COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Statistical Learning Methods for Process Data

Jingchen Liu

Department of Statistics
Columbia University

July 19, 2021

Paper-pencil tests

Trigonometry

1. Solve for θ. $0 \le \theta < 2\pi$

a) $2\cos^2\theta - 1 = 0$

$\cos^2\theta = \frac{1}{2} \Rightarrow \cos\theta = \pm \frac{1}{\sqrt{2}}$.

on $[0, 2\pi)$: $\theta = \pi/4, \frac{3}{4}\pi, \frac{5}{4}\pi, \frac{7}{4}\pi$.

b). $3\tan^2\theta - 1 = 0$.

$\tan^2\theta = \frac{1}{3} \Rightarrow \tan\theta = \pm 1/\sqrt{3}$

on. $[0, 2\pi)$: $\theta = \frac{\pi}{6}, \frac{5}{6}\pi, \frac{7}{6}\pi, \frac{11}{6}\pi$.

Standard tests

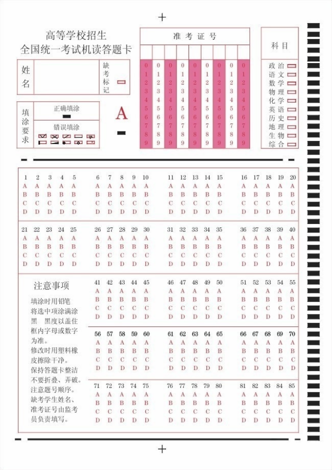
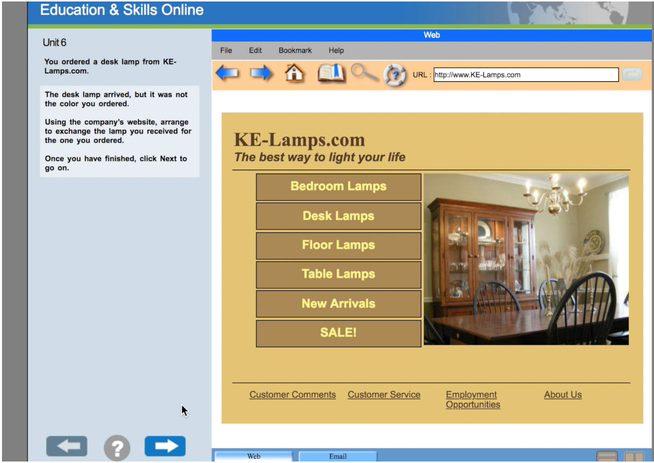


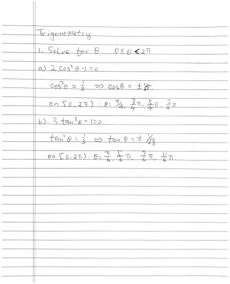
Figure: Exams when I was a student

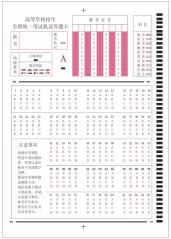
Process Data

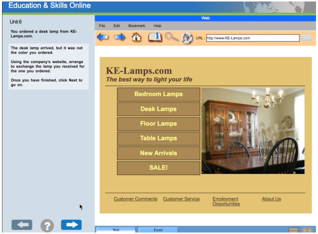


Event	Time
Start	0.0
Click_CS	2.9
Click_ObtNo	12.1
Button_ObtNo	16.3
Auth_No_Close	18.2
Email	21.4
Web	26.8
Back	28.1
Click_RF	33.3
Reason_Wrong	36.3
Combobox1	39.4
On_AuthBox	40.8
ASCII_7	41.7
ASCII_8	42.2
ASCII_3	43.1
ASCII_4	43.8
Off_AuthBox	44.9
Submit	50.6
Submit_Close	53.5
Next	60.4
Next_OK	62.2

Paper-pencil test, standard test, computer-based interactive test







Process response

Process response

Process data

Process Data Research

- New problems:
  - problem-solving strategy analysis
  - cognitive structures
  - ...
- Existing problems:
  - assessment
  - differential item functioning
  - computerized adaptive testing
  - adaptive learning, etc.
  - ...

Process Data Research

- New problems:
  - problem-solving strategy analysis
  - cognitive structures
  - ...
- Existing problems:
  - assessment
  - differential item functioning
  - computerized adaptive testing
  - adaptive learning, etc.
  - ...

Process Data Research

- New problems:
  - problem-solving strategy analysis
  - cognitive structures
  - ...
- Existing problems:
  - assessment
  - differential item functioning
  - computerized adaptive testing
  - adaptive learning, etc.
  - ...

Process Data Research

- New problems:
  - problem-solving strategy analysis
  - cognitive structures
  - ...
- Existing problems:
  - assessment
  - differential item functioning
  - computerized adaptive testing
  - adaptive learning, etc.
  - ...

Process Data Research

- New problems:
  - problem-solving strategy analysis
  - cognitive structures
  - ...
- Existing problems:
  - assessment
  - differential item functioning
  - computerized adaptive testing
  - adaptive learning, etc.
  - ...

Process Data Research

- New problems:
  - problem-solving strategy analysis
  - cognitive structures
  - ...
- Existing problems:
  - assessment
  - differential item functioning
  - computerized adaptive testing
  - adaptive learning, etc.
  - ...

Process Data Research

- New problems:
  - problem-solving strategy analysis
  - cognitive structures
  - ...
- Existing problems:
  - assessment
  - differential item functioning
  - computerized adaptive testing
  - adaptive learning, etc.
  - ...

Process Data Research

- New problems:
  - problem-solving strategy analysis
  - cognitive structures
  - ...
- Existing problems:
  - assessment
  - differential item functioning
  - computerized adaptive testing
  - adaptive learning, etc.
  - ...

Process Data Research

- New problems:
  - problem-solving strategy analysis
  - cognitive structures
  - ...
- Existing problems:
  - assessment
  - differential item functioning
  - computerized adaptive testing
  - adaptive learning, etc.
  - ...

Process Data Research

- New problems:
  - problem-solving strategy analysis
  - cognitive structures
  - ...
- Existing problems:
  - assessment
  - differential item functioning
  - computerized adaptive testing
  - adaptive learning, etc.
  - ...

