App 2 | App 2  |        |         |        |
| App 3 | App 3  |        |         |        |
| App 4 | App 4  |        |         |        |

Non-analytic  Analytic
Scope: **analytic** evaluation with **exogenous** aggregation

**Exogenous aggregation**: predefined rules or algorithms

**Human aggregation**: evaluator combines and weights attributes
  - No more (or even less) accurate than simple rules  [Kahneman, Sibony, and Sunstein 2021]
Scope: **analytic** evaluation with **exogenous** aggregation

“People trust that the complex characteristics of applicants can be best assessed by a sensitive, equally complex human being. This does not stand up to scientific scrutiny”  
[Highhouse 2008]
Holistic vs segmented

- **Holistic**: assign a subset of applications to each reviewer
- **Segmented**: assign a subset of attributes to each reviewer
Comparison

- **Holistic**: assign a subset of applications to each reviewer
- **Segmented**: assign a subset of attributes to each reviewer

|    | 1     | 2     | 3     |
|----|-------|-------|-------|
| Desiderata | Calibration | Efficiency | Fairness |
Comparison

- **Holistic**: assign a subset of applications to each reviewer
- **Segmented**: assign a subset of attributes to each reviewer

| Desiderata | Method   | 1   | 2   | 3   |
|------------|----------|-----|-----|-----|
| Calibration| Experiment| Efficiency | Simulation | Theory |
1. Calibration

• Accuracy of estimating percentile of each applicant with respect to the entire pool of applicants

• 1 attribute

• Give workers a set of numbers between 0-300

• Ask workers to estimate percentile using 5 bins

Number: 244

Reminder: answering this question accurately will give you a bonus.

| 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---------------|--------|--------|--------|----------------|
| ○             | ○      | ○      | ○      | ○              |

Group 1: Holistic
See 5 numbers

Group 2: Segmented
See 20 numbers
(5 numbers per page)
Experimental Results

**Observation 1:** 20Q-group workers have lower error for later pages (p<0.01).
- **Page 1:** 0.95 (± 0.06)
- **Page 4:** 0.74 (± 0.06)

**Observation 2:** 20Q-group has lower error than 5Q-group (p<0.01).
- **20Q-group:** 0.84 (± 0.05)
- **5Q-group:** 1.14 (± 0.06)

**Conclusion:** Reviewers in segmented evaluation have better calibration, due to seeing more applicants.
2. Efficiency

• Adaptively allocate efforts to evaluate attributes
• 2 attributes with correlation $\sigma$
• Holistic: Evaluate attribute 2 only if attribute 1 is in top $\tau$-fraction

**Observation 1:** Tradeoff between efficiency and accuracy

**Observation 2:** When correlation $\sigma$ is high, small values of $\tau$ give:
• significant saving in efficiency
• marginal decrease in accuracy

**Conclusion:** Holistic evaluation is more efficient, due to evaluators adaptively allocating efforts.
3. Fairness

- Biased evaluators against certain disadvantaged groups
- Multiplicative discount $\beta \leq 1$ if

biased reviewer + disadvantaged app. + protected attr.

| attr 1 | attr 2 |
|--------|--------|
| App 1  |        |
| App 2  |        |
| App 3  |        |
| App 4  |        |

protected attr.

disadvantaged app.

biased reviewer
Theoretical results

- 2 reviewers (1 biased, 1 unbiased)
- 2 attributes with identical values from PowerLaw(\(\delta\)) [Kleinberg & Raghavan 2018]
- 50% disadvantaged applicants.

**Theorem (informal).**

a) Both attributes are protected. Extreme discount \(\beta = 0\).

Segmented evaluation is better if and only if

\[
\delta < \frac{\log 3}{\log 2} - 1 \approx 0.58.
\]

b) One attribute is protected. Any discount \(\beta\).

Segmented evaluation is always better than holistic evaluation.

**Conclusion:** Which evaluation scheme is more fair depends on specific settings.
Take-aways:

Complexities in using segmented vs. holistic allocation for evaluation tasks.

| Desiderata   | Factor                                         | Which better? |
|--------------|------------------------------------------------|---------------|
| Calibration  | Learning info about population                 | Segmented     |
| Efficiency   | Allocating effort adaptively                   | Holistic      |
| Fairness     | Distributing impact of biased evaluators       | Depends       |
Take-aways:

Complexities in using segmented vs. holistic allocation for evaluation tasks.

| Desiderata   | Factor                                      | Which better? | Other factors?               |
|--------------|---------------------------------------------|---------------|------------------------------|
| Calibration  | Learning info about population              | Segmented     | Ordering effect              |
| Efficiency   | Allocating effort adaptively                | Holistic      | Switching costs              |
| Fairness     | Distributing impact of biased evaluators    | Depends       | Restricting biasing info from reviewers |

Thanks! :)
jingyanw@gatech.edu