**VINCENT ALEVEN, HUMAN-COMPUTER INTERACTION INSTITUTE AND CARNEGIE LEARNING INC.**

Three possible projects; the first and third seem most likely to succeed. There is also likely to be involvement from either Steve Ritter or Steve Fonzali from Carnegie Learning. Data for all three projects is existing data from their DataShop facility, and/or data collected by Carnegie Learning Inc.

* **Investigate whether prerequisite relations among math topics can be detected in standard longitudinal log data from tutors.** Given a math "topic" or skill or unit, what math topics or skills or units do you need to master in order to (smoothly) learn that topic or skill or unit? The main idea would be to treat correlations between performance/learning on some tutor units, or on certain skills, as evidence of prerequisite relations.  Although this question may appear to be fairly well-defined, my hunch is that it will nonetheless be a challenging and rich investigation: For one thing, the data may not be ideal for answering the question.  It would really help to have data with topics/skills/units assigned to students in random order, but that would not fly, so the team would have to "make do" with data from the standard curricular sequence. (Still has good potential, though.)  As well, there are questions of what metrics to use (large space of possibilities, including AFM-derived measures) and at what level to look for prerequisite relations (units? skills? topics? problem types?). Also, when are correlations strong enough to be considered evidence of prerequisite relations?  (Over and above other variables that might cause correlations between work on tutor units, most of which we will not have in the data?)  This work would have potentially interesting implications for EDM and education more generally. (Some lit review would need to be part of the project.)  I am not aware of any really good empirical investigations of prerequisite relations, so any surprising findings could have profound consequences for how we teach math in middle school, say.  Though the data set used will likely be about middle school mathematics, the approach might generalize. More concretely, these investigations could become a foundation of a notion I would like to develop, namely, "adaptive practice of prior knowledge." Perhaps that term is self-explanatory enough for now ... I could say more if needed. I have been wanting to pursue this idea for years now, but have not gotten to it.  I know Carnegie Learning is interested as well. (I would need to check with them what data set they could make available for this project.)
* **Develop a method (or parts thereof) for summarizing individual students' performance with tutoring software over a session or even longer, based on tutor log data, in a way that is useful for teachers.** This one is somewhat ill-defined, but potentially very interesting. Imagine a scenario in which a class of middle school students (perhaps in classes that are remote only) uses tutoring software at home. At the start of her day, the teacher reviews the progress (or lack thereof) of her students, as they worked on the tutoring software, as they were expected to do, yesterday. What would be a good way for the software to present, say on a to-be-created dashboard, to the teacher, a brief summary of each student's work, distilled from tutor log data, in a way that the teacher can digest in a very short amount of time? The teacher needs to gauge which students are productively learning, and needs to decide, for each student, whether and what kind of further (personalized) communication with the student is needed (e.g., quick text message or voice message with encouragement, or schedule Zoom consultation later in the day).  Say, in under a minute per student.  One broad idea is to show the teacher, for each student, or for each student whose work session may not have been fully productive, trends and examples. That is, trends (or overview stats) would summarize progress and struggle, the latter perhaps divided into resolved and unresolved struggle - as distilled from the log data. Examples - i.e., examples of the particular student's attempts at solving certain problems in the tutoring software, extracted from the log data  - would also illustrate success or struggle, though might probably focus in particular on struggle that may not have been resolved (as that is where the student might need help the most). The trends provide context, the examples specifics. Only a very small number of examples can be presented in the available time (perhaps 2 or 3, possibly with more available at the teacher's request). We know from past work on Lumilo (Ken Holstein's smart glasses for teachers) that carefully-selected examples can be very helpful for teachers - a good way of communicating what students might need help with. But that work focused on in-the-moment helping, whereas the current scenario is focused on a teacher's reviewing their students' recent past work, which would no doubt change what kinds of analytics about trends, and what ways of