Four aspects

- Research objective
- Empirical findings
- Technical details: references
  http://www.scientifichpc.com/processdata/pub.html
- Implementation: sessions 3 and 4

Four aspects

- Research objective
- Empirical findings
- Technical details: references http://www.scientifichpc.com/processdata/pub.html
- Implementation: sessions 3 and 4

Four aspects

- Research objective
- Empirical findings
- Technical details: references
  http://www.scientifichpc.com/processdata/pub.html
- Implementation: sessions 3 and 4

Four aspects

- Research objective
- Empirical findings
- Technical details: references
  http://www.scientifichpc.com/processdata/pub.html
- Implementation: sessions 3 and 4

Content of the overview

- Feature extraction: $(\theta_1, ..., \theta_K) \in \mathrm{R}^K$
- Partial scoring
- Removing differential item functioning

Content of the overview

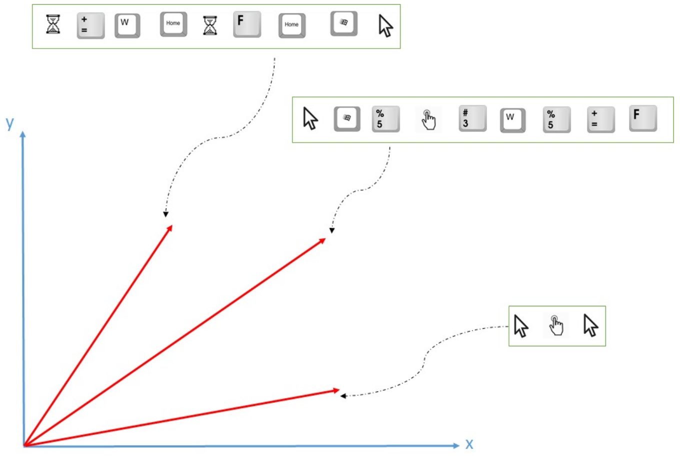
- Feature extraction: $(\theta_1, \ldots, \theta_K) \in \mathbb{R}^K$
- Partial scoring
- Removing differential item functioning

Content of the overview

- Feature extraction: $(\theta_1, \ldots, \theta_K) \in \mathrm{R}^K$
- Partial scoring
- Removing differential item functioning

Feature Extraction

Embedding



Embedding

- Process response:

```
action:Start, Click_cs, Click_ObtNo, ..., Next, Next_OK
time:  0.0  , 2.9     , 12.1       , ..., 60.4, 62.2
```

- Summarize the process to $(\theta_1, ..., \theta_k) \in \mathbb{R}^k$.

A large variety of items

- Email handling/classification, spread sheet handling, scheduling, web browsing/comprehension, etc.
- Learning/interactive with a system, designing experiments, etc.

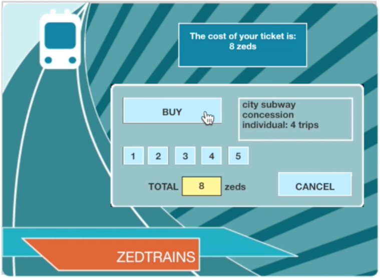
Process Data

TICKETS

A train station has an automated ticketing machine. You use the touch screen on the right to buy a ticket. You must make three choices.

- Choose the train network you want (subway or country).
- Choose the type of fare (full or concession).
- Choose a daily ticket or a ticket for a specified number of trips. Daily tickets give you unlimited travel on the day of purchase. If you buy a ticket with a specified number of trips, you can use the trips on different days.

The BUY button appears when you have made these three choices. There is a CANCEL button that can be used at any time BEFORE you press the BUY button.



Question TICKETS

You plan to take four trips around the city on the subway today. You are a student, so you can use concession fares. Use the ticketing machine to find the cheapest ticket and press BUY.

Once you have pressed BUY, you cannot return to the question.

Very noisy

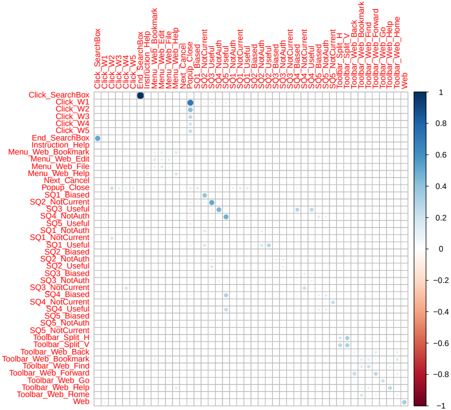


Figure: Lag 1 autocorrelation: $\text{cor}(a_t, a_{t+1})$

Very noisy

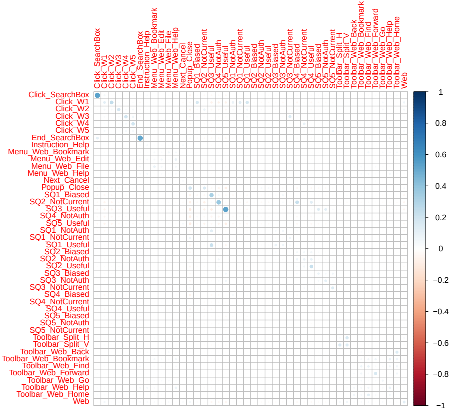


Figure: Lag 2 autocorrelation: $\text{cor}(a_t, a_{t+2})$

Objective

- Denoising: aggregate the process to strengthen the signal.
- Formatting: $\theta \in \mathbb{R}^K$, for $K \geq 100$.
- Dimension reduction.

Latent structure extraction

$a_1 \quad a_2 \quad a_3 \; \ldots\ldots \qquad\qquad a_j \in \mathcal{A} = \{1, \ldots, m\}$

- Process length varies among individuals in the range of [3,1000]
- The number of possible actions m varies among items in the range of $[20, 300]$.
- $(a_1, \ldots, a_{n_i}) \Rightarrow (\theta_1, \ldots, \theta_k)$

Evaluation criteria

- Process features: $(a_1, \ldots, a_{n_i}) \Rightarrow \theta = (\theta_1, \ldots, \theta_k)$
- Benchmark: $(a_1, \ldots, a_{n_i}) \Rightarrow r \in \{\checkmark, \times\}$: task accomplishment
- Does θ contain more information than r? How much information do we loose?

