



國立中央大學
National Central University



International Symposium on Grids & Clouds 2017

Occupation recommendation with major programs for adolescents

Ankhtuya Ochirbat
Prof. Timothy K. Shih

Presenter: O. Ankhtuya

2017.03.08

Outline

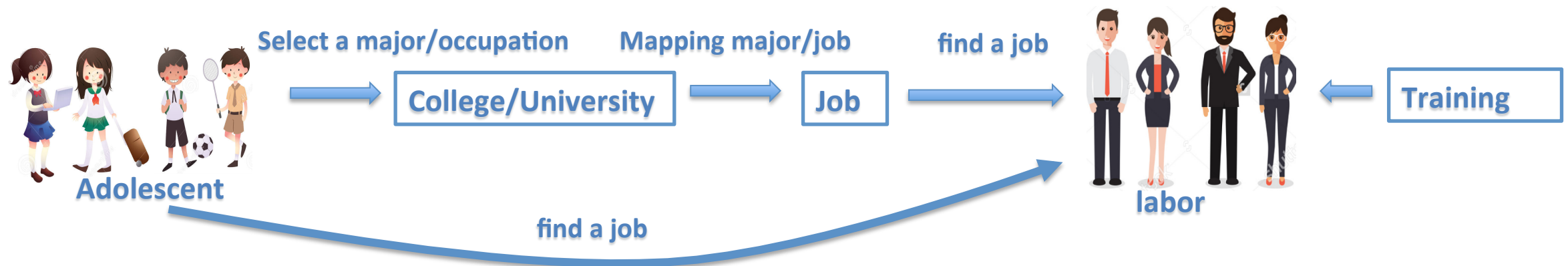
- Introduction
- Problem Statement
- Data Collection
- Methods
- Result
- Conclusion

Introduction

- **Young people** in high school or collage **make critical decisions** regarding what to study and **which career** path to pursue.
- Many of them end up switching to other majors because of **mismatching major choice** and **lack of processing information** through the professional study.
 - Such changes are **wasteful in time** and **resources** and they produce **emotional** and **economical stresses**.

Academic program/college major

- Adolescents first intend and select the **occupation** that they will work through it in the future.
- But the some **occupations** are not clear **to map into the academic program** to study or vice versa.

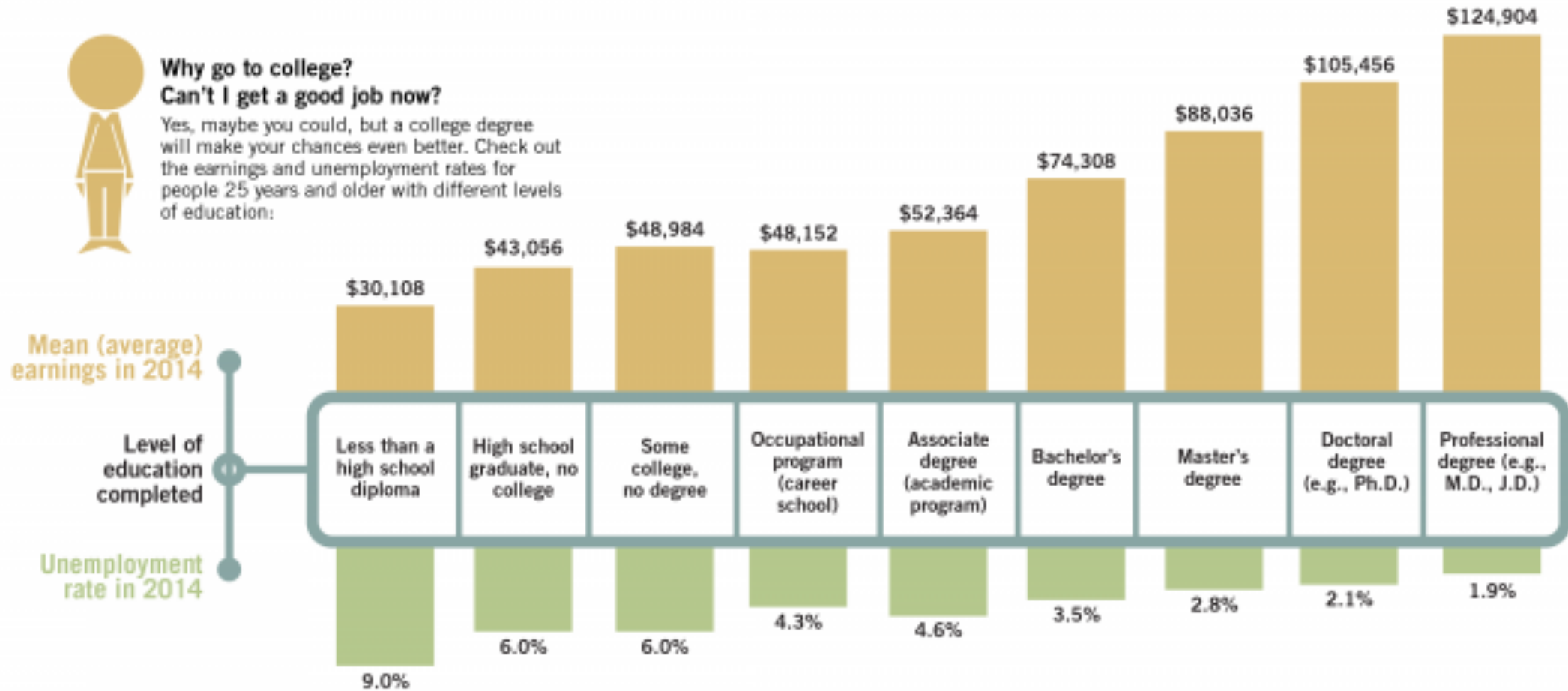


Why go to college?



Why go to college? Can't I get a good job now?

Yes, maybe you could, but a college degree will make your chances even better. Check out the earnings and unemployment rates for people 25 years and older with different levels of education:



Source: Bureau of Labor Statistics, Current Population Survey, unpublished tables 2015

Problem Statement

- But the main issue of difficulty making the **decisions of major choices** for the students is a lack of knowledge and information about **majors and its mapping to occupations.**
- Hence, it is essential to build **Automatic Recommendation System.**



Related works (1/3)

- D. Tsogzolmaa, S. Delgermaa, and P. Ulziisaikhan (2013) investigated **how to choose major properly among high school students** in Mongolia.
 - **27.4%** of students answered choosing the major by their **interest**,
 - **52.9%** of students replied the major choice made by **family member** or relatives' suggestion, and
 - **15.3%** of students answered that they would consider a **labor market**, a possibility of the **finding the job** as well as other factors.

