SUPPORTING ADULT LEARNING THROUGH TUTORING GROUPS

Pooja Addla, Nikita Tejwani, and Kan Yamane

Department of Learning Analytics, Columbia University

HUDK 4050: Core Methods in Educational Data Mining

Professor Zhichun Liu

December 21, 2021

INTRODUCTION

The General Educational Development (GED) test is a high school equivalency diploma that was originally created by The American Council on Education during World War II to help veterans who were unable to complete high school to still earn credentials (Rose, 2013). All though there has been a consistent decline in the number of highschool dropouts since the late 70s (Stark & Noel, 2015, still, roughly 40 million Americans don't have a high school degree or GED certificate (Rose, 2013–especially marginalized populations such as racial minorities (Hispanics), immigrants and disabled adults (Ryan & Bauman, 2016). This data is devastating given that the high school degree and GED diploma are important credentials: in 2013, 73% of jobs required a high school diploma or equivalent (Torpey & Watson, 2014).

There are over 4000 Adult Education programs that exist in the US (Rose, 2013), but the focus of this paper is the GED preparation program at *Community Impact*. The program supports adult learners in developing literacy, numeracy, social studies skills, and science skills, which in turn prepare them to pass the GED. It is supported by the Workforce Innovation and Opportunity Act (WIOA) which aims to prepare Americans, especially those belonging to marginalized populations, to obtain and keep high quality employment. WIOA funds, which are essential to *Community Impact*'s work with adult learners, come with reporting requirements that enable WIOA's oversight entities to assess the effectiveness of a program.

The most common assessment for high school equivalency programs in New York is the TABE (Test of Adult Basic Education). The TABE exam includes a reading, language, and mathematics pre-test and post-test to be administered at the beginning and end of a learning period. Organizations are required to maintain records of all students' TABE scores, as well as detailed reports generated by the software on which the TABE is administered. *Community*

Impact and other high school equivalency programs tend to view this requirement as an inconvenience, printing out reports after students have tested and filing them away. While the WIOA testing requirement can be inconvenient, it also provides organizations with a wealth of data that could be used to design targeted interventions.

Therefore, in this paper, we will start by reporting findings from our secondary research on adult education to form a clear research question. Then, we outline the path for a research project that answers this question by sourcing relevant data, conducting analyses using Python, and making recommendations to *Community Impact*. These recommendations can be used to enhance the support that *Community Impact*, and adult education programs in general, are able to provide for their students. Data-driven recommendations in adult education programs such as *Community Impact* are pivotal not only because of the resulting socioeconomic wellbeing, but also because they lead to reduced crime and improved health (Rose, 2013).

LITERATURE REVIEW

In order to identify an intervention that will benefit adult students, we conducted an analysis of existing research on the challenges that are specific to adult students. Our review showed that older students have difficulty generating learning strategies and may not have feelings of belonging due to negative experiences in K-12 educational settings (Kotok et al., 2016). Subsequent research advocates for social scaffolding that can address both of these challenges.

Older adults do not self-generate effective learning strategies as successfully as younger adults as stated by the production deficit hypothesis (Craik & Byrd, 1982; Naveh-Benjamin et al., 2007; Perlmutter & Mitchell, 1982), which could be because of their restricted cognitive resources as compared to younger adults (Naveh-Benjamin et al., 2005; Shaw & Craik, 1989). This shows that most of them may need additional support and “social scaffolding” (Siegler & Alibali, 2020), which is a type of assistance used with children to provide them useful hints and nudges in the right direction, help them think about specific tasks and assist with problem solving models. By doing so and enabling adult learners to set clear goals, build on prior knowledge and leverage extrinsic motivators (such as better job prospects and income), tutoring can be an effective strategy for adult learners (Bailey & Card, 2009; Youde, 2018). In a study by Johnson et al. (2017), it was found that learners were able to expand on their initial answers and follow-up by asking meaningful questions when they interacted with tutors further proving that targeted tutoring is helpful for better learning.