Assessing latent variable

- y: a different variable, such as literacy score.
  - $\hat{y}_\theta$: prediction based on θ

    versus
  - $\hat{y}_r$: prediction based on r

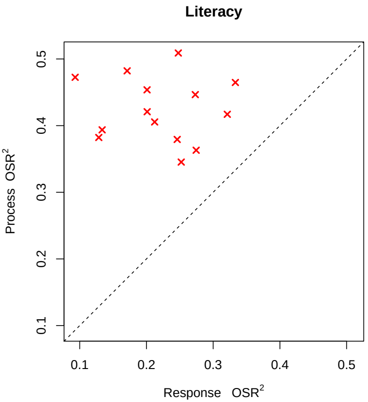


Figure: $cor^2(y, \hat{y}_r)$ versus $cor^2(y, \hat{y}_\theta)$

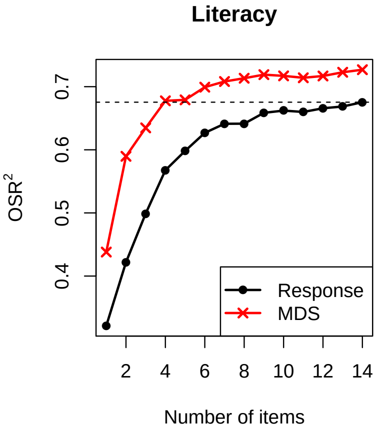


Figure: $cor^2(y, \hat{y}_{r_1,\ldots,r_k})$ versus $cor^2(y, \hat{y}_{\theta_1,\ldots,\theta_k})$

Latent structure extraction

- Multidimensional scaling

  Tang, X., Wang, Z., He, Q., Liu, J., and Ying, Z. (2019) Latent Feature Extraction for Process Data via Multidimensional Scaling. *Psychometrika*.

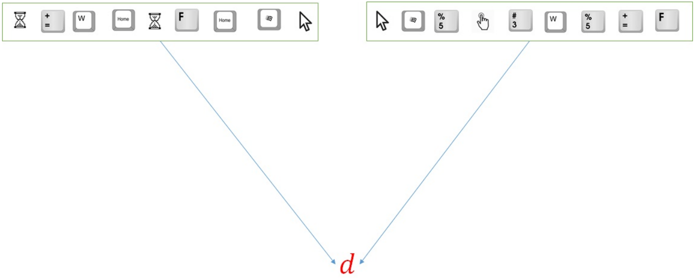
- Autoencoder

  Tang, X., Wang, Z., Liu, J., and Ying, Z. (2019) An Exploratory Analysis of the Latent Structure of Process Data via Action Sequence Autoencoder. *British Journal of Mathematical and Statistical Psychology*.

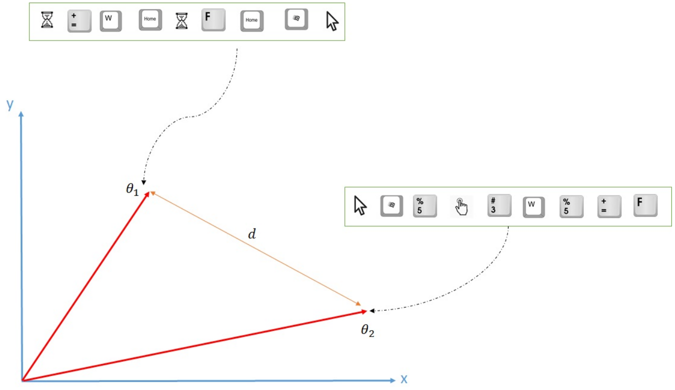
- R package

  Tang, X., Zhang, S., Wang, Z., Liu, J., and Ying, Z. (2021) ProcData: An R Package for Process Data Analysis. *Psychometrika*. To appear.

Multidimensional scaling



Multidimensional scaling



Multidimensional scaling

- Two response processes: $\mathrm{a}_i = (a_{i1}, ..., a_{in_i})$, $\mathrm{a}_j = (a_{j1}, ..., a_{jn_j})$

$$(\mathrm{a}_i, \mathrm{a}_j) \to d_{ij} = d(\mathrm{a}_i, \mathrm{a}_j)$$

- The distance Gómez-Alonso and Valls (2008)

$$d(\mathrm{a}_i, \mathrm{a}_j) = \frac{f(\mathrm{a}_i, \mathrm{a}_j) + g(\mathrm{a}_i, \mathrm{a}_j)}{n_i + n_j},$$

- Common actions:

$$f(\mathrm{a}_i, \mathrm{a}_j) = \frac{\sum_{a \in C_{ij}} \sum_{k=1}^{K_{ij}^a} |s_i^a(k) - s_j^a(k)|}{\max\{n_i, n_j\}},$$

- Uncommon actions:

$$g(\mathrm{a}_i, \mathrm{a}_j) = \sum_{a \in U_{ij}} n_i^a + \sum_{a \in U_{ji}} n_j^a,$$

Multidimensional scaling

- Two response processes: $\mathbf{a}_i = (a_{i1}, \ldots, a_{in_i})$, $\mathbf{a}_j = (a_{j1}, \ldots, a_{jn_j})$

$$(\mathbf{a}_i, \mathbf{a}_j) \to d_{ij} = d(\mathbf{a}_i, \mathbf{a}_j)$$

- The distance Gómez-Alonso and Valls (2008)

$$d(\mathbf{a}_i, \mathbf{a}_j) = \frac{f(\mathbf{a}_i, \mathbf{a}_j) + g(\mathbf{a}_i, \mathbf{a}_j)}{n_i + n_j},$$

- Common actions:

$$f(\mathbf{a}_i, \mathbf{a}_j) = \frac{\sum_{a \in C_{ij}} \sum_{k=1}^{K_{ij}^a} |s_i^a(k) - s_j^a(k)|}{\max\{n_i, n_j\}},$$

- Uncommon actions:

$$g(\mathbf{a}_i, \mathbf{a}_j) = \sum_{a \in U_{ij}} n_i^a + \sum_{a \in U_{ji}} n_j^a,$$

Multidimensional scaling

- Two response processes: $\mathrm{a}_i = (a_{i1}, \ldots, a_{in_i})$, $\mathrm{a}_j = (a_{j1}, \ldots, a_{jn_j})$

$$(\mathrm{a}_i, \mathrm{a}_j) \to d_{ij} = d(\mathrm{a}_i, \mathrm{a}_j)$$

- The distance Gómez-Alonso and Valls (2008)

$$d(\mathrm{a}_i, \mathrm{a}_j) = \frac{f(\mathrm{a}_i, \mathrm{a}_j) + g(\mathrm{a}_i, \mathrm{a}_j)}{n_i + n_j},$$