Related works (2/3)

- According to the **M.Erdenechimeg (2013)** study showed that
 - **4.8%** of students – **vocational training center**,
 - **91.2%** of students – college/university.
- But in the real case, **43%** of high school graduates in **2013 year** were **enrolled in colleges or universities**, the National statistical office of Mongolia reported.

Related works (3/3)

- B.Bolormaa, B. Oyunsuren, Ch. Altangerel, and Ch. Tsolmon (2015) concluded that 20.7% of 324 students have not yet chosen the major; the reason is that students have **a lack of occupational information** a lot, and schools have **not** provided the **guidance** and **counseling**.
- Hence, they recommend that **Online automate system**, which can provide occupational information including contexts, requirements, tasks and competence needed for the particular majors/occupations and the guidance/counseling for students are needed.

What is Recommender techniques?

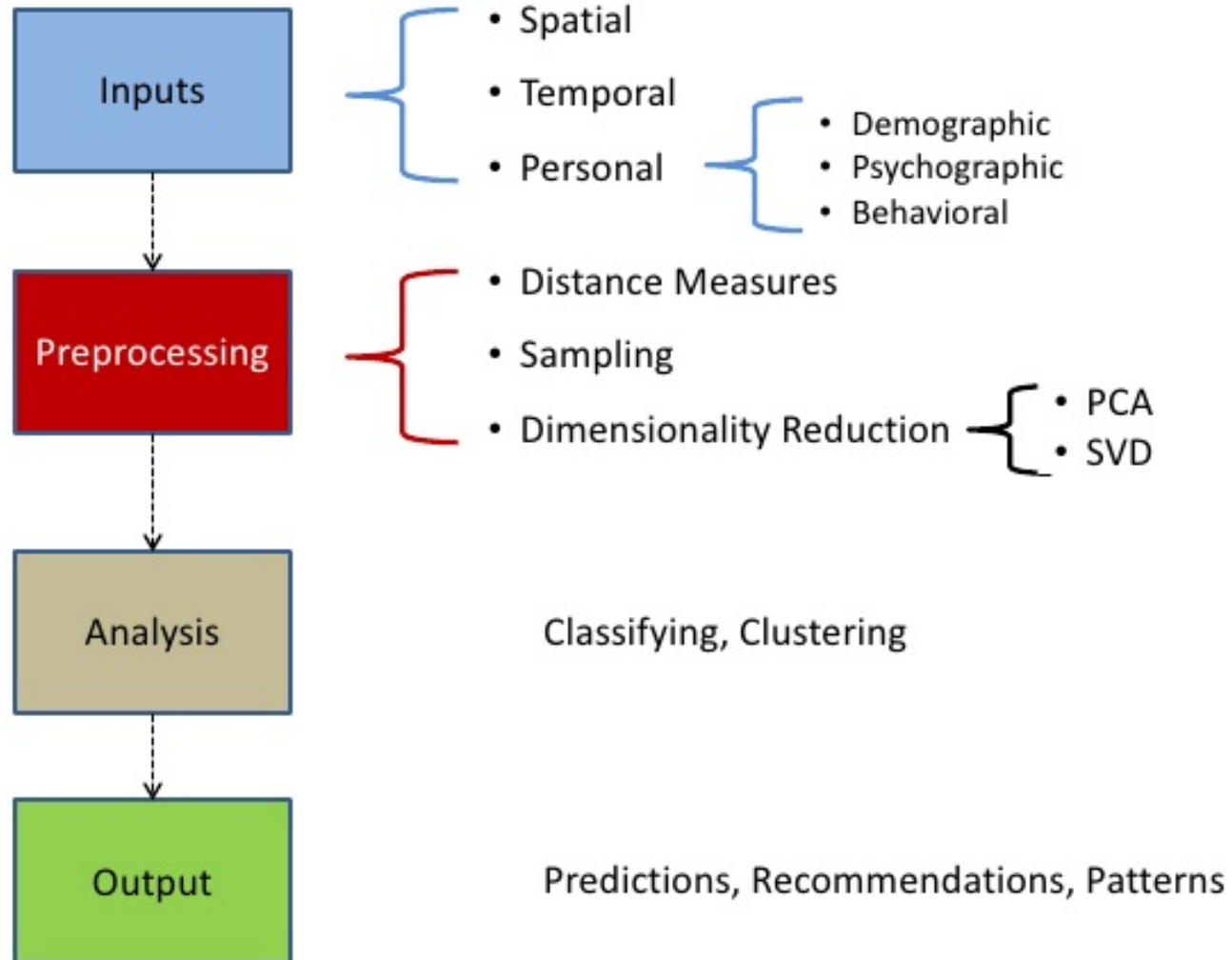
Recommender techniques are information agents that attempt to predict which items out of large pool a user may be interested in and recommend the best ones to a target user.

Examples:



The New York Times

Recommender Systems



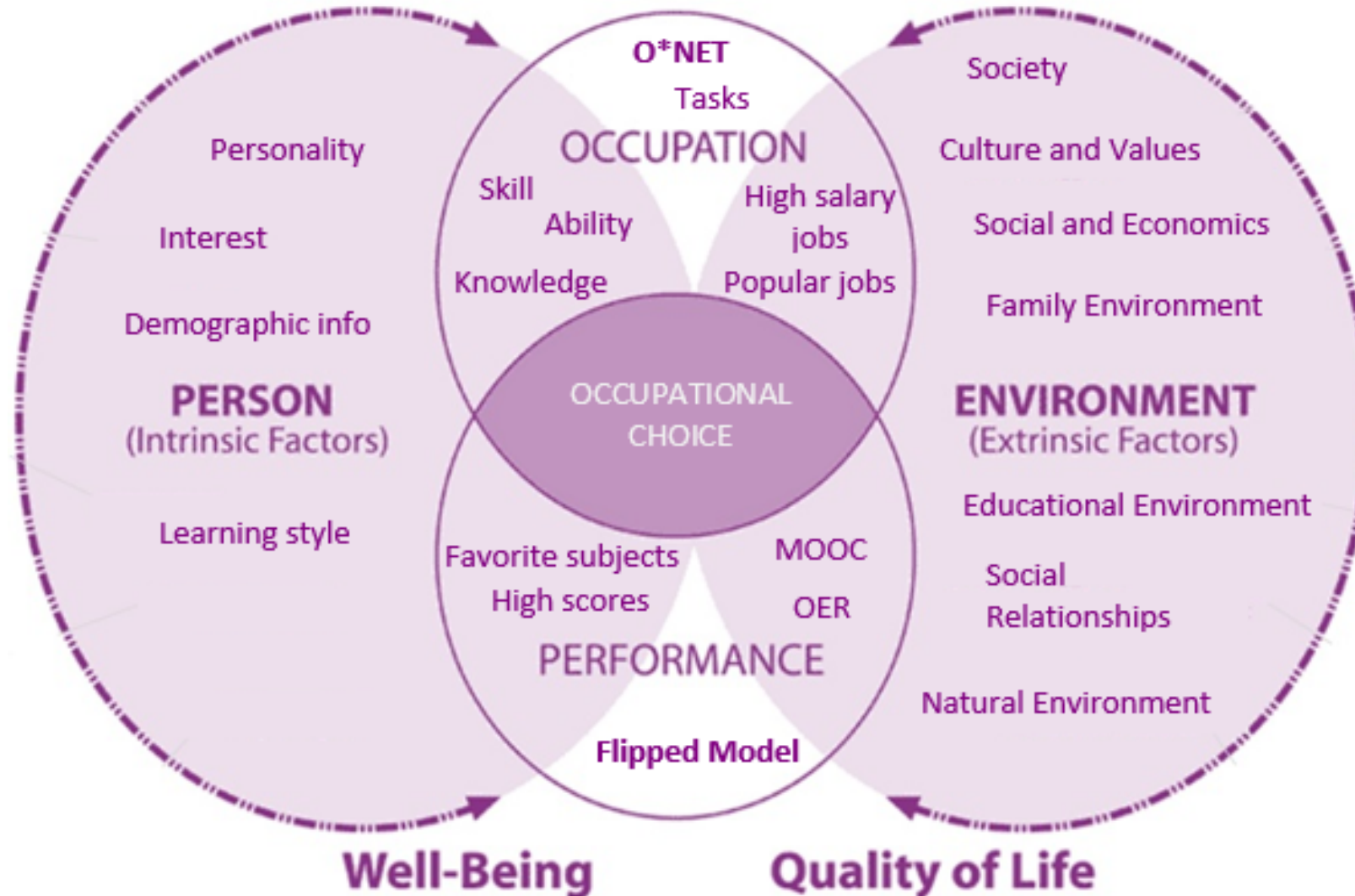
Person-Environment-Occupation Performance (PEOP) model

www.LifelongLearningWithOT.wordpress.com



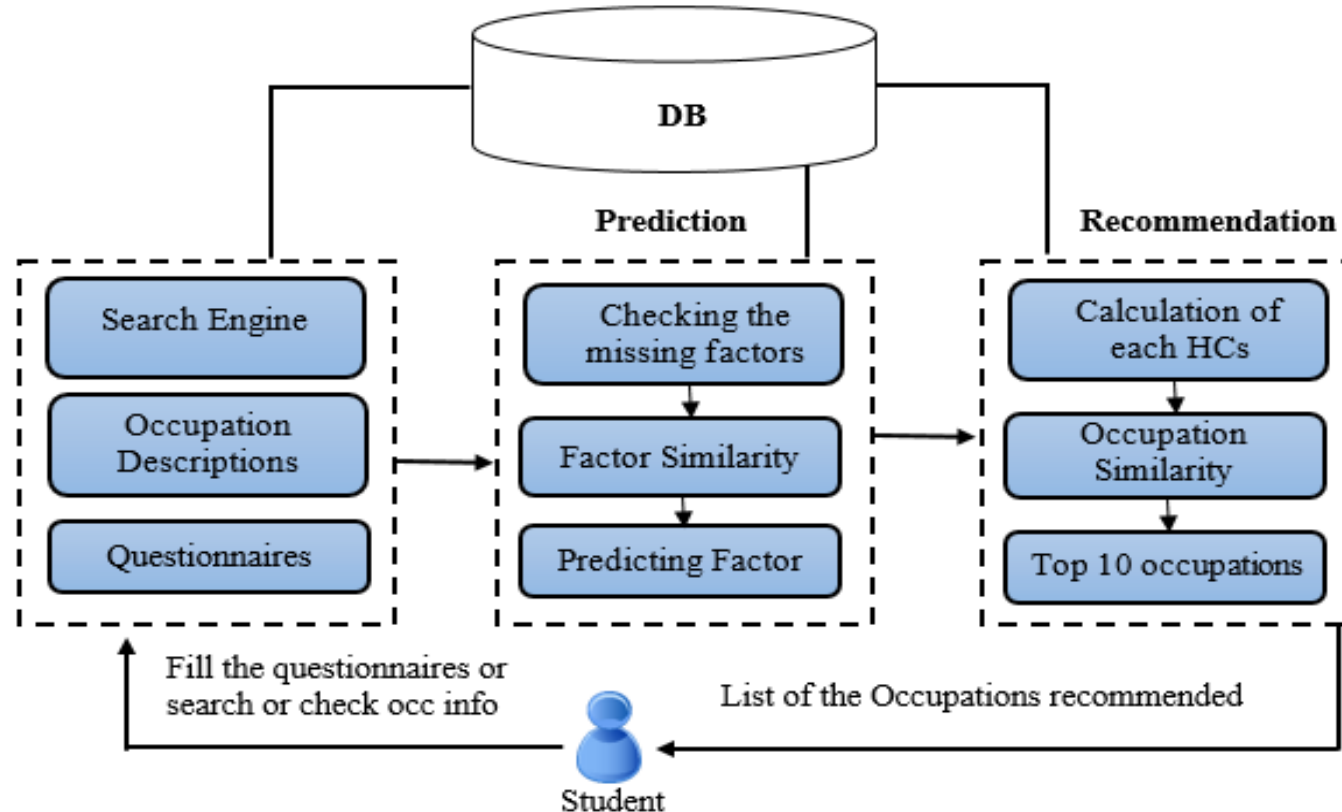
References: Christiansen CH, Baum CM & Bass-Haugen J. (2005). *Occupational therapy: Performance, participation and well-being* (3rd ed). Thorofare NJ: SLACK incorporated
 Duncan EES (2012) *Foundations for Practice in Occupational Therapy* (5th Ed) Edinburgh: Churchill Livingstone