Lastly, according to the literature, sense of belonging is strongly associated with retention rates (Kotok et al., 2016) and adult students from racial minorities often struggle with self-image and belonging, putting them at risk of dropping out. On the other hand, Pihlainen et al. (2021) states that peer support and individualized instruction can help increase intrinsic motivation for adult learners. In light of the above evidence, it is clear that grouping adult learners based on similarities (scores, demographics, etc.) can help with healthy cohort development, companionship and decrease achievement gaps.

PROBLEM STATEMENT

Existing research emphasizes the importance of peer support, which can exist in the form of tutoring groups. These groups can help with distortion of self-image due to issues that students may have experienced in previous education settings, while also permitting informal social scaffolding from peers. As a result, we ask: “What are the ‘optimal’ number of tutoring groups for students in GED classes at *Community Impact*?" Quantifying the number of tutoring groups needed can aid *Community Impact* in allocating resources on a group-by-group basis.

ANALYTICAL PLAN

The data source that we will be using in this analysis is the TABE reports of each student in the program. Below is the set of data within each report that will be used for this analysis.

1. The difficulty of the test that the student took for each subject (Reading, Mathematics, Language). Each subject test has four levels (Easy, Medium, Difficult, and Advanced), and the level of the test that the student took for each subject is noted on the report.
2. The overall scale scores for each subject. Student performance is quantified as a scale score, which takes into account individual performance and the difficulty of the test, based on other's performances.
3. Score by domain: Each subject is divided into smaller domains, and the performance of the student in each domain is also reported. For example, in the mathematics exam, students will receive individual scores for domains such as "Operations and Algebraic thinking" and "Geometry".
4. Proficiency in each domain: Based on the above scores, a student's proficiency in each domain is categorized as 'Non-Proficiency', 'Partial Proficiency', or 'Proficiency'.

Overall, the report gives a thorough overview of the student's performance in each subject and specific domain. The original report is provided in pdf format. Due to the report being provided in pdf format, thorough processing and cleaning is necessary before the actual analysis. A Python script will extract text from each .pdf file and subsequently clean the data, which is in a single string after extraction. The following steps are taken for each pdf: 1) Student names are removed from the text file, 2) Spaces are added after line breaks to improve human readability, and 3) The string is saved as a .txt file. While the data is in string form, it is possible to locate the variables of exam level, scale score, score for each domain, and proficiency for each domain. Figure 1 (in the Appendix) provides an example of how exam level and scale score can be found and transformed into a table.

Once we have finished extracting and cleaning the data, we can move onto the analytical process. Here, we will make use of the K-means clustering method combined with principal component analysis. In short, we will divide the students into similar clusters based on their academic performances. The variables of interest are: the exam levels for each subject, their performance for each subject (the scale score), score for each domain, and proficiency for each domain. To reduce the number of dimensions, we will apply the principal component analysis to first transform the data. After we have applied K-means clustering on the transformed data, we can also identify which variables are most relevant by investigating the eigenvalues of the principal components with highest explained variance.

HYPOTHESIZED RESULTS

Even without cluster analysis, there are a few patterns that have been observed among students at Community Impact. A few students each semester come after having completed high school outside of the United States. These students typically bond with each other quickly during classes and obtain their GED within a year. There are also a few students each semester who have very high scores in the Mathematics section of the TABE and very low scores in the Reading and Language sections. These students have difficulty with spoken English and have trouble progressing in their English and Math classes as a result. A majority of students, unlike these two groups, have low TABE scores in all sections of the TABE test. Many of these students make slow progress through the program and take longer than the expected two years. They have low attendance, limited interactions with their fellow students, and low homework completion rates.

We expect the cluster analysis to show one group of students with high scores in all TABE subjects and another group of students with only high Mathematics scores. These groups would represent students who have a high school diploma outside of the United States, and students with high math proficiency and limited English proficiency, respectively. We hope to gain additional insight into differences between the remaining students who have low scores in all TABE subjects. If subgroups are found within this group, we can perform informal qualitative analyses to identify reasons for these differences and corresponding support that can be provided.