- Common actions:

$$f(\mathrm{a}_i, \mathrm{a}_j) = \frac{\sum_{a \in C_{ij}} \sum_{k=1}^{K_{ij}^a} |s_i^a(k) - s_j^a(k)|}{\max\{n_i, n_j\}},$$

- Uncommon actions:

$$g(\mathrm{a}_i, \mathrm{a}_j) = \sum_{a \in U_{ij}} n_i^a + \sum_{a \in U_{ji}} n_j^a,$$

Multidimensional scaling

- Two response processes: $\mathrm{a}_i = (a_{i1}, \ldots, a_{in_i})$, $\mathrm{a}_j = (a_{j1}, \ldots, a_{jn_j})$

$$(\mathrm{a}_i, \mathrm{a}_j) \to d_{ij} = d(\mathrm{a}_i, \mathrm{a}_j)$$

- The distance Gómez-Alonso and Valls (2008)

$$d(\mathrm{a}_i, \mathrm{a}_j) = \frac{f(\mathrm{a}_i, \mathrm{a}_j) + g(\mathrm{a}_i, \mathrm{a}_j)}{n_i + n_j},$$

- Common actions:

$$f(\mathrm{a}_i, \mathrm{a}_j) = \frac{\sum_{a \in C_{ij}} \sum_{k=1}^{K_{ij}^a} |s_i^a(k) - s_j^a(k)|}{\max\{n_i, n_j\}},$$

- Uncommon actions:

$$g(\mathrm{a}_i, \mathrm{a}_j) = \sum_{a \in U_{ij}} n_i^a + \sum_{a \in U_{ji}} n_j^a,$$

Multidimensional scaling

- Two response processes: $\mathrm{a}_i = (a_{i1}, \ldots, a_{in_1})$, $\mathrm{a}_j = (a_{j1}, \ldots, a_{jn_2})$

$$(\mathrm{a}_i, \mathrm{a}_j) \to d_{ij}, \quad 1 \le i, j \le n$$

- Distance matrices $D = (d_{ij})_{n \times n}$
- Latent variable $\theta_i \in \mathbb{R}^k$.

$$d_{ij} = \varphi(\theta_i, \theta_j) + \varepsilon_{ij}$$

where $\varphi(\theta_i, \theta_j) = |\theta_i - \theta_j|$.

- Optimization

$$\min_{\theta_1, \ldots, \theta_n} \sum_{i<j} |d_{ij} - \varphi(\theta_i, \theta_j)|^2.$$

- Tang, X., Wang, Z., He, Q., Liu, J., and Ying, Z. (2019) Latent Feature Extraction for Process Data via Multidimensional Scaling. *Psychometrika*.

Multidimensional scaling

- Two response processes: $\mathrm{a}_i = (a_{i1}, \ldots, a_{in_1})$, $\mathrm{a}_j = (a_{j1}, \ldots, a_{jn_2})$

$$(\mathrm{a}_i, \mathrm{a}_j) \to d_{ij}, \quad 1 \leq i, j \leq n$$

- Distance matrices $D = (d_{ij})_{n \times n}$

- Latent variable $\theta_i \in \mathbb{R}^k$.

$$d_{ij} = \varphi(\theta_i, \theta_j) + \varepsilon_{ij}$$

where $\varphi(\theta_i, \theta_j) = |\theta_i - \theta_j|$.

- Optimization

$$\min_{\theta_1, \ldots, \theta_n} \sum_{i<j} |d_{ij} - \varphi(\theta_i, \theta_j)|^2.$$

- Tang, X., Wang, Z., He, Q., Liu, J., and Ying, Z. (2019) Latent Feature Extraction for Process Data via Multidimensional Scaling. *Psychometrika*.

Multidimensional scaling

- Two response processes: $\mathrm{a}_i = (a_{i1}, \ldots, a_{in_1})$, $\mathrm{a}_j = (a_{j1}, \ldots, a_{jn_2})$

$$(\mathrm{a}_i, \mathrm{a}_j) \to d_{ij}, \quad 1 \le i, j \le n$$

- Distance matrices $D = (d_{ij})_{n \times n}$

- Latent variable $\theta_i \in \mathbb{R}^k$.

$$d_{ij} = \varphi(\theta_i, \theta_j) + \varepsilon_{ij}$$

where $\varphi(\theta_i, \theta_j) = |\theta_i - \theta_j|$.

- Optimization

$$\min_{\theta_1, \ldots, \theta_n} \sum_{i<j} |d_{ij} - \varphi(\theta_i, \theta_j)|^2.$$

- Tang, X., Wang, Z., He, Q., Liu, J., and Ying, Z. (2019) Latent Feature Extraction for Process Data via Multidimensional Scaling. *Psychometrika*.

Multidimensional scaling

- Two response processes: $\mathrm{a}_i = (a_{i1}, \ldots, a_{in_1})$, $\mathrm{a}_j = (a_{j1}, \ldots, a_{jn_2})$

$$(\mathrm{a}_i, \mathrm{a}_j) \to d_{ij}, \quad 1 \leq i, j \leq n$$

- Distance matrices $D = (d_{ij})_{n \times n}$
- Latent variable $\theta_i \in \mathbb{R}^k$.

$$d_{ij} = \varphi(\theta_i, \theta_j) + \varepsilon_{ij}$$

where $\varphi(\theta_i, \theta_j) = |\theta_i - \theta_j|$.

- Optimization

$$\min_{\theta_1, \ldots, \theta_n} \sum_{i<j} |d_{ij} - \varphi(\theta_i, \theta_j)|^2.$$

- Tang, X., Wang, Z., He, Q., Liu, J., and Ying, Z. (2019) Latent Feature Extraction for Process Data via Multidimensional Scaling. *Psychometrika*.

Multidimensional scaling

- Two response processes: $\mathbf{a}_i = (a_{i1}, ..., a_{in_1})$, $\mathbf{a}_j = (a_{j1}, ..., a_{jn_2})$

$$(\mathbf{a}_i, \mathbf{a}_j) \to d_{ij}, \quad 1 \leq i, j \leq n$$

- Distance matrices $D = (d_{ij})_{n \times n}$
- Latent variable $\theta_i \in \mathbb{R}^k$.

$$d_{ij} = \varphi(\theta_i, \theta_j) + \varepsilon_{ij}$$

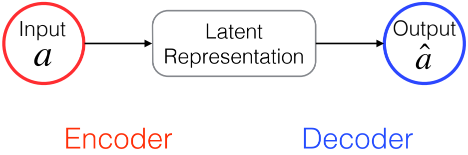
where $\varphi(\theta_i, \theta_j) = |\theta_i - \theta_j|$.

