Supporting Adult Learners through Tutoring Groups

HUDK 4050: Educational Data Mining

Pooja Addla, Nikita Tejwani, Kan Yamane

Contextual Overview

◎ Setting: Community Impact's High School Equivalency Program
◎ Goal of the program
  ○ Support adult literacy
  ○ Prepare non-traditional students to pass the GED
  ○ Enable students to attain better socioeconomic status

Community Impact
AT COLUMBIA UNIVERSITY

TABE Overview

◎ Student intake involves diagnostic exam: TABE (Test of Adult Basic Education)
- ○ Subjects: Reading and Mathematics
- ○ Levels: Easy, Medium, Difficult, Advanced
- ○ TABE reports help us determine class placement and readiness for the GED exam

Sample TABE Report Pt. 1

DATA RECOGNITION
DRC
CORPORATION

Individual Profile: STUDENT, SAMPLE

Report Criteria

ID:	12345	State:	
Test Name:	TABE 11 ALL	District:	SAMPLE DISTRICT
Report:	ALL	School:	SAMPLE SCHOOL
Report Date:	10-18-2019		

Test Results	Test Date	Level	Number of Points		Items Attempted	Scale Score	SEM	NRS Level	MSG
			Total	Obtained					
Reading	10/25/2018	M	47	44	40	575+	52	4	Y
Mathematics	10/26/2018	M	39	31	35	570	20	4	Y
Language	10/26/2018	M	39	30	35	552	19	4	N

Sample TABE Report Pt. 2

Performance on Domains	Number of Items	Number of Points		Performance Category		
		Total	Obtained	Non-Proficiency	Partial Proficiency	Proficiency
Reading						
Key Ideas and Details	18	18	17			✓
Craft and Structure	17	20	19			✓
Integration of Knowledge and Ideas	5	9	8			✓
Mathematics						
Measurement and Data	6	6	5			✓
Numbers and Operations - Fractions	7	7	5		✓	
Numbers and Operations - Base Ten	5	6	5		✓	
Operations and Algebraic Thinking	4	5	5			✓
Geometry	4	6	5			✓
Expressions and Equations	4	4	2		✓	

FORM	DOMAIN	PERFORMANCE	DEMONSTRATED SKILLS	AREAS FOR NEXT FOCUS
M	**Reading**			
	Key Ideas and Details	Proficiency	• Summarize an informational text • Identify the main idea of a very complex text • Make an inference connecting ideas in a text • Make an inference based on a section of text • Use details to support inferences regarding connections in a text	• Support an inferred central idea with evidence from a moderately complex text • Make distinctions between ideas in a moderately complex text • Analyze the connections among ideas in a very complex text • Use evidence to support the analysis of connections among ideas in a very complex text
	Craft and Structure	Proficiency	• Describe the structure of multiple paragraphs of a moderately complex text • Describe the structure of a section of very complex text (technical document) • Describe the structure of very complex text • Use evidence to support comparisons of the structure of moderately to very complex texts • Use evidence to support the comparison of points of view across multiple texts on the same topic	• Analyze the function of a text feature • Determine the point of view of a moderately-very complex text • Use evidence to support the determination point of view • Determine authors' purposes across complex texts
		Proficiency	• Explain the author's use of evidence to support an idea in very complex text • Use evidence to support the explanation of an author's point about a text • Use text evidence from multiple texts to support the use of quantitative information • Use multiple pieces of evidence to support the explanation of an author's point about a text	• Determine the relevance of evidence use to support a claim • Compare claims made across texts and support comparison with multiple pieces of evidence
M	**Mathematics**			
	Measurement and Data	Proficiency	• Use properties of complementary and supplementary angles to find missing angle measures in diagrams • Find the missing dimension of a rectangular prism when given the other dimensions and the volume • Find volumes of rectangular prisms by counting unit cubes and by multiplying the side lengths (using the volume formula) • Use line plots to solve multi-step addition, subtraction, multiplication, and division problems	• Use the properties of angles to write and solve equations in one variable to find missing angle measures in diagrams • Create rectangular prisms with different dimensions and volumes that are the same • Convert from a larger unit of measure to a smaller unit of measure

Practical Implications of TABE

How can we use this data to:

- inform design of interventions to improve their performance in the GED?

- better support these learners?

RESEARCH QUESTION

What are the "optimal" number of classes/tutoring groups?
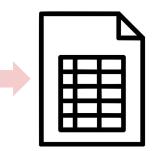
Data Pre-Processing Overview

