Mining Reading Comprehension Within Educational Objective Frameworks

Terry Peckham
Department of Computer Science
University of Saskatchewan
110 Science Place
Saskatoon, SK., Canada
tep578@mail.usask.ca

Gordon McCalla
Department of Computer Science
University of Saskatchewan
110 Science Place
Saskatoon, SK., Canada
mccalla@cs.usask.ca

ABSTRACT
In this paper we explore patterns in student behavior as they answer questions about documents they are reading. In earlier work [4] we showed that as students answer a question online, they can be categorized into one of 4 different clusters of "reading-scanning-scrolling" behaviors. Further, their reading-scanning-scrolling behavior category predicts the quality of their answer to that particular question based on the level of that question in Bloom's Taxonomy. We have performed a second experiment that confirms these earlier results. In a third exploratory experiment we also show how the reading-scanning-scrolling clusters already discovered can be refined for use with another taxonomy, the Marzano Taxonomy. We are currently exploring whether other clusters can be found to help understand student behavior in terms of the Marzano Taxonomy.

Keywords
K-means clustering, Bloom's Taxonomy, Marzano's Taxonomy

1. INTRODUCTION
Educational objective taxonomies form a pedagogical framework for understanding student learning. Within the classroom environment, these taxonomies are utilized to challenge the teachers and instructors to move beyond simple low level learning.

Bloom's Taxonomy of Educational Objectives and its subsequent revision by Anderson [1] is a widely used taxonomy within the classroom. It is comprised of three major domains, the cognitive, affective and psychomotor. The cognitive domain is comprised of six hierarchical categories ranged from the easiest cognitive tasks to the most difficult cognitive tasks. The categories, from lowest to highest, are knowledge, comprehension, application, analysis, synthesis and evaluation (as revised by Anderson et. al. [1]).

In response to shortfalls found within Bloom's Taxonomy [2], Marzano and Kendall [3] in 2007 introduced their taxonomy of educational objectives. Marzano's premise is that knowledge use is affected by three systems: the cognitive system, the metacognitive system and the self-system [3]. When an individual is faced with some new situation, the self-system must determine if it is better to continue with the current behavior or to adapt some new behavior. The metacognitive system then tries to set the goals that are needed to achieve the desired outcome and then monitor those goals. The cognitive system processes all the necessary information required to complete the task that is obtained from the knowledge system [3].

Each of Bloom's categories for the cognitive domain can map over to one of the categories for Marzano's cognitive domain. However, there is no one-to-one mapping possible between these domains [2]. So in practice we will find that a problem categorized as Bloom level 2 (understanding) may equate to Marzano's level 2 (comprehension) or to Marzano's level 1 (knowledge) depending on the context of the problem.

This paper extends our earlier work [4]. In particular, we wanted to confirm the results of our first experiment. Additionally, we wanted to see if we could move from the Bloom taxonomy to the Marzano taxonomy, and whether this would lead to a more refined predictive capability.

2. METHODOLOGY AND RESULTS
Our initial and confirmatory experiment was performed to determine if there were useful patterns of student usage that could be found within a simple learning content management system [4]. Students were given multiple documents that contained novel information and then were asked multiple questions to determine how they had learned the material presented. The students were allowed to freely move between the various articles and questions presented to them and could freely interact with the content they were expected to learn. Following the trace methodological approach, all of the interactions/events in the system were captured and time-stamped. The events captured included mouse clicks, mouse wheel movements, button clicks, typing, and so on. Over the two experiments, a total of 50 participants were tested generating over 63,738 events.

Based on the timestamps of these events, we were able to measure when students were reading (slowest), scanning, or scrolling (fastest) through the document. The time cutoffs used to differentiate between the reading, scanning and scrolling categories were consistent with other document navigation literature,, as discussed in [4]. In the first experiment we found no significant differences between the clusters until the level of knowledge needed to answer a question in terms of Bloom's Taxonomy was factored in [4]. Then, over many k-mean clustering iterations we discovered 6 clusters that allowed us to predict the quality of the students' answers to questions based on their Bloom level, with the 4 most predictive as follows: Light Reading Cluster (50% reading, 30% scanning, 20% scrolling) (50:30:20), Light Medium Reading Cluster (60:30:10), Heavy Medium Reading Cluster (70:30:10), Heavy Reading Cluster (80:10:10). In experiment 2, we used the clusters found in [4] predictively as metrics and checked to see if we would still obtain significant differences between the clusters.

Table 1 shows that for experiment 2 all of the levels tested have significant differences. This shows that the clusters from experiment 1 hold up well in predicting students' answers to questions in experiment 2, thus confirming the results of the first experiment.