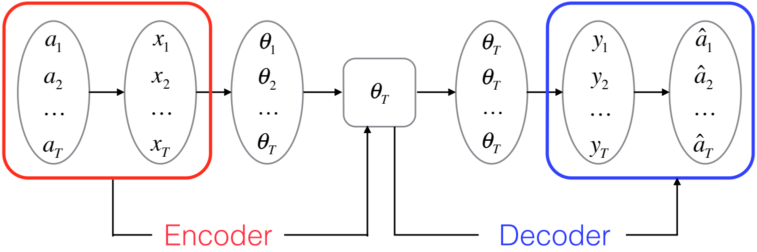
- Optimization

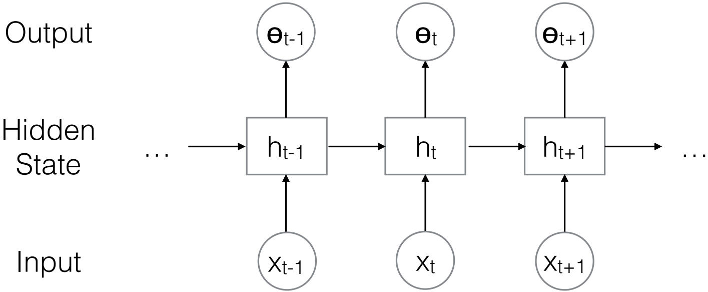
$$\min_{\theta_1,\ldots,\theta_n} \sum_{i<j} |d_{ij} - \varphi(\theta_i, \theta_j)|^2.$$

- Tang, X., Wang, Z., He, Q., Liu, J., and Ying, Z. (2019) Latent Feature Extraction for Process Data via Multidimensional Scaling. *Psychometrika*.

Autoencoder




a_1
a_2
...
a_T
x_1
x_2
...
x_T
θ_1
θ_2
...
θ_T
θ_T
θ_T
θ_T
...
θ_T
y_1
y_2
...
y_T
$\hat{a}_1$
$\hat{a}_2$
...
$\hat{a}_T$
Encoder
Decoder


Output
Hidden State
Input
ot-1
ot
ot+1
...
ht-1
ht
ht+1
...
xt-1
xt
xt+1

Autoencoder

- Autoencoder via tenorflow
- Tang, X., Wang, Z., Liu, J., and Ying, Z. (2019) An Exploratory Analysis of the Latent Structure of Process Data via Action Sequence Autoencoder. *British Journal of Mathematical and Statistical Psychology*.

Criterion

- a $\Rightarrow \theta \in \mathbb{R}^K$: multidimensional scaling or autoencoder
- a $\Rightarrow r \in \{\checkmark, \times\}$: task accomplishment

Assessing latent variable

- How much information did we lose to r?
- $\hat{r}_\theta$: prediction of task accomplishment based on θ.
- To what extend θ captures the information in r.

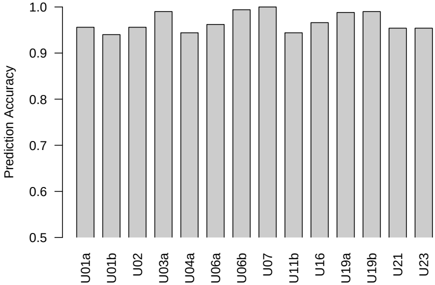


Figure: $P(r = \hat{r}_\theta)$ based on MDS

Assessing latent variable through prediction

- How much additional information do we gain?
- y: a different variable, such as numeracy score.
  - $\hat{y}_\theta$: prediction based on θ

    Versus
  - $\hat{y}_r$: prediction based on r

cor(y, $\hat{y}$)

