

# Student perceptions of privacy principles for learning analytics

2016. 11. 23. (수) iLED 세미나  
발표자: 송아름이

# CONTENTS

I . introduction

II . 학습분석을 위한 정보보호 원칙

III . method

VI . results

V . discussion

# Introduction

- ▶ 방대한 교육학적 정보는 학습과 교수를 강화하고 관리상의 의사결정을 improve하는데 기회를 제공
- ▶ 고등교육분야에서 학습분석의 채택은 빠르게 증가하고 있으나 ethical and privacy issues가 주요 고려사항으로 대두되고 있음
- ▶ 학습분석은 투명하고 편견 없는 정보의 수집, 사용, 분석을 포함하여 조직의 원칙과 가치에 정렬되어야 하고, 모든 이해관계자에게 beneficial
- ▶ 현재의 학습분석에서의 privacy와 관련된 연구들은 인터넷 보안과 의료 환경과 같은 다른 분야의 가이드라인을 언급하고 있음

**이 탐험적 연구는 학습분석과 관련된 정보보호 원칙과 관련된 학생들의 인식을 조사함으로써 이러한 research gap을 다룸**

# privacy principles for learning analytics

- ▶ 학습자들이 사적인 정보를 보호하려는 이유
  - 경쟁적 이유
  - 학습자가 자신의 정보를 공유하고 싶지 않기 때문(country-specific)
- ▶ 발전된 테크놀로지와 학습분석 시스템은 고등교육기관이 다양한 맥락에서 실시간으로 데이터를 수집하는 것을 가능케 함
  - 통계와 역동적인 교육적 정보의 방대한 양은 학습 프로세스, 학습환경, 교육학적 의사결정을 실시간으로 모델링하고 예측하고 최적화 하는데 이용
  - 학생들은 그들의 우편번호나 이전의 학문적 수행과 같이 행정맥락에서 드러냈던 데이터를 학습환경맥락에서 드러내는 것 (또는 그 반대로도) 을 원치 않을 것
- ▶ 누가 어떤 데이터에 접근하는지, 어디에 그리고 얼마나 오래 그 데이터를 저장할건지, 그리고 어떤 절차들과 알고리즘이 가능한 데이터의 further한 사용을 위해 실행되는지를 정의할 필요

# Research question and hypotheses

## Hypothesis1.

학생들은 학습에 대한 광범위하고 다양한 지원을 제공하는 학습분석시스템을 좋아할 것이다.

## Hypothesis2.

학생들은 학습분석시스템에서 데이터를 공유하는데 보수적일 것이다.

## Hypothesis3.

학생들의 데이터 공유의 상세한 수준(3a)과 데이터 통제(3b)는 학습분석시스템을 이용하는데 있어 선호도와 관련이 있을 것이다.

# Methods

- ▶ one-group quasi-experimental design(준실험설계)
- ▶ Online laboratory study implemented on the university's sever and conducted in June 2015
- ▶ N = 333 valid responses(223female, 107male). The average age of the participants was 22.75 years
- ▶ German University's bachelor's program(80%), masters program(20%)

# Learning analytics system

- ▶ The first example was based on the Course Signals project and included simple visual aids.
  - 디자인 원리는 자기조절학습(self-regulated learning)의 촉진에 근거
  - 학습자들의 정보는 세가지 시각적 신호로 표시(빨강, 노랑, 초록)
- ▶ The Second example included a dashboard showing general Information
  - 성찰적 사고(reflective thinking)가 핵심 디자인 원리
  - 학습자들의 정보는 그래프와 차트로 표현
- ▶ The third example provides detailed insights into learning and performance
  - 디자인 원리로 다양한 요소들 사용(Prompting, metacognitive awareness, adaptive and personalized feedback)
  - 학습자들의 정보는 텍스트, 그래프, 아이콘, 개인화된 prompt