Case study: PEOP model & Occupational choice



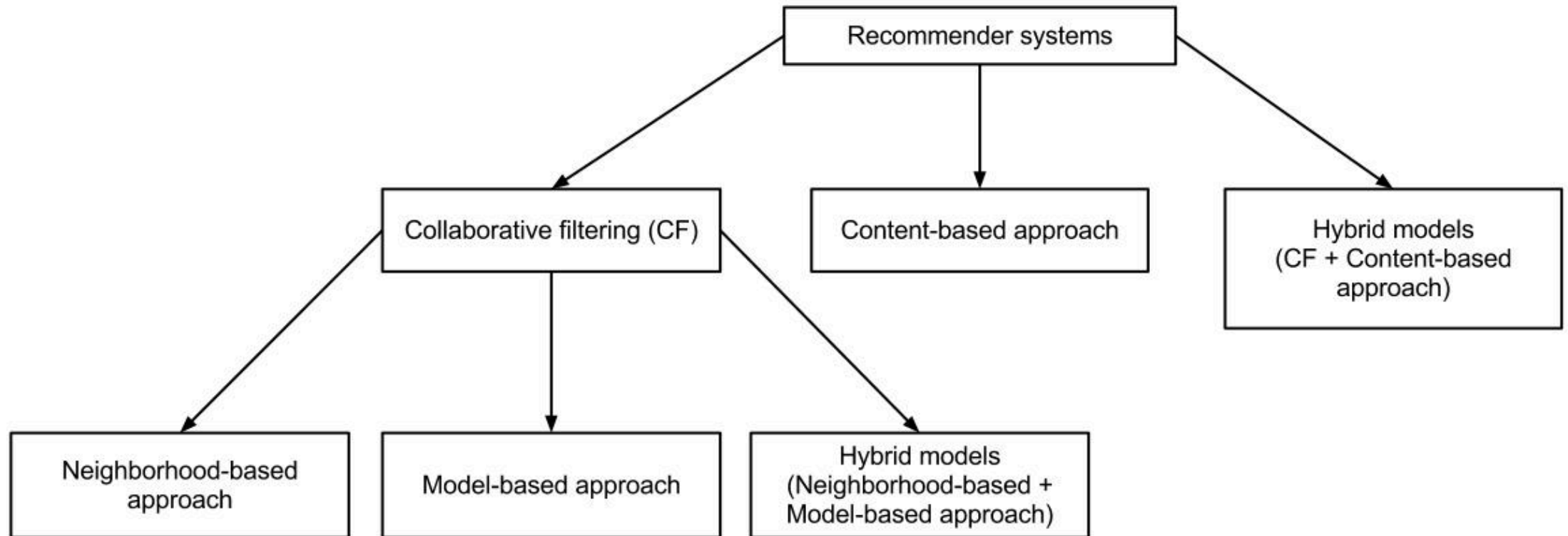
Occupation Recommendation

- **Occupation Recommender System (ORS)** is to assist adolescents in discovering the occupation interests.



Methods

- **Recommender Techniques category:**



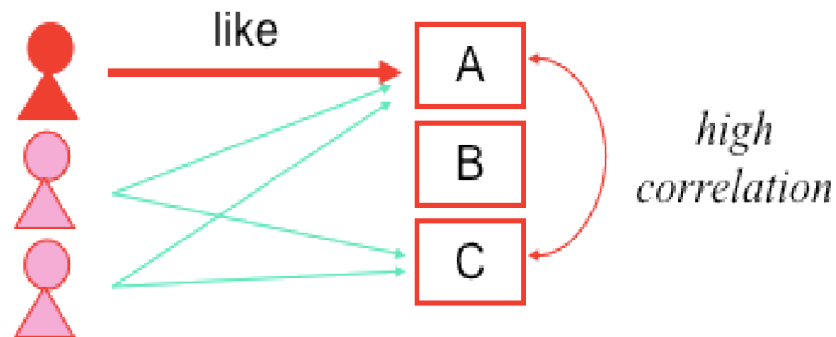
Collaborative Filtering (CF)

- CF techniques use the datasets of preferences for **items** by **users** to **predict** additional topics or products the active user might like.
- CF consist of three main categories:
 - memory-based,
 - model-based, and
 - hybrid CF algorithms

Collaborative Filtering

- **Memory-based Collaborative Filtering**

- It utilizes the entire user(student)-item(occupation) dataset to generate a prediction.
- Uses user-item ratings matrix
- Makes item-to-item correlations (similarity)
- Finds items that are highly correlated
- Recommends items with highest correlation



Data normalization

- I converted those scores into **fuzzy values** in range [0,1] that can be representing in graphical view which is a membership function using Trapezoid function.

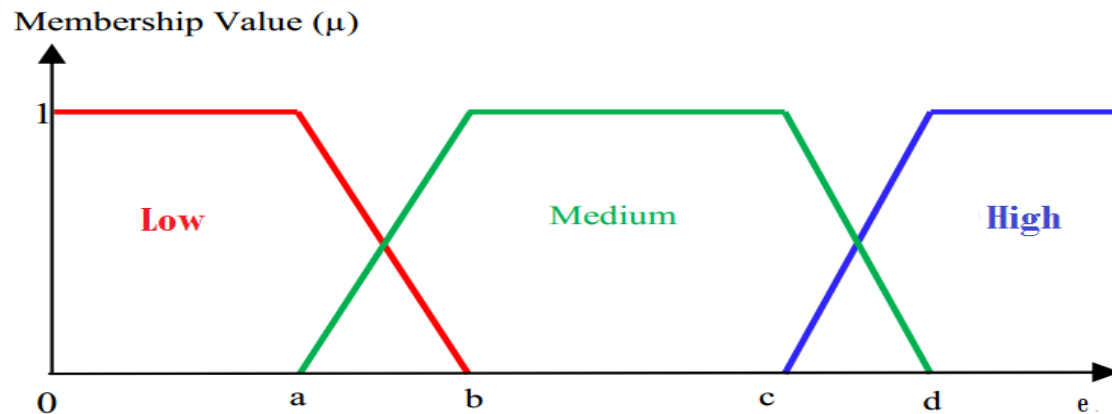


Figure 3. Membership function for interests of students and occupations

$$\mu_{\mu S}^O(x) = \begin{cases} 1 & x < a \\ \frac{x-a}{b-a} & a \ll x \ll b \\ 1 & b \ll x \ll c \\ \frac{d-x}{d-c} & c \ll x \ll d \\ 1 & x > d \end{cases} \quad (1)$$