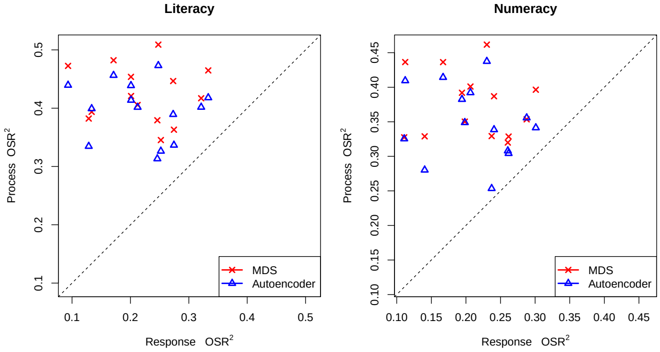


Figure: $\hat{y}_\theta$ versus $\hat{y}_r$

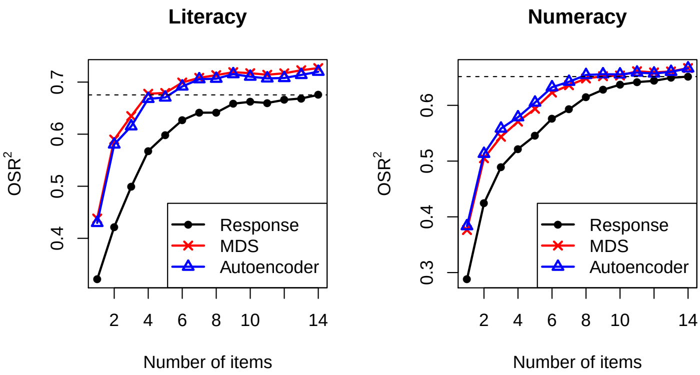


Figure: $\hat{y}(\theta_1, ..., \theta_k)$ versus $\hat{y}(r_1, ..., r_k)$

Demographic variables

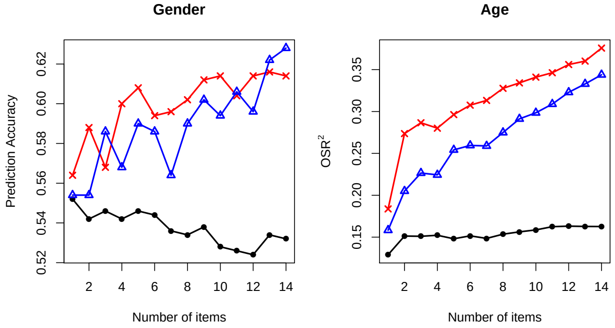


Figure: Prediction of age and gender

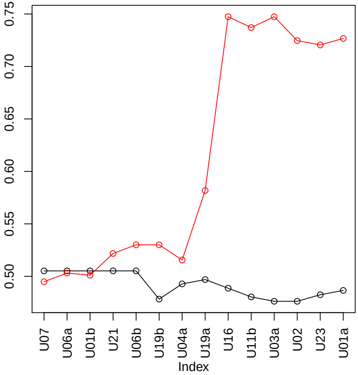


Figure: Prediction of country

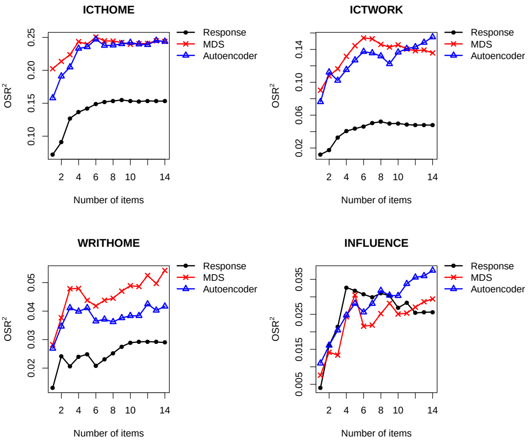


Figure: ICTHOME,ICTWORK,WRITHOME,INFLUENCE

Closer look at the latent variables – principle components

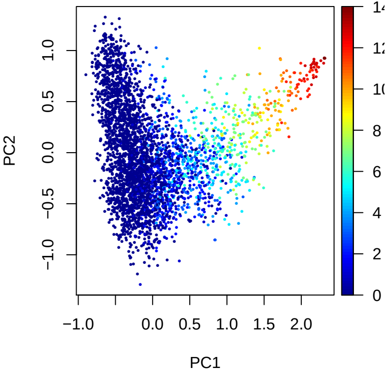


Figure: Principle components of θ. PC1 correlation with number of skipped items: 0.88

Latent variable in subpopulations

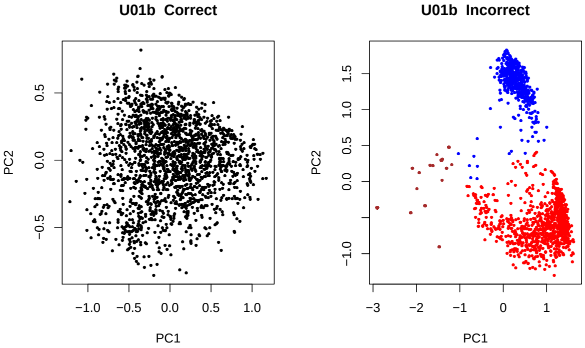
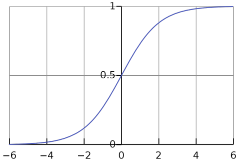


Figure: Correct vs incorrect

Applications

- Partial score: improving assessment accuracy
- Removing differential item functioning (DIF)

Partial scores



- IRT model

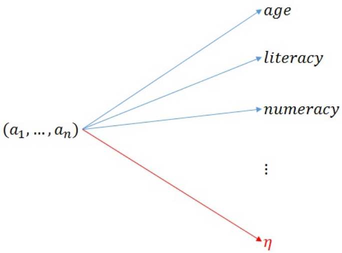
$$P(r_j = 1|\eta) = \frac{e^{a_j(\eta - b_j)}}{1 + e^{a_j(\eta - b_j)}}$$

- Assessment (grading/scoring) via maximum likelihood estimate

$$\hat{\eta}(r_1, ..., r_J) = \arg\max_{\eta} \prod_j P(r_j|\eta)$$

- Process-data-based scores: $\hat{\eta}(\mathbf{a}_j)$

Partial scores



- Difficulty: validity
- Partial scores based on the entire response process
- Scores guided by y.
- Generalization: assisting any measurable characteristics

Partial scores guided by response

- Train a scoring rule $f(\mathsf{a})$ towards η.
- $r = g(\eta) + \varepsilon \sim \mathsf{a}$, ε is a-predictable.
- $g(\eta) + \epsilon \sim \mathsf{a}$
- Proxies
  - $\hat{\eta}(r_{-j}) = \eta + \epsilon$
  - a: process features θ

$$\hat{\eta}(r_{-j}) \sim \theta$$

Partial scores guided by response

- Train a scoring rule $f(\text{a})$ towards η.
- $r = g(\eta) + \varepsilon \sim \text{a}$, ε is a-predictable.
- $g(\eta) + \epsilon \sim \text{a}$
- Proxies
  - $\hat{\eta}(r_{-j}) = \eta + \epsilon$
  - a: process features θ

$$\hat{\eta}(r_{-j}) \sim \theta$$

Partial scores guided by response

- Train a scoring rule $f(\mathrm{a})$ towards η.
- $r = g(\eta) + \varepsilon \sim \mathrm{a}$, ε is a-predictable.
- $g(\eta) + \epsilon \sim \mathrm{a}$
- Proxies
  - $\hat{\eta}(r_{-j}) = \eta + \epsilon$
  - a: process features θ

$$\hat{\eta}(r_{-j}) \sim \theta$$

Partial scores guided by response

- Train a scoring rule $f(\mathrm{a})$ towards η.
- $r = g(\eta) + \varepsilon \sim \mathrm{a}$, ε is a-predictable.
- $g(\eta) + \epsilon \sim \mathrm{a}$
- Proxies
  - $\hat{\eta}(r_{-j}) = \eta + \epsilon$
  - a: process features θ

$$\hat{\eta}(r_{-j}) \sim \theta$$

Partial scores guided by response

- Train a scoring rule $f(\mathrm{a})$ towards η.
- $r = g(\eta) + \varepsilon \sim \mathrm{a}$, ε is a-predictable.
- $g(\eta) + \epsilon \sim \mathrm{a}$
- Proxies
  - $\hat{\eta}(r_{-j}) = \eta + \epsilon$
  - a: process features θ

$$\hat{\eta}(r_{-j}) \sim \theta$$

Partial scores guided by response

- Train a scoring rule $f(a)$ towards η.
- $r = g(\eta) + \varepsilon \sim a$, ε is a-predictable.
- $g(\eta) + \epsilon \sim a$
- Proxies
  - $\hat{\eta}(r_{-j}) = \eta + \epsilon$
  - a: process features θ

$$\hat{\eta}(r_{-j}) \sim \theta$$

Partial scores guided by response

1. Extract process feature θ_j for each item
2. Compute IRT score $\eta(r_{-j})$
3. Regression: $\hat{\eta} = f(\theta_j)$
4. Score: $f(\theta_j)$