# Instruments

- ▶ Control over data (COD)scale - 5점 Likert 척도
  - 학습분석시스템에서 데이터의 접근, 통제, 사용에 초점
- ▶ Sharing of data(SOD) questionnaire - Thurstone 척도(동의/비동의)
  - 학습분석시스템에서 개인의 구체적인 정보의 공유와 관련
- ▶ Acceptance and use of learning analytics system
  - 수용성과 사용에 있어서 세가지 학습분석 시스템의 등수를 결정
  - “그 학습분석시스템이 나의 학습을 촉진했다”,  
“그 학습분석시스템은 사용하기에 너무 복잡했다”
- ▶ Demographic information
  - 나이, 성별, 대학입학자격, 전공, 현재 수강과목 등등

# Procedures

- ▶ 2016년 6월, 2주 동안 학생들은 실험연구에 참여하도록 초대받았다. 위에 언급된 모든 도구들은 온라인 플랫폼(세가지 예시의 학습분석 시스템 또한 포함한)으로서 실행되었다.
- ▶ 우선, 참가자들은 디지털 대학 시스템에서 학습분석과 개인적 데이터의 사용에 대한 일반적 정보를 받는다.
- ▶ 그리고 참가자들은 세가지 다른 학습분석 시스템에 직면하게 된다. 그들에게 짧은 시간 시스템에 익숙하게 한 후, 각각의 학습분석시스템에서 수용성과 학습에 대해 기대되는 이용에 대해 세가지 다른 학습분석 시스템의 등수와 비교를 묻는다.
- ▶ 그리고 참가자들은 데이터스케일의 통제를 완료하고, 데이터 공유에 대한 질문, 마지막으로 참가자들은 그들의 인구통계학정보를 적는다