Bloom Level	F	P	F-Critical
1	23.137	1.04E-6	3.09
2	33.245	2.47E-7	3.19
3	21.237	.005796	6.60
4	50.535	.000854	6.60
5	25.128	1.18E-6	3.15

Table 1 One way ANOVA for Bloom Level Experiment 2

Again as in[4], the clustering does not predict an exact grade on a question but provides a more coarse grained prediction of a student's performance. For example, question 2, experiment 2 asked for a student to recollect two pieces of information. The heavy reading cluster almost always involved the student achieving a failing grade while those students who performed more scanning obtaining a grade greater than 75%. Those students who performed more scanning and who did not receive higher grades did so because they misinterpreted the question.

	50,30,20	60,30,10	70,20,10	80,10,10
50,30,20	-	0.19626	0.15202	0.13407
60,30,10	0.25348*	-	0.1896	0.17554
70,20,10	0.3588*	0.10529	-	0.12412
80,10,10	0.4651*	0.21159*	0.1063	-

Table 2 Tukey-Kramer Analysis Bloom Level 2 Experiment 2

Table 2 demonstrates the differences between the clusters for experiment 2. Again we see that there are significant differences but those differences tend to be between the 50:30:20 and the 80:10:10 clusters. In the second experiment the participants consisted primarily of individuals that are heavy computer users. This contrasts with the participants in experiment 1 that were primarily novice computer users. The participants from the second experiment tended to either perform heavy reading or the other extreme with the highest scanning and scrolling ratios. The middle two clusters were under-represented in the second experiment. The reason may be that the more advanced computer users have found strategies which allow for successful information processing in online environments.

Recently, Marzano's Taxonomy [3] has become popular, partly because it has finer grained sub-categories. This left us wondering if we could make predictions using Marzano's Taxonomy similar to those we did using Bloom. We decided to look at this in a third exploratory experiment where we recast the data from the first two experiments in terms of Marzano's categorizations..

To this end, questions used in experiment 1 and experiment 2 were re-categorized in terms of Marzano's Taxonomy. .Since in Marzano the cognitive domain only contains 4 main levels, there was a slight generalization from Bloom to Marzano. Table 3 shows that statistically significant predictions could be made about students' performance on questions at the first 3 levels of Marzano using the questions from experiment 2 Level 4 of Marzano did not show up as statistically significant. As with the first experiment, there weren't sufficient numbers of students to obtain significant values. However, when we combined the questions from both experiments 1 and 2, we can even make significant predictions at Marzano's level 4 ($F = 43.86$, F-Critical $= 3.00$, $p = 6.77E-10$).

Marzano's cognitive domain contains 4 main levels that can, in turn, be subdivided into 14 sublevels (see Table 3). These sublevels offer a much more fine-grained level of detail compared to Bloom. We found that our questions from the earlier experiments covered 8 of the 14 subcategories of Marzano's Taxonomy. In particular, all three sublevels within Marzano level 1 were represented. We could predict the quality of student answers to sublevel 1 questions with statistical significance, but could not do so with the other two sublevels of Marzano level 1. Nevertheless, this does give hope that we can make predictions at this more fine-grained levels offered by the Marzano Taxonomy, although likely larger studies will be needed, with more students, to discover the relevant clusters.

Marzano Level	F	F-Critical
1 (Sublevel 1, 2, 3)	120.98	2.73
2 (1, 2)	62.31	3.07
3 (1, 2, 3, 4, 5)	52.71	3.91
4 (1, 2, 3, 4)	0.60	3.58
All Levels Combined	1.40	2.67

Table 3 Tukey-Kramer Analysis Marzano Experiment 2

CONCLUSIONS

Our three experiments lead to the following general conclusions. First, the patterns discovered in the first experiment seem to hold well for the second experiment. This provides more confidence that they actually represent real behavioral differences, and that it is worthwhile to look at student activities in terms of the Bloom level of the tasks they are trying to accomplish. Second, the patterns to some degree survived when the questions were relabeled in terms of the Marzano taxonomy. This points out that there are strong correspondences between Bloom and Marzano, and even opens up the idea of perhaps formally exploring these correspondences in future through mining actual student behavior as they solve problems at various levels of the two taxonomies. Third, the promise of Marzano's taxonomy, with its more refined categorizations, to explain why the reading-scanning-scrolling behaviors lead to the various outcomes that they do has yet to be fully validated

REFERENCES

[1] Anderson, L.W., Krathwohl, D.R. and Bloom, B.S., (Eds), 2000, Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Longman.

[2] Aworuwa, B. & Nkoge, B. (2007). The New Taxonomy of Educational Objectives and Implications for Designing Instruction for Distance Learning Delivery. *In T. Bastiaens & S. Carliner (Eds.), E-Learn* 2007 (pp. 1394-1398).

[3] Marzano, R. J., & Kendall, J. S. (Eds.). (2007). *The new taxonomy of educational objectives*. Corwin Press.

[4] Peckham, T., & McCalla, G. (2012). Mining Student Behavior Patterns in Reading Comprehension Tasks. *Proc. EDM 2012, Crete, Greece (pp. 87-94).*