Similarity computation

- **Fuzzy similarity** measures a similarity between item i and j is given by

$$sim(i, j) = \frac{\sum_{k=1}^m \min(x_{ik}, x_{jk})}{\sum_{k=1}^m \max(x_{ik}, x_{jk})} \quad (6)$$

Where $j = 1, \dots, n$

n = total numbers of occupation

m = number of factors in vocational interest

i = is the current student

x_i, x_j = Scores of student and occupation in interest

Prediction/Recommendation computation

- **Prediction: weighted sum** method computes the **it**.
 - Each **rating** is weighted by the corresponding **similarity** $S_{i,N}$ between **items** i and N . $R_{u,N}$ is the **rating** of **user** u on **item** N . **Prediction** $P_{u,i}$ as

$$P_{u,i} = \frac{\sum_{\text{all similar items } N} (S_{i,N} * R_{u,N})}{\sum_{\text{all similar items } N} (|S_{i,N}|)} \quad (7)$$

- **Top-N recommendation:**

Algorithm 1: Top n occupation recommendation algorithm

Inputs: studentID, occupations, n, factors

Output: Top-n recommended occupations for studentID

- 1: for every occupation i derived from questionnaire for studentID
 - 2: for every occupation j in database
 - 3: compute a similarity ss between i and j
 - 4: add studentID's preference for j , weighted by ss , to a running average
 - 5: return the top-n occupations, ranked by weighted average
-

Model-based CF

- **Model-based CF** algorithms involve building a model based on the dataset of ratings.

Regression model.

Regression model

- **The model** is developed using **regression analysis** to find patterns based on training data.
- **Vocational interests** (six factors - RIASEC) are predicted by using **personality** and **learning style** of students in (10) - (15).

$$R = 0.47 * openness + 0.06 * sensory + 0.07 * visual + 1.02 \quad (10)$$

$$I = 0.61 * openness + 0.05 * visual + 1.17 \quad (11)$$

$$A = 0.78 * openness + 0.06 * visual + 0.66 \quad (12)$$

$$S = 0.34 * Extraversion + 0.19 * Agreeableness + 0.14 * Conscientiousness + 0.27 * neuroticism + 0.36 * openness + 0.07 * active - 1.01 \quad (13)$$

$$E = 0.35 * Extraversion + 0.20 * neuroticism + 0.47 * openness + 0.04 * active + 0.03 * visual - 0.44 \quad (14)$$

$$C = 0.35 * Extraversion + 0.23 * neuroticism + 0.42 * openness + 0.06 * active + 0.04 * visual - 0.38 \quad (15)$$

Data Collection - Person

➤ **Variables associated with adolescents include:**

- Demographic Information;
- Personality;
- Interests;
- Learning style; and
- Intended major;

Welcome to **OCR**

Please provide your general information below:

Surname:

Firstname:

Gender: Male Female

Grade:

Suggested major by parent:

Intendend major:

Data Collection - Person

- **Participants**

- **190** students were participated in the study.
 - This study was carried out during the first semester of the 2014/2016 academic year in Mongolia.
 - **81** in grade 10, **107** in grade 11, and **2** in grade 12.
 - The participants' age range is from 15 to 19 years.
 - Female = 107 and male = 80

Data Collection - Person

- **Personality – Big Five Inventory**
(BFI - Goldberg, 1992, 1993)

- The **BFI** has the **44 questions** with responses made on a **Likert scale** format ranging from 1 = strongly disagree to 5 = strongly agree.

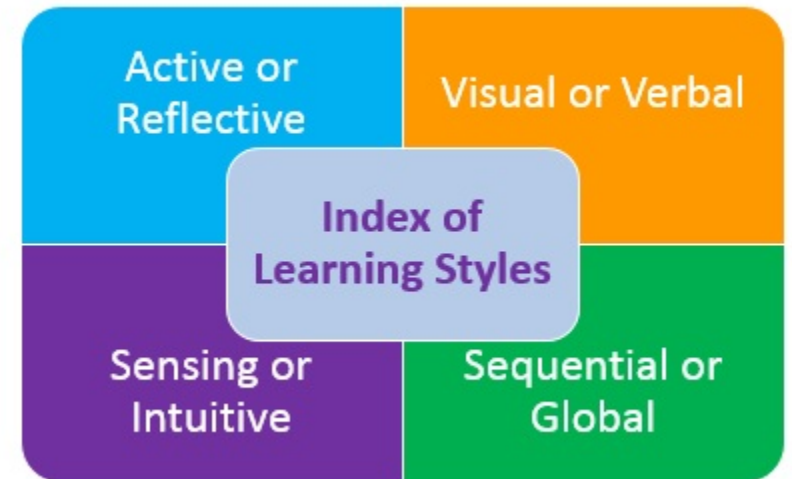
EXTRAVERSION	• Friendliness, Gregariousness, Assertiveness, Activity Level, Excitement-Seeking, Cheerfulness
AGREEABLENESS	• Trust, Morality, Altruism, Cooperation, Modesty, Sympathy
CONSCIENTIOUSNESS	• Self-efficacy, Orderliness, Dutifulness, Achievement-striving, Self-discipline, Cautiousness
NEUROTICISM	• Anxiety, Anger, Depression, Self-consciousness, Immoderation, Vulnerability
OPENNESS TO EXPERIENCE	• Imagination, Artistic Interests, Emotionality, Adventurousness, Intellect, Liberalism

Data Collection - Person

- **Index of Learning style**

(ILS - Soloman & Felder, 1999)