Evaluation criterion – improving reliability

- 14 PSTRE items in total
- Randomization
  - Session 1 (7 items): $\hat{\eta}_1$ IRT estimate, $\tilde{\eta}_1$ process data estimate
  - Session 2 (7 items): $\hat{\eta}_2$ IRT estimate
- Compare $cor(\hat{\eta}_1, \hat{\eta}_2)$ against $cor(\tilde{\eta}_1, \hat{\eta}_2)$

Evaluation criterion – improving reliability

- 14 PSTRE items in total
- Randomization
  - Session 1 (7 items): $\hat{\eta}_1$ IRT estimate, $\tilde{\eta}_1$ process data estimate
  - Session 2 (7 items): $\hat{\eta}_2$ IRT estimate
- Compare $cor(\hat{\eta}_1, \hat{\eta}_2)$ against $cor(\tilde{\eta}_1, \hat{\eta}_2)$

Evaluation criterion – improving reliability

- 14 PSTRE items in total
- Randomization
  - Session 1 (7 items): $\hat{\eta}_1$ IRT estimate, $\tilde{\eta}_1$ process data estimate
  - Session 2 (7 items): $\hat{\eta}_2$ IRT estimate
- Compare $cor(\hat{\eta}_1, \hat{\eta}_2)$ against $cor(\tilde{\eta}_1, \hat{\eta}_2)$

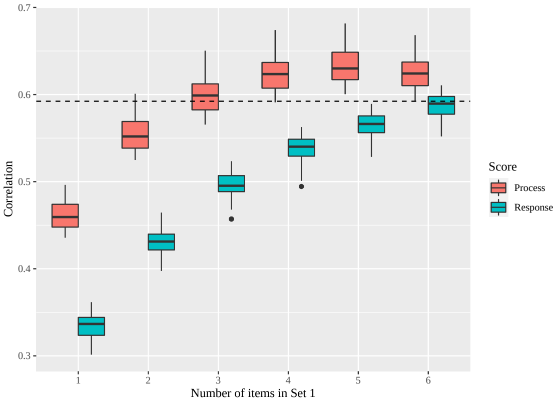
Evaluation criterion – improving reliability

- 14 PSTRE items in total
- Randomization
  - Session 1 (7 items): $\hat{\eta}_1$ IRT estimate, $\tilde{\eta}_1$ process data estimate
  - Session 2 (7 items): $\hat{\eta}_2$ IRT estimate
- Compare $cor(\hat{\eta}_1, \hat{\eta}_2)$ against $cor(\tilde{\eta}_1, \hat{\eta}_2)$

Evaluation criterion – improving reliability

- 14 PSTRE items in total
- Randomization
  - Session 1 (7 items): $\hat{\eta}_1$ IRT estimate, $\tilde{\eta}_1$ process data estimate
  - Session 2 (7 items): $\hat{\eta}_2$ IRT estimate
- Compare $cor(\hat{\eta}_1, \hat{\eta}_2)$ against $cor(\tilde{\eta}_1, \hat{\eta}_2)$

Out-of-sample Reliability



Differential item functioning (DIF)

- About differential item functioning
- Literature: identifying DIF
- Process data: removing DIF

Framework

- Response Y with item response function

$$r \sim \eta, x_1, ..., x_m$$

$$f(r|\eta, x_1, ..., x_m)$$

- Assessment model

$$r \sim \eta$$

- Observed item response function

$$f(r|\eta) = \int f(r|\eta, x_1, ..., x_m)\pi(x_1, ..., x_m|\eta)dx_1...dx_m$$

- Two groups: 1 and 2.

$$f_g(r|\eta) = \int f(r|\eta, x_1, ..., x_m)\pi_g(x_1, ..., x_m|\eta)dx_1...dx_m$$

- DIF: unbalanced distributions $\pi_g(\eta, x_1, ..., x_m)$ in groups 1 and 2.

Framework

- Response Y with item response function

$$r \sim \eta, x_1, ..., x_m$$

$$f(r|\eta, x_1, ..., x_m)$$

- Assessment model

$$r \sim \eta$$

- Observed item response function

$$f(r|\eta) = \int f(r|\eta, x_1, ..., x_m)\pi(x_1, ..., x_m|\eta)dx_1...dx_m$$

- Two groups: 1 and 2.

$$f_g(r|\eta) = \int f(r|\eta, x_1, ..., x_m)\pi_g(x_1, ..., x_m|\eta)dx_1...dx_m$$

- DIF: unbalanced distributions $\pi_g(\eta, x_1, ..., x_m)$ in groups 1 and 2.

Framework

- Response Y with item response function

$$r \sim \eta, x_1, ..., x_m$$

$$f(r|\eta, x_1, ..., x_m)$$

- Assessment model

$$r \sim \eta$$

- Observed item response function

$$f(r|\eta) = \int f(r|\eta, x_1, ..., x_m)\pi(x_1, ..., x_m|\eta)dx_1...dx_m$$

- Two groups: 1 and 2.

$$f_g(r|\eta) = \int f(r|\eta, x_1, ..., x_m)\pi_g(x_1, ..., x_m|\eta)dx_1...dx_m$$

- DIF: unbalanced distributions $\pi_g(\eta, x_1, ..., x_m)$ in groups 1 and 2.

Framework

- Response Y with item response function

$$r \sim \eta, x_1, ..., x_m$$

$$f(r|\eta, x_1, ..., x_m)$$

- Assessment model

$$r \sim \eta$$

- Observed item response function

$$f(r|\eta) = \int f(r|\eta, x_1, ..., x_m)\pi(x_1, ..., x_m|\eta)dx_1...dx_m$$

- Two groups: 1 and 2.

$$f_g(r|\eta) = \int f(r|\eta, x_1, ..., x_m)\pi_g(x_1, ..., x_m|\eta)dx_1...dx_m$$

- DIF: unbalanced distributions $\pi_g(\eta, x_1, ..., x_m)$ in groups 1 and 2.

