Impact of Automated Scoring and Feedback on Scientific Argumentation in Earth Science

Through the Log Data Analysis

Mengxiao Zhu, Liyang Mao, Ou Lydia Liu
Educational Testing Service
Princeton, NJ USA
mzhu, lmao, lliu @ets.org

Amy Pallant;
Concord Consortium
Concord, MA USA
apallant@concord.org

Extended Abstract—

In formative assessment, constructed response questions are typically used for scientific argumentation, but students seldom receive timely feedback while answering these questions. The development of natural language processing (NLP) techniques makes it possible for the researchers using automated scoring engine to provide real-time feedback to students. As is true for any new technology, it is still unclear how automated scoring and feedback may impact learning in scientific argumentation. In this study, we will take advantage of log analytical tools to examine the granularity of students' interactions with automated scores and feedback and investigate the association between various students' behaviors and their science performance. Innovative log analysis makes it possible to capture the rich information and identify patterns from students' log data. Specifically, this study addresses the following research questions:

1. How do students use automated scoring and feedback?

2. Are students' performance improved with automated scoring and feedback?

The benefit of using log-data is two-fold. First, it tracks all activities conducted by students while answering questions and interacting with the automated scoring engine. This complete log enables the researchers to recover all events that happened. Second, log data is automatically collected by the project server without interfering with student activities, which minimizes the impact of data collection on student activities.

In this pilot study, data were collected from 42 students who engaged in a network-based earth science assessment on climate change developed by Concord Consortium. There are eight argument blocks in the assessment. The answers for the scientific argumentation items in these blocks were graded by the c-rater-ML, an automated scoring engine develop at ETS for scoring short-text items and providing instant feedback. The log data includes time-stamped information on student activities, such as, visiting to a certain page, visiting to a certain question, or submission of answers for automated feedback. Analysis of the log data can help us identify the patterns of student actions such as time on task, activity navigation, and answer revisions after getting feedback.

Through log data analysis, we first recovered and visualize the pattern of students navigating through the argument items. Analysis shows that even though students varied on the number of times they checked the automated feedback and made revisions, only *11.9%* of students did not made any changes after getting instant feedback. On average, students made one revision ($\mu = 0.96$) for each argumentation block. The maximum number of feedback checking and revising is 24, in which case the student improved his/her score from an argument item from 3 to the highest possible score of 6. A boxplot of the number of times students checked the feedback and attempted to answer the items is shown in Figure 1. To collect information on students' perceived usefulness of the automated scores and feedback, students were asked to answer one question to rate the usefulness of the scores and feedback they received after each submission. The rating ranges from "not at all", "somewhat", to "very". Among all ratings, the majority (*78.46%*) are either "somewhat" (*42.13%*) or "very" (*36.33%*). We also generated variables on student activities, including time spent answering questions, number of attempts on answering questions, time spent reading feedback, time spent revising answers, and number of attempts on revising answers/checking feedback. Students' performance on argument items before and after feedback and revisions were also extracted from the log data. Preliminary analyses show that checking feedback and making revisions can significantly improve students' final scores.