- People take in and process information **in different ways** based on their individual preferences.
- ILS contains 44 **two choices questions**.
- **4 dimensions**

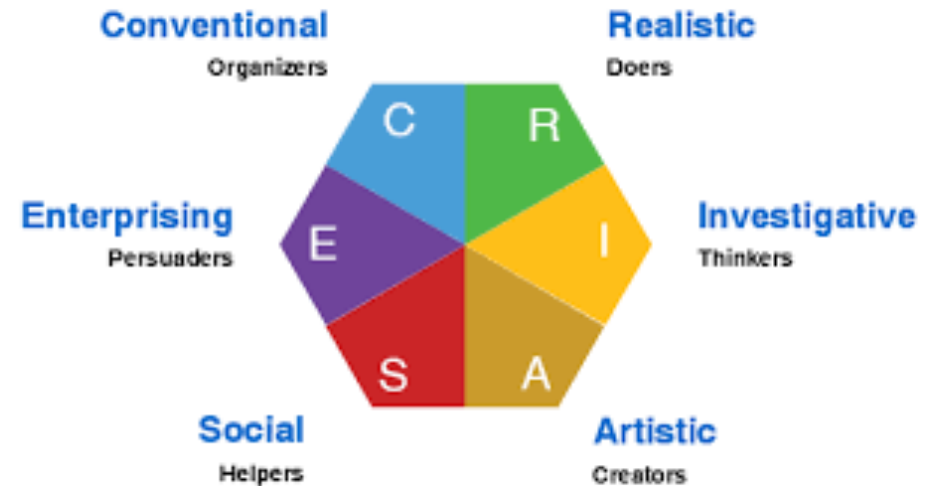


Data Collection - Person

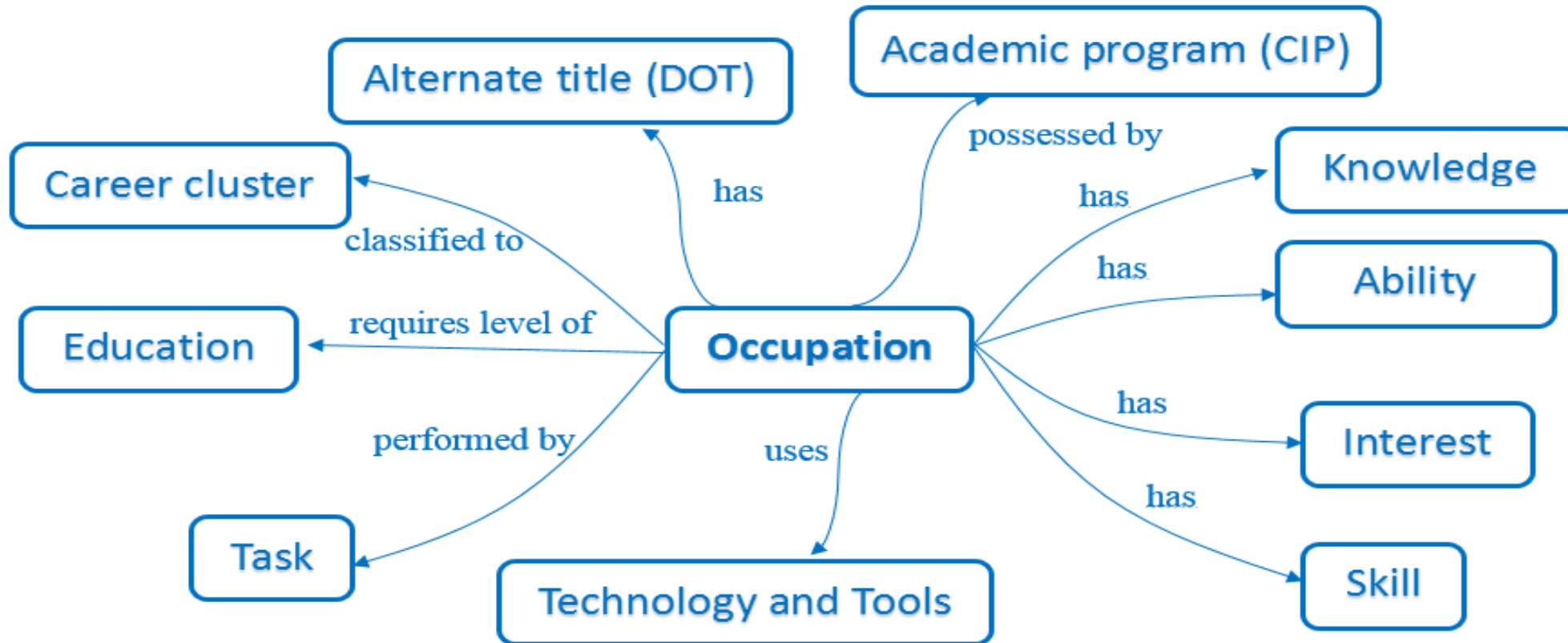
- **Vocational Interest:**

- **Holland code model**

- (**RIASEC** - Holland, 1985a, b, 1997; Holland & Gottfredson, 1992)
- It has the **106-item** with responses made on a **5-point scale** (strongly dislike to strongly like). There are **six types**.
- All occupations can be classified according to their alignment with these six types.



Data Collection - Occupation



How to connect and measure the correlation between student and occupation?