Original Report (PDF Format)		Text File		Data Table

Data Extraction Overview

| Original Report (PDF Format) | Text File | Data Table |

Data Extraction - Steps

Score reports are downloaded as pdfs

1. Python script used to **extract text** from a folder called 'Students' containing pdf reports
2. Student **names were removed** per CI policy
3. **Spaces were added** between words
4. **Reports saved** as .txt files without names in a folder called 'NamesRedacted'

```python
student_num = 0

for filename in os.listdir('Students'):

    #Open a single pdf file with TABE results and extract text
    path = 'Students/' + filename
    pdfFileObj = open(path,'rb')
    pdfReader = PyPDF2.PdfFileReader(pdfFileObj)
    text = ''
    for i in range(0, pdfReader.numPages):
        text += pdfReader.getPage(i).extractText()

    #Add spaces where they are needed
    new_text = ''
    for i, letter in enumerate(text):
        if i and (letter.isupper()) and (not(text[i-1].isspace())):
            new_text += ' '
        new_text += letter
    has_name = True

    #Remove student names, found after the words 'Individual Profile'
    while has_name:
        index = new_text.find('Individual Profile')
        if index == -1:
            has_name = False
        else:
            new_text = new_text[:index] + new_text[index + 45:]

    #Write a new pdf file with the student data, minus the student name
    name = 'NamesRedacted/Student' + str(student_num)
    FileObject = open(name, 'w')
    FileObject.write(new_text)
    student_num = student_num + 1
```

Data Transformation Overview

Original Report (PDF Format) → **Text File** → **Data Table**

Variables of Interest

Test Results	Test Date	Level	Number of Points		Items Attempted	Scale Score	SEM	NRS Level	MSG
			Total	Obtained					
Reading	10/25/2018	M	47	44	40	575+	52	4	Y
Mathematics	10/26/2018	M	39	31	35	570	20	4	Y
Language	10/26/2018	M	39	30	35	552	19	4	N

Performance on Domains	Number of Items	Number of Points		Performance Category		
		Total	Obtained	Non-Proficiency	Partial Proficiency	Proficiency
Reading						
Key Ideas and Details	18	18	17			✓
Craft and Structure	17	20	19			✓
Integration of Knowledge and Ideas	5	9	8			✓
Mathematics						
Measurement and Data	6	6	5			✓
Numbers and Operations - Fractions	7	7	5		✓	
Numbers and Operations - Base Ten	5	6	5		✓	
Operations and Algebraic Thinking	4	5	5			✓
Geometry	4	6	5			✓
Expressions and Equations	4	4	2		✓	

Variables of Interest

FORM	DOMAIN	PERFORMANCE	DEMONSTRATED SKILLS	AREAS FOR NEXT FOCUS
M	**Reading**			
	Key Ideas and Details	Proficiency	• Summarize an informational text • Identify the main idea of a very complex text • Make an inference connecting ideas in a text • Make an inference based on a section of text • Use details to support inferences regarding connections in a text	• Support an inferred central idea with evidence from a moderately complex text • Make distinctions between ideas in a moderately complex text • Analyze the connections among ideas in a very complex text • Use evidence to support the analysis of connections among ideas in a very complex text

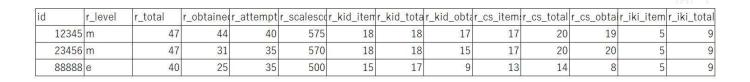# Converting the Text to a Table

Total Obtained Reading10/15/2021 M472240485122 N Mathematics10/15/2021 E351635448121
:ails to support key ideasl Explain sequence of an event in a moderatelycomplex textl I(
inmoderately complex textl Compare information expressed in multipleformatsl Use evide
of measure for agiven objectl Estimate the length of an object beforemeasuring the obj(

Extract by finding position of the string "Reading"

"M472240485122"

Parse each score by position

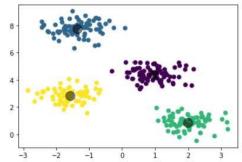| r_level	r_total	r_obtained	r_attempted	r_scalescore
m	47	22	40	485

id	r_level	r_total	r_obtained	r_attempt	r_scalesc	r_kid_item	r_kid_tota	r_kid_obta	r_cs_item	r_cs_total	r_cs_obta	r_iki_item	r_iki_total
12345	m	47	44	40	575	18	18	17	17	20	19	5	9
23456	m	47	31	35	570	18	18	15	17	20	20	5	9
88888	e	40	25	35	500	15	17	9	13	14	8	5	9

⋮

Data Processing

Group similar students together

◎ ICAP Framework (Chi, 2014) asserts that joint dialogue improves learning

◎ Teachers can differentiate instruction more easily for delineated tutoring groups

How to identify groups of similar students

◎ K-means clustering
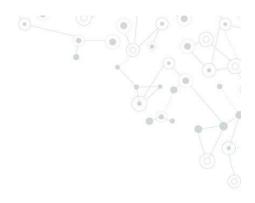
◎ Evaluation through sum within clusters distance

Suggestions for future research

1. What are the variables that are correlated with student's TABE score performances?

   a. Relationship between gender and TABE scores
   b. Relationship between age and TABE scores
   c. Relationship between socioeconomic status/race and TABE scores
   d. Relationship between years since dropping out and TABE scores (overall and subject-wise)
   e. Relationship between the number of years that the student has spent in the program and their TABE scores
   f. Relationship between the number of times the student has taken the TABE test and their latest scores

Suggestions for future research (contd.)

2. Can interventions targeted to support students based on their demographics (e.g.: support for single parents) improve TABE scores, and subsequently, GED scores?

3. Predicting GED performance (score or P/F) based on TABE scores. In the event they perform well in TABE but not so in GED, understanding why and exploring text anxiety.

Thank you.