selecting examples, might be most effective. The project would focus on these latter questions - i.e., development of (related) methods for distilling trends and selecting a very small number of . The success of these methods might be measured in studies with teachers, which may be somewhat unusual for a stats project, but could be interesting. The work would build on extensive prior work in my lab (including Ken Holstein's work mentioned above) on developing a teacher dashboard for use with intelligent tutoring software. If the above is too open-ended, one way to make it specific would be to narrow it to the issue of distinguishing smooth work from resolved struggle from unresolved struggle, within a session on the tutor, and possibly at the  level of skills (knowledge components). (Past work in EDM has focused quite a bit on detecting wheel spinning and struggle, but not on past struggle, and whether it was resolved or not.) One step more ambitious would be to add methods for selecting examples.
* **Develop a way to detect small (?) learning discontinuities within tutor log data to measure effects of out-of-tutor events.**  A common scenario is one in which middle school students are using tutoring software in class (in person or online). The teacher "monitors" the class and helps those students who appear to need more help than the software gives them, usually with very short individualized sessions focused on the given tutor problem that the student is working on.  A question that keeps coming up (for one thing, teachers would like to know) is:  How helpful are these teacher interventions to students? Do they put students on a different learning trajectory, with respect to the skills that were discussed? (And perhaps other skills.) How might we measure their effect?     There may be easier and harder ways. A simpler way may be to look at the error rate on subsequent opportunities for the skill (is the student not stuck anymore?), or even the BKT estimate.  A slightly more intriguing idea might be to see if there is evidence of faster learning (e.g., in terms of AFM slopes - though perhaps taking into account more information about each opportunity than just right or wrong) following the intervention. Exactly how that would work is not very clear to me, so that's one of the challenges for this project. We have data sets that include teacher interactions with students, in the context of learning with tutoring software. In addition to using real data, it might perhaps be useful to explore this problem with "minimally synthetic" data.  E.g., leaving out data points from log data (e.g., skill opportunities) might be a way of simulating the effect of potentially beneficial out-of-tutor events. Adding data points could, likewise, be a way of simulating the effect of ineffective learning within the tutor. (The question then becomes whether these discontinuities could be detected automatically.) There might be other ways to use synthetic data as well, but this one might have the advantage of maintaining realism. See this as brainstorming, though, there may be better ideas to be had. This investigation would have many applications, including  measuring effects of student-teacher or student-student (e.g., collaborative learning) or student-parent discussions, for the purpose of giving teachers feedback on their interactions with students (e.g,. on a teacher dashboard), or for research purposes (e.g., is letting students work collaboratively for brief periods of time a useful way of combating student struggle?).  The most relevant prior work that comes to mind is Ido Roll's work on help seeking, in which he used a "local measure of learning" to study the effect of asking for a hint from the tutor). Other relevant work is that by Ryan Baker on detecting the moment of learning.  Perhaps there are connections to be made with past work on detecting wheel spinning, off-task behavior, and possibly work by Käser et al. on - asymptotic convergence.

**PITTSBURGH PUBLIC SCHOOLS**

Two possible projects. Contacts are Steve Greene, Deb Friss & Anthony Periell of PPS. Data is a combination of internal PPS tracking data and data from the National Student Clearinghouse (NSC). Likely some of this work would require a confidentiality agreement with respect to individual student records. These project descriptions need a bit more elaboration, but the underlying projects seem to be of a reasonable magnitude.

* **College and Trade School Enrollment Indicators.**  Build a good prediction model for enrollment in college, trade school, or other post-secondary activity. Predictor variables would come from internal PPS data; outcome measures from NSC data. Contact person is Chanelle Labash.
* **Promise Use and post-secondary retention and mobility.** What predicts use of Promise scholarships, for eligible students? What factors affect retention at and mobility among post-secondary institutions, for students who use Promise scholarships? Again, the predictor variables are all internal PPS data, the outcomes (which might be quite diverse) are the NSC.