- Vocational Interest (Holland code model) RIASEC - 6 factors

Occupation description		Range: 1-7						Student Interest from Questionnaire						
ocode	title	r	i	a	s	e	c	studentID	r	i	a	s	e	c
11-1011.00	Chief Executives	1.33	2.00	2.67	3.67	7.00	5.33	669	3.74	2.81	4.06	4.10	3.63	3.67
11-1011.03	Chief Sustainability Officers	1.00	4.33	2.67	2.33	7.00	4.33	673	3.32	3.69	3.18	3.60	2.95	3.73
11-1021.00	General and Operations Managers	1.33	1.33	1.00	3.33	7.00	3.67	677	3.37	3.88	3.76	3.75	3.21	3.53
11-1031.00	Legislators	1.00	3.67	3.67	4.67	7.00	3.00	681	3.95	3.69	3.41	3.90	3.53	3.73
11-2011.00	Advertising and Promotions Managers	1.67	2.00	5.33	2.33	7.00	4.67	685	3.53	3.81	3.94	4.40	3.42	3.73
11-2011.01	Green Marketers	1.00	5.33	4.33	2.33	5.33	3.00	689	2.68	2.88	3.82	3.40	3.47	2.73
11-2021.00	Marketing Managers	1.00	2.33	3.67	2.67	7.00	5.33	693	2.74	3.69	4.29	4.10	3.79	3.60
11-2022.00	Sales Managers	3.00	2.00	2.00	3.67	7.00	4.67	697	3.16	2.56	3.24	3.50	2.79	3.67
11-2031.00	Public Relations and Fundraising Managers	1.33	1.33	5.00	3.67	7.00	3.67	701	4.53	4.63	4.65	4.85	4.05	4.40
11-3011.00	Administrative Services Managers	2.00	2.33	1.00	2.67	7.00	5.33	705	3.63	4.50	3.88	4.50	4.00	3.27
11-3021.00	Computer and Information Systems Managers	4.00	4.33	1.67	1.67	6.67	5.33	709	3.79	4.00	4.12	4.40	4.00	3.67
11-3031.01	Treasurers and Controllers	1.67	2.67	1.00	2.67	6.00	7.00	713	3.53	3.88	4.06	4.45	4.00	4.27
								717	3.37	3.63	3.47	3.60	2.84	3.67
								721	3.95	3.69	3.47	3.80	3.47	3.47

Result (1/5)

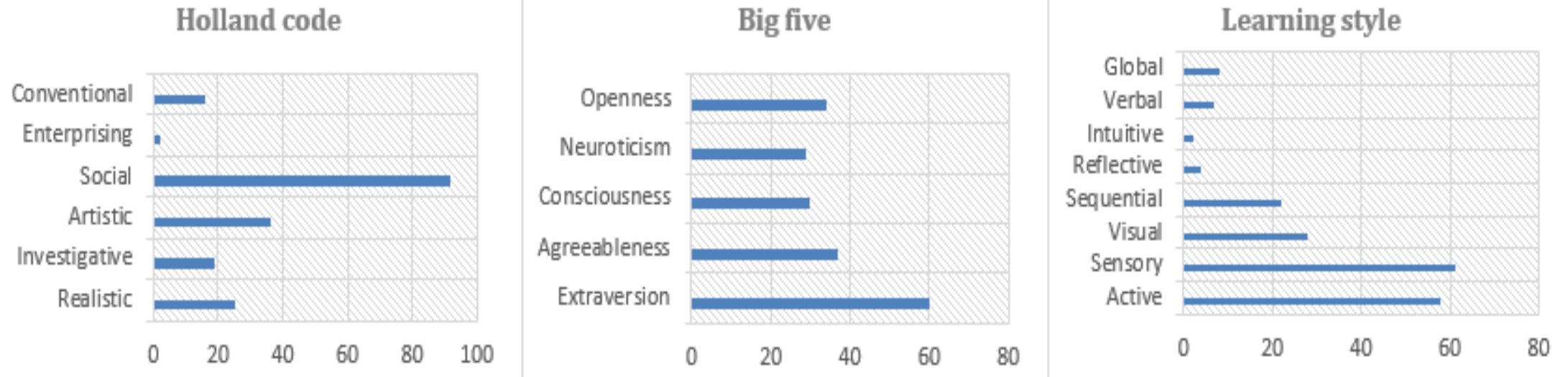


Figure 17 Frequencies of students on Vocational interests, Big five personalities, and Learning styles

Result (2/5)

Table 9 Pearson's correlations between suggested, intended, predicted, and recommended majors

Majors	1.	2.	3.	4.
1. Suggested major by parents	1			
2. Intended major of students	.362**	1		
3. Predicted major	-.020	.164*	1	
4. Recommended major by Holland model	.086	.141	.228**	1

* $p < .05$. ** $p < .01$.

Result (3/5)