DISCUSSION

Following the literature review and analysis plan, we aim to identify the optimal number of tutoring groups that will provide socio-emotional support as well as social scaffolding to students. Principal Component Analysis (PCA) can be used to identify variables that explain the most variance between students, and K-means clustering can be used to identify groups of students that have commonalities such as previous education experience and language proficiency. As a result of these analyses, students who share characteristics can be grouped together. Their teachers can create interventions for groups who are known to make slow progress through the program, such as those students who struggle with understanding spoken

English. The interventions, in combination with peer support, can help students in mastering concepts more quickly.

Additionally, being aware of these groups can be useful to *Community Impact* in considering the allocation of human and financial resources. *Community Impact* can prevent attrition resulting from socio-emotional factors and invest in necessary resources, such as additional tutors or supplemental textbooks. There are a few limitations to using TABE results to group students. Some dimensions of academic competency cannot be captured on the TABE test. Therefore, some groupings may contain students who need either significantly more or significantly less support than their group mates. We hope that in this event, the small size of *Community Impact*'s program will make it easy for teachers to identify the 'misplaced' student. We also anticipate that this event might provide opportunities for different forms of peer support (ex. the student who needs less support than their group mates can allocate some of their time to tutoring their group mates). Another limitation of the data is the fact that characteristics related to students' identities are not captured. To highlight demographic commonalities between students in a group, data can be taken from registration forms instead of TABE reports in order to capture demographic variables.

Finally, it is worth noting that this analysis is but a first step in the measures that can be taken to support organizations such as *Community Impact* through learning analytics. As mentioned above, a range of interesting questions can be asked by combining the current TABE data with other datasets, such as information on student demographics and amount of time spent in the *Community Impact* program. For example, it would be worthwhile to investigate the variables most correlated to student TABE performance; another interesting question would be to predict performance on the actual GED test based on prior TABE performance.