Framework

- Response Y with item response function

$$r \sim \eta, x_1, ..., x_m$$

$$f(r|\eta, x_1, ..., x_m)$$

- Assessment model

$$r \sim \eta$$

- Observed item response function

$$f(r|\eta) = \int f(r|\eta, x_1, ..., x_m)\pi(x_1, ..., x_m|\eta)dx_1...dx_m$$

- Two groups: 1 and 2.

$$f_g(r|\eta) = \int f(r|\eta, x_1, ..., x_m)\pi_g(x_1, ..., x_m|\eta)dx_1...dx_m$$

- DIF: unbalanced distributions $\pi_g(\eta, x_1, ..., x_m)$ in groups 1 and 2.

Removing DIF

- Ideal solution: use the correct item response function $f(r|\eta, x_1, \ldots, x_m)$.
- $x_1, \ldots, x_m$ unobserved
- Use process data features as proxies, $\theta_1, \ldots, \theta_K$.

Removing DIF – technical aspects

- Over fitting: including all process feature

$$f(r|\eta, \theta_1, \ldots, \theta_K) = f(r|\theta_1, \ldots, \theta_K)$$

- Variable selection: minimum amount of process data so that

$$\|f_1(r|\eta, \theta_{i_1}, \ldots, \theta_{i_l}) - f_2(r|\eta, \theta_{i_1}, \ldots, \theta_{i_l})\| \approx 0$$

Why are process features good proxies?

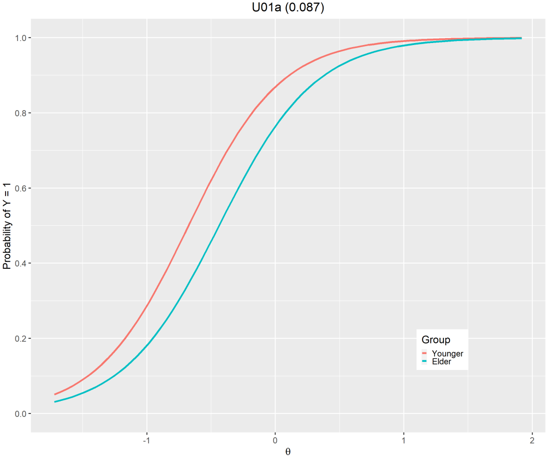
- Process features contain sufficient information to remove DIF

$$r = f(\theta_1, \ldots, \theta_K)$$

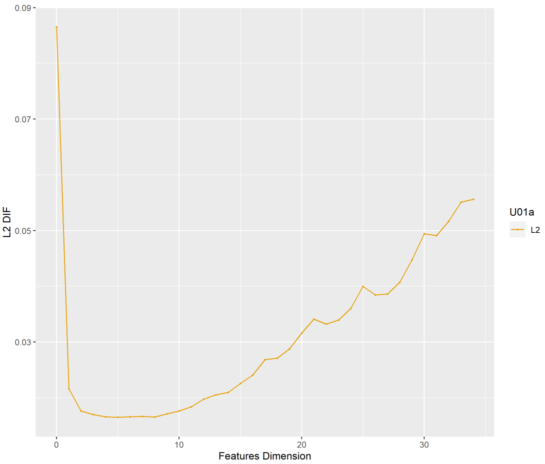
- Overfitting
- No extra information to estimate η
- DIF versus information: include minimum amount of process data to maintain non-differentiation of item functioning.
- A forward search algorithm

$$max_j \| f_1(r|\eta, \theta_{i_1}, \ldots, \theta_{i_k}, \theta_j) - f_2(r|\eta, \theta_{i_1}, \ldots, \theta_{i_k}, \theta_j) \|$$

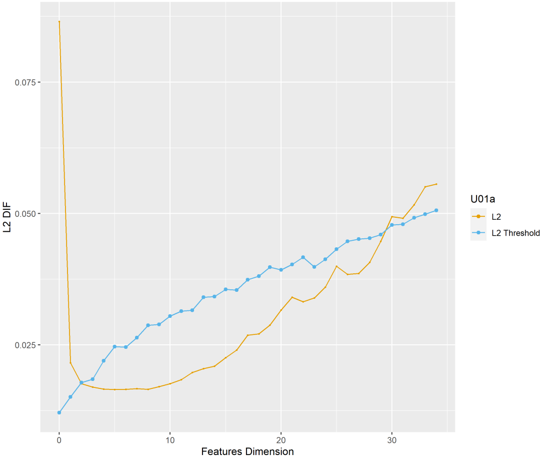
Empirical results



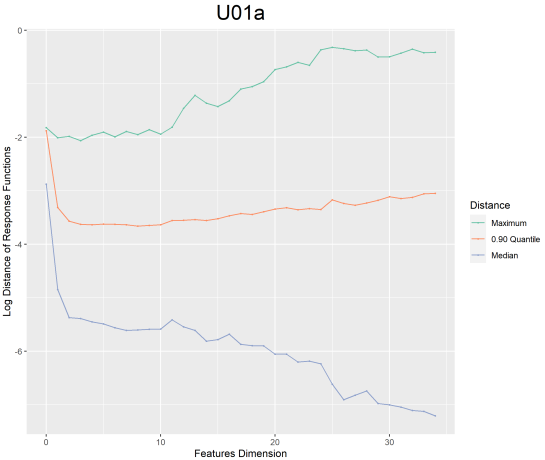
Empirical results



Empirical results



Empirical results



References

- http://www.scientifichpc.com/processdata/pub.html
- Feature extraction
  - Tang, X., Wang, Z., He, Q., Liu, J., and Ying, Z. (2020) Latent Feature Extraction for Process Data via Multidimensional Scaling. Psychometrika.
  - Tang, X., Wang, Z., Liu, J., and Ying, Z. (2020) An Exploratory Analysis of the Latent Structure of Process Data via Action Sequence Autoencoder. British Journal of Mathematical and Statistical Psychology.
  - Tang, X., Zhang, S., Wang, Z., Liu, J., and Ying, Z. (2021) ProcData: An R Package for Process Data Analysis. *Psychometrika*. To appear.
- Partial Scoring:
  - Zhang, S., Wang, Z., Qi, J., Liu, J., and Ying, Z. Accurate Assessment via Process Data.
- Removing differential item functioning

Acknowledgement

Zhiliang Ying

Zhi Wang, Jitong Qi, Brian Ling

OECD, Educational Testing Service: Qiwei He

National Science Foundation, Army Research Office