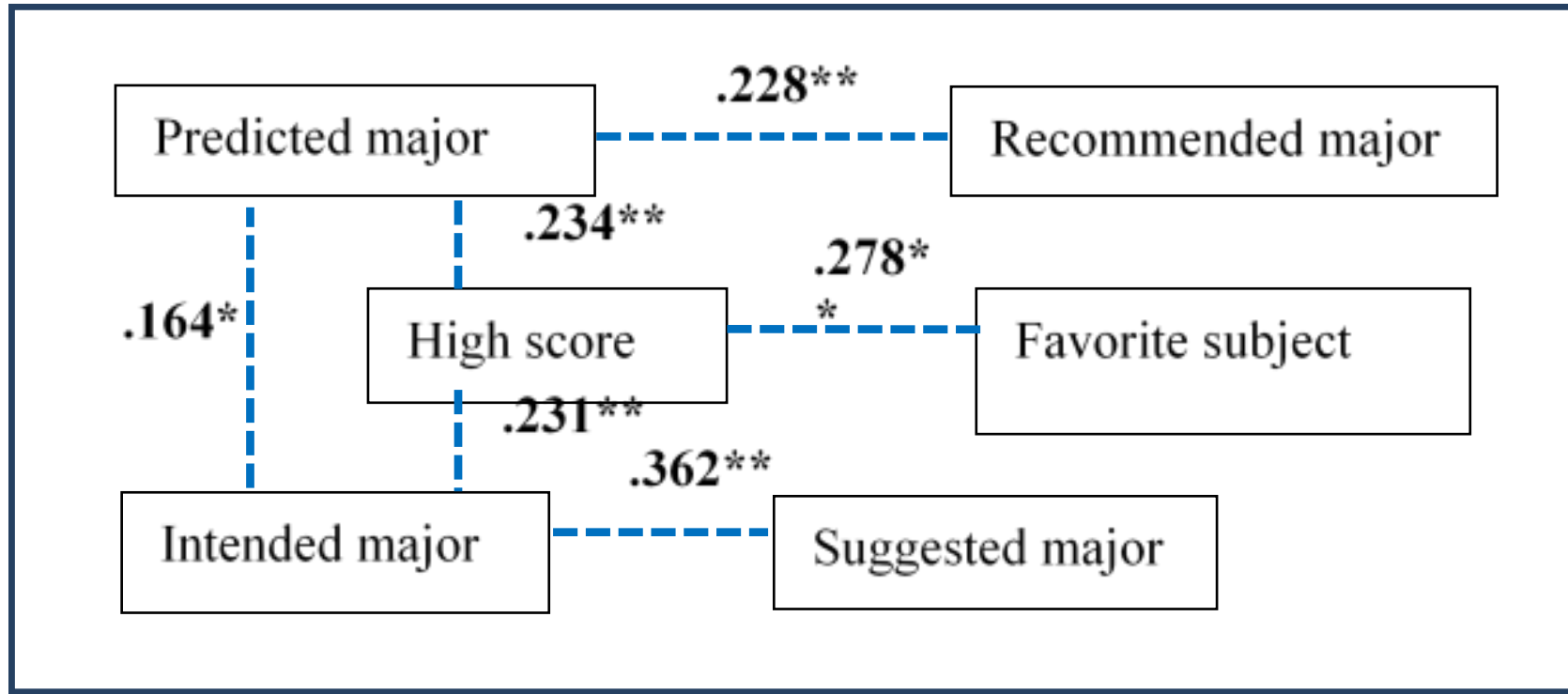


Figure 19 Correlations between high score, favorite subject and four kinds of majors. Dashed lines indicate the correlations. (*p < .05, **p < .01)

Result (4/5)

Table 10 Influence of the choice of the major

Questions of the influence of the major choice of the student	Yes	No	Not filled
Will you discuss with your parents when you make a critical decision?	133 (70%)	20 (10.5%)	37(19.5%)
Will you follow the suggestion of your parent, if your parent asks you to choose their suggested major, even you don't like?	30 (15.5%)	121 (64%)	39 (20.5%)
Why did you choose this intended major?			37 (19.4%)
According to my interest	130		
Easily finding a job	41		
Family (specially parent) suggested	38		
Nowadays this major is very popular	32		
High salary	21		
Teacher suggested	3		
My friends are going to choose	0		

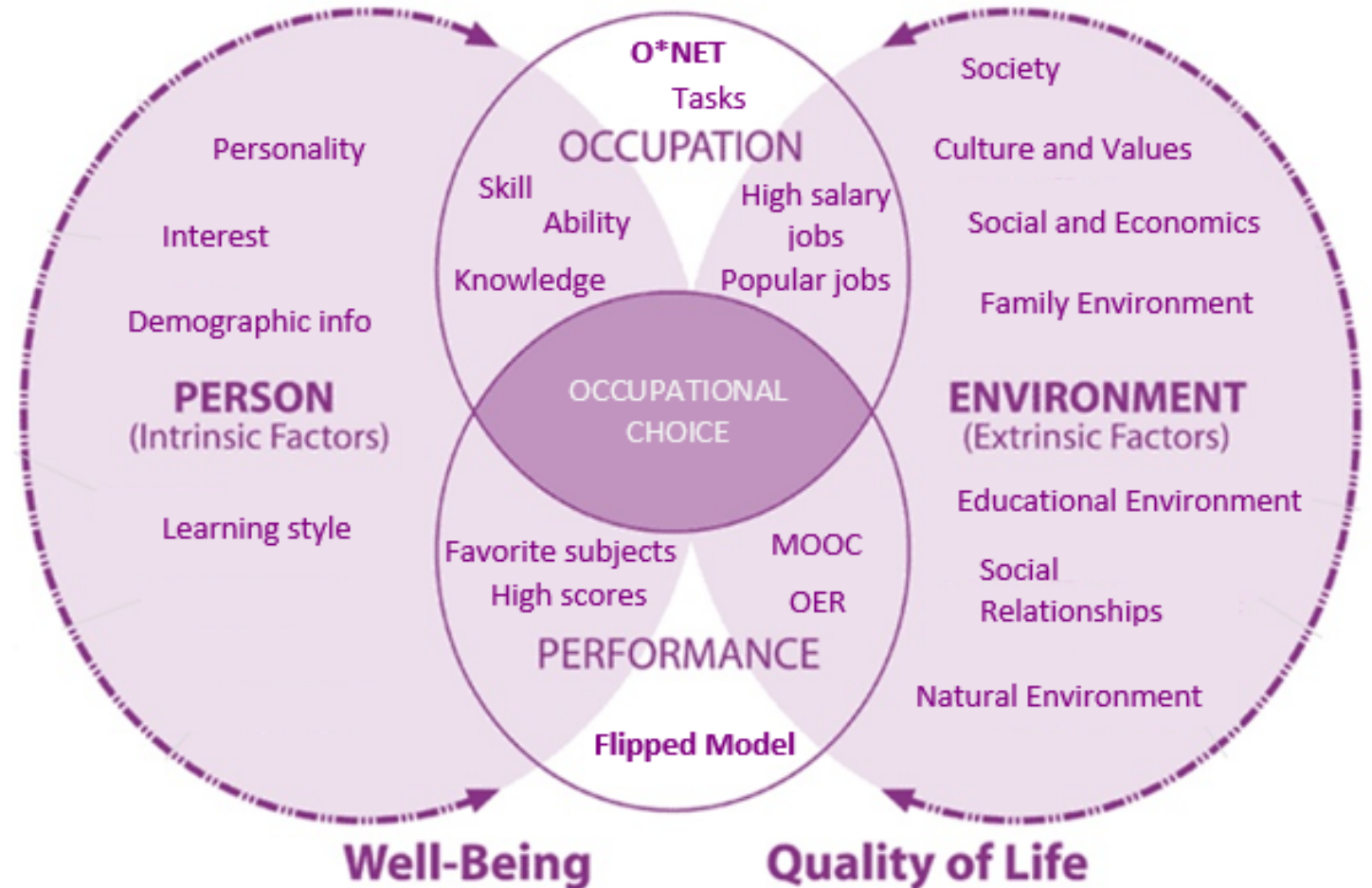
Result (5/5)

Method	R	I	A	S	E	C	Total
Regression Model	0.22	0.12	0.06	0.08	0.13	0.11	0.1215
HC Model	0.17	0.13	0.11	0.09	0.14	0.12	0.1261

Table 5: Influence of the choice of the major. Lower value is better result. RMSE is used.

Future Works

- Open Government Data
- MOOC
- Open Educational Resource
 - Wikipedia
- Linked Open Data





Thank you!