# Results

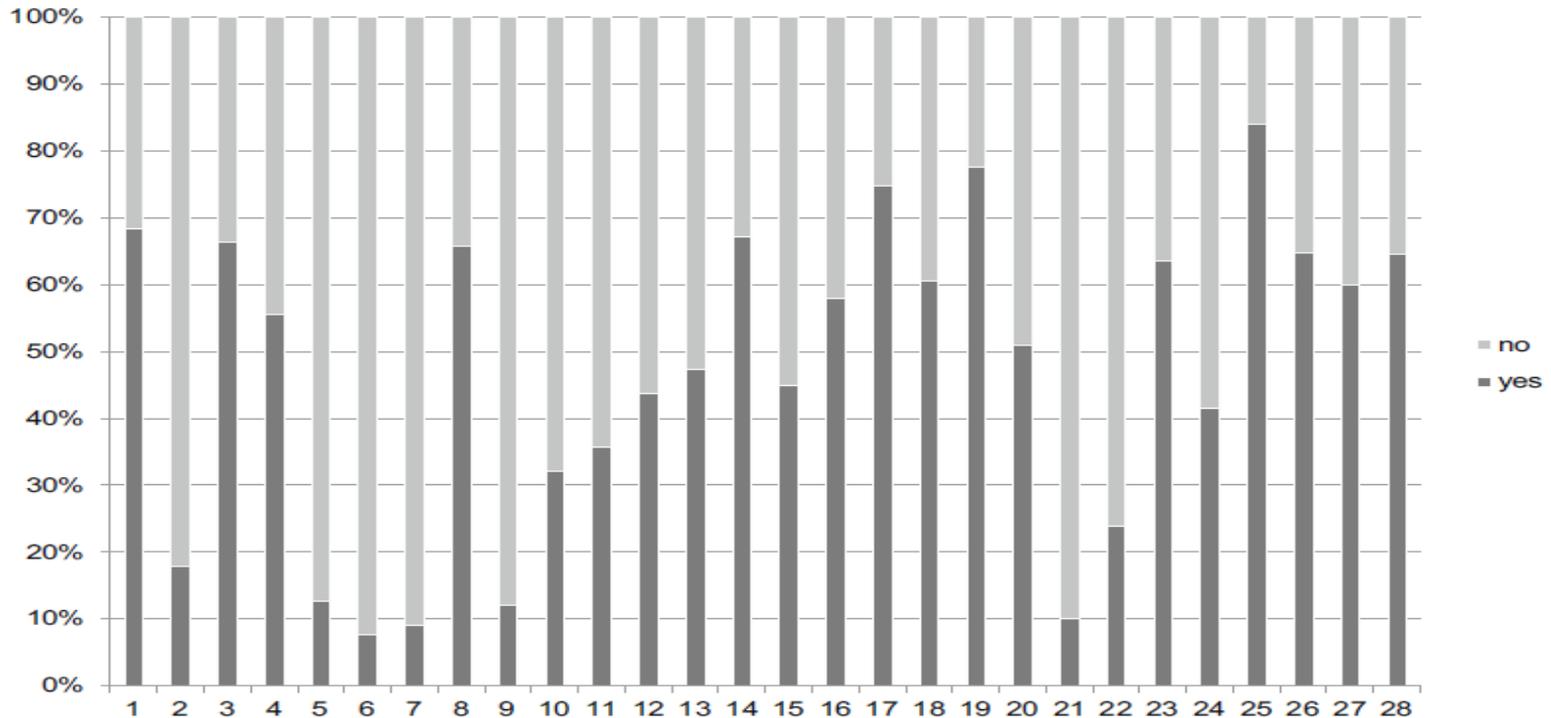
## Hypothesis1: 학습분석시스템의 사용

- 세 학습분석시스템의 사용에 대한 학생들의 평가는 상당한 차이가 있었다.
- example3이 example2보다 상당히 높게 평가되었다.
- example3이 example1보다 상당히 높게 평가되었다.

학생들은 학습에 광범위하고 다양한 지원을 제공하는 학습분석시스템(example3)을 다른 시스템들(example) 보다 선호한다.

# Results

## Hypothesis2: 학습분석시스템에서의 데이터 공유



**Fig. 1** Sharing of data in learning analytics systems (1 name, 2 address, 3 email, 4 date of birth, 5 marital status, 6 medical data, 7 income, 8 prior knowledge, 9 user path, 10 online times, 11 downloads, 12 course-specific discussion activity, 13 semantic analysis of posts, 14 test scores, 15 higher education entrance qualification grade, 16 school history, 17 motivation test results, 18 interest test results, 19 learning strategies test results, 20 intelligence quotient, 21 externally produced data, e.g., social media, 22 parents' educational level, 23 academic achievements, 24 occupation other than university studies, 25 course enrollment, 26 self-test scores, 27 general discussion activity, 28 library activity statistics)

# Results

## Hypothesis2: 학습분석시스템에서의 데이터 공유

- 많은 학생들이 학습분석을 목적으로 그들의 코스등록데이터(84%), 학습전략 테스트결과(78%), 동기 테스트 결과(75%)를 공유하는 것에 동의하였다.
- 반대로, 학생들은 그들의 의료 데이터(92%), 수입(91%), 외부적으로 생산된 정보들인 소셜 미디어(90%), 결혼유무(87%)등은 공유하는 것을 원치 않았다.
- 학생들은 시험점수나 코스등록과 같은 학교와 관련된 정도는 공유하지만, 개인정보와 온라인 상 행동의 trail(user path, online times, downloads, etc.)은 공유하는 것에 동의하지 않았다.
- 데이터 공유에 있어서 세가지 학습분석시스템에서의 차이점을 검증한 결과, 차이가 있었으며 추가적인 검증을 통해 참가자들은 2번째 시스템에서 1번, 3번 시스템보다 상당히 더 많은 데이터를 공유하고자 했다.

**학생들은 학습분석시스템을 지원하기 위해 데이터를 공유하는데 있어서 보수적이다.**

# Results

## Hypothesis3: 정보공유와 학습분석시스템 사용 간의 관계

- 세가지 학습분석시스템에서 참가자들의 데이터에 대한 통제(COD)는 세가지 학습분석시스템의 수용성과 기대된 사용(ALA)과 정적인 상관을 가진다.
- 추가적으로, 참가자의 데이터에 대한 통제(COD)는 세가지 학습분석시스템을 위해 데이터를 기꺼이 공유하고자 하는것(SOD)과 정적인 상관을 가진다.
- 마지막으로, 학습분석시스템의 수용성과 기대되는 사용(ALA)과 학습분석시스템을 위해 데이터를 공유하고자 하는 것(SOD)사이에는 정적상관이 있다.