APPENDIX

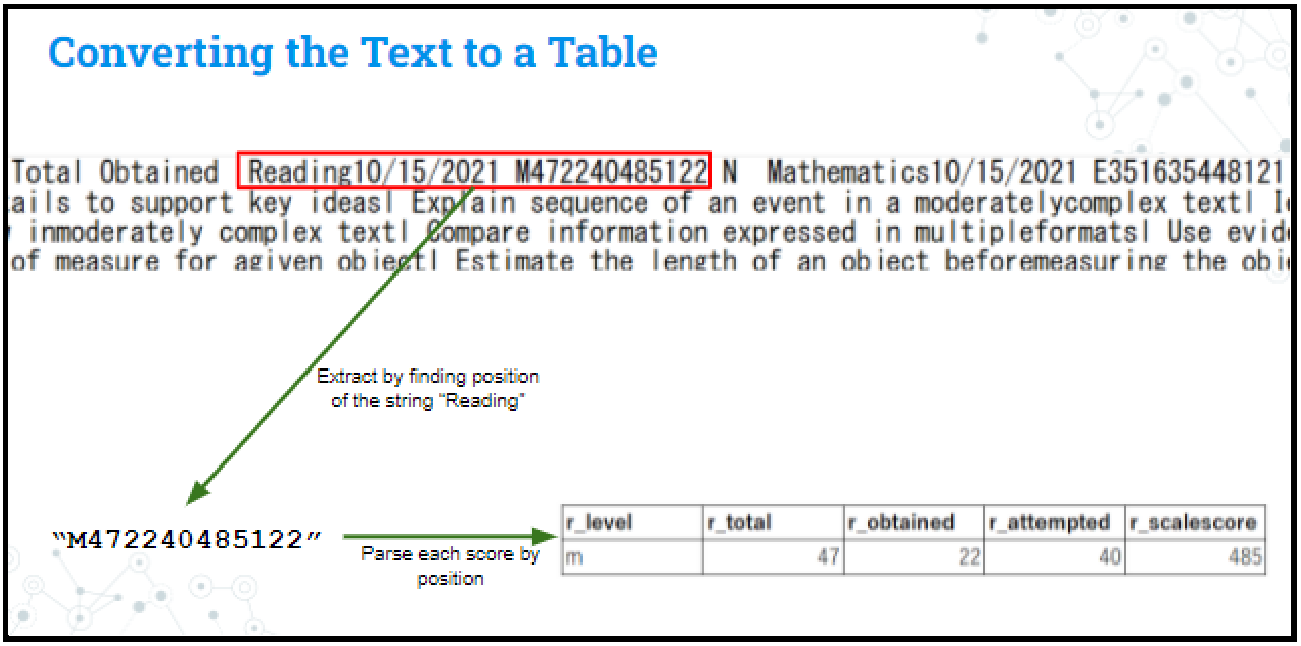

Figure 1: Transformation of substring to data table

REFERENCES

Bailey, C. J., & Card, K. A. (2009). Effective pedagogical practices for online teaching: Perception of experienced instructors. Internet and Higher Education, 12(3-4), 152–155. doi:10.1016/j.iheduc.2009.08 doi: 10.1016/j.iheduc.2009.08.002

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 191–211). New York, NY: Plenum Press.

Johnson, C.S. (2017). Collaborative technologies, higher order thinking and self-sufficient learning: A case study of adult learners. *Association for Learning Technology*. doi: http://dx.doi.org/10.25304/rlt.v25.1981

Kotok, S., Ikoma, S., & Bodovski, K. (2016). School climate and dropping out of school in the era of accountability. American Journal of Education, 122(4), 569–599.

Naveh-Benjamin, M., Brav, T. K., & Levy, O. (2007). The associative memory deficit of older adults: The role of strategy utilization. *Psychology and Aging*, 22,202–208. http://dx.doi.org/10.1037/0882-7974.22.1.202

Naveh-Benjamin, M., Craik, F. I. M., Guez, J., & Kreuger, S. (2005). Divided attention in younger and older adults: Effects of strategy and relatedness on memory performance and secondary task costs. *Journal Of Experimental Psychology: Learning, Memory, and Cognition*, 31, 520–537. http://dx.doi.org/10.1037/0278-7393.31.3.520

Perlmutter, M., & Mitchell, D. B. (1982). The appearance and disappear-ance of age differences in adult memory. In F. I. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 127–144). New York, NY: Plenum Press. http://dx.doi.org/10.1007/978-1-4684-4178-9_7

Rose, M. (2013). Second chances: The value of adult education and the GED. Phi Delta Kappan, 94(6), 45–49. https://doi-org.tc.idm.oclc.org/10.1177/003172171309400612

Ryan, C. L., & Bauman, K. (2016, March). Educational attainment in the United States: 2015 (Current Population Reports P20-578). *Washington, DC: U.S. Department of Commerce, Economics and Statistics Administration, Census Bureau*. Retrieved from https://www.census.gov/content/dam/Census/library/publications/2016/demo/p20-578.pdf

Shaw, R. J., & Craik, F. I. M. (1989). Age differences in predictions and performance on a cued recall task. *Psychology and Aging*, 4,131–135. http://dx.doi.org/10.1037/0882-7974.4.2.131

Siegler, R. S., & Alibali, M. W. (2020). Children's Thinking, 5th Edition, Upper Saddle River, NJ: Prentice Hall.

Stark, P., & Noel, A. M. (2015). Trends in high school dropout and completion rates in the United States: 1972–2012 (NCES 2015-015). *Washington, DC: U.S. Department of Education, National Center for Education Statistics*.

Torpey, E., & Watson, A. (2014). Education level and jobs: Opportunities by state. *Career Outlook*. Washington DC: U.S. Bureau of Labor Statistics. Retrieved from https://www.bls.gov/careeroutlook/2014/article/education-level-and-jobs.htm

Youde, A. (2018). Andragogy in blended learning contexts: Effective tutoring of adult learners studying part-time, vocationally relevant degrees at a distance. *International Journal of Lifelong Education*, *37*(2), 255–272. doi: 10.1080/02601370.2018.1450303