# Results

## Hypothesis3: 정보공유와 학습분석시스템 사용 간의 관계

- 데이터에 대한 통제(COD)와 정보 공유(SOD)가 세가지 학습분석시스템의 수용성과 사용(ALA)을 상당히 예측하는지 회귀분석을 실시한 결과,
- 데이터에 대한 통제(COD)의 중요성과 데이터를 공유에 대한 willingness(SOD)는 학습분석시스템의 수용성과 사용(ALA)를 긍정적으로 예측하였다.

학생들의 데이터 공유의 상세한 수준(3a)과 데이터 통제(3b)는 학습분석시스템의 수용과 기대되는 사용에 관련이 있다.

# Discussion

- ▶ 참가자들은 학습분석 시스템이 학습을 위해 폭넓고 다양한 지원을 제공해야 함을 지시: 자기평가, 역동적인 내용 반응성, 시각적 신호, 사회적 상호작용을 위한 제언, 개인화된 학습활동, 예측적 코스이수, 적응적 스캐폴드 등
- ▶ 학습자들은 모든 데이터를 공유하는 것을 원치 않으나. 만약 학습분석시스템이 풍부하고 의미 있는 정보를 제공한다면, 더 많은 데이터를 공유하는 것에 열려있음
- ▶ 학습분석 데이터의 관리와 보호에 대한 가이드라인은 학습분석시스템의 신뢰를 구축하고 높일 수 있는 중요한 매체(데이터를 공유에 대한 동의를 얻는 것에 대한 전략들 실행될 필요)
- ▶ 누가 데이터에 대한 접근권한을 가지는지, 어디에 얼마나 오랫동안 데이터가 저장되는지, 어떻게 분석 및 추론이 수행되는지, 학생들이 데이터가 수집되는 것을 아는지 등에 대한 질문은 차후 연구에서 논의될 필요

# Implications

- ▶ 학생들의 데이터를 공개하는 것을 꺼리는 경향은 학습분석시스템의 사용에 있어서 제한점을 낳는다. (학습분석시스템은 풍부한 정보와 관련돼 있기에 중요한 정보의 부족은 biased되거나 useless한 결과를 초래)
- ▶ 학습분석시스템의 수용성을 높이기 위해, 모든 이해관계자들이 가능한 빨리 개입하는 것이 중요
- ▶ 학습분석시스템의 수용성과 관련해서 투명성은 가장 중요한 요인으로 고려 되어야 함  
: 기본적 정보(기본알고리즘, 로우데이터, 저장, 액세스 등) 를 제공하는 question button, FAQ, 경고시스템 제공
- ▶ 고등교육기관은 이해관계자들에게 학습분석시스템이 복잡한 결정을 요구한다는 것을 알리기 위해 모델 (fig.2.) 을 이용해야 함

# Implications

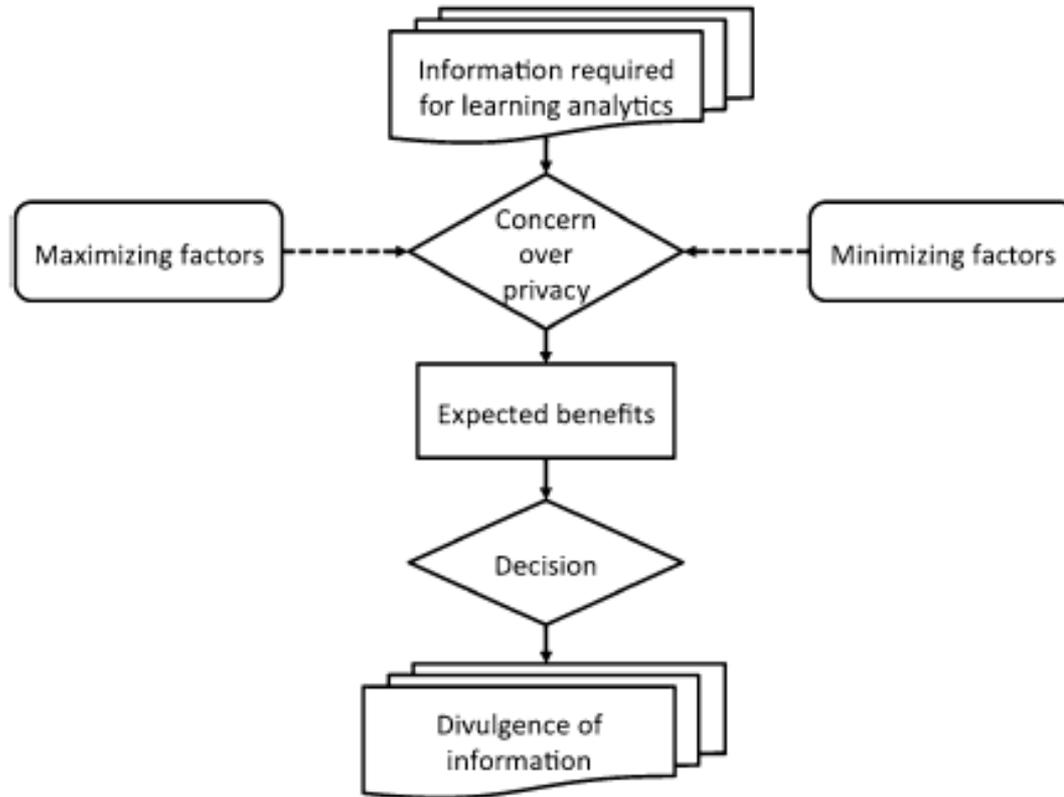


Fig. 2 Deliberation process for sharing information for learning analytics systems

- ▶ 학생들은 그들의 걱정을 학습분석시스템에 요구되는 구체적인 정보에 근거하여 평가
- ▶ 리스크를 최소화하는 요인: 학습분석시스템 또는 기관에 대한 신뢰, 스스로 관리를 할 수 있는 데이터에 대한 통제
- ▶ 리스크를 최대화하는 요인: 불투명성, 학습분석시스템이나 기관에 대한 부정적 평판

# Limitation and future work

- ▶ 결과는 진단도구와 수집 데이터에 정직하고 정확한 응답을 가정한 self-report data에 의존, bias 또는 응답자의 경향성을 포함할 가능성 존재
- ▶ 한 기관에서만 샘플 추출, 윤리와 프라이버시는 다른 문화점 관점에 의존하기므로 미래연구는 국제적 비교가 가능한 다 기관에 걸쳐 수행 필요
- ▶ 참가자들은 학습분석시스템의 이용한 사전경험이 없으며 제한된 시간에 linear하게 오직 세가지 학습분석 시스템에만 노출. 미래 연구는 학습분석 시스템을 경험한 학습자들에게 다른 학습분석시스템을 empirical vary한 순서로 제시할 필요
- ▶ 미래연구는 다양한 이해관계자(teacher, tutor, learning designer, 를 포함하여 윤리적이고 프라이버시의 관점에서 질적 요소를 포함하여 mixed-method하게 설계하여 수행 되어야 함

# Conclusion

- ▶ 학습분석을 통한 personalized, dynamic, timely한 피드백의 유용성은 학습자의 자기조절 학습을 지원하며 동기와 성공을 증가시킬 수 있으나 자동화된 시스템은 학습자의 비판적 사고와 자율적인 학습을 방해
- ▶ 접근을 통제하는 절차와 교육학적 데이터 사용의 요구는 학습분석이 이뤄지기 전에 수행되어야 하며 익명화된 개인 정보의 저장과 처리는 아주 적은 단계들로 이루어져야 함
- ▶ 학습분석은 개인 정보와 개인의 학습 기록에 관한 통찰을 드러낼 수 있으나 그들은 편견 없는 종합적이고 유효한 것으로 간주되지는 않음

**QNA**



**THANKS A LOT**