The wealth of data collected provides new opportunities for Hierarchical EDITLINK to post CHG We use Multi step (in for a given pattern short answer score gains from pre-2013)

The dataset comes from an experiment we ran with 98 middle school students who used Betty’s Brain system to learn thermoregulation phenomena

We used the sum of the students’ thermoregulation multiple choice and thermoregulation short answer score gains from pre- to post-test, $S_{gain}$, to divide them into:

- high performing group vs. low performing group

The histogram of students distribution over $S_{gain}$ (the figure on the right) is approximately bimodal. We used the median gain (5.0) as the boundary for the two groups

### Related Works

**Sequential Pattern Mining (SPAM):** SPAM with bitmap representation for sequences performs a Depth-First Search traversal to find patterns that exceed a pre-defined frequency threshold (Ayres, 2002)

**Differential Sequence Mining (DSM):** DSM finds frequent patterns that are good to differentiate between different groups of students (Kinnebrew, 2013)

### MFH-SPAM algorithm

**Multi-Feature Hierarchical Sequential Pattern Mining:** MFH-SPAM extends SPAM algorithm by adding two new pattern extension steps and forms a multi-dimensional pattern extension schema:

- Sequence-extension step (S-step), a flat extension originally in SPAM
- Hierarchical-extension step (H-step) extends the last action hierarchically
- Feature-extension step (F-step) extends the last action by associating with features

### Output Analysis

**Algorithm Workflow**

Pattern Selection:
- Even for moderate size of datasets, MFH-SPAM finds a huge number of frequent patterns
- We adopt the idea from DSM to find frequent patterns that are good differentiators between different student groups. For example:
  - High performers vs. Low Performers

**Pattern Selection Metrics**

**Classifier Wrapper (CW):** It uses the $F_1$-score of a 5-fold cross validation for a Decision Tree Classifier as the metric to select the best set of frequent patterns

**Information Gain (IG):** The computational complexity of CW, which builds decision tree multiple times and results in redundant Information Gain computation is very high. We directly apply Information Gain as a new selection metric

- The Information Gain for a given pattern $P_1$ is computed from the reduction of Shannon Entropy when $P_1$ becomes known, where Shannon Entropy measures homogeneity for a sample data and it is given by:
  \[ H(P_1) = \sum_{i=1}^{n} p_i \log_2 p_i \]
  where n is the number of student groups, $p_i$ is the probability of $P_1$ been used by students in group i
- We use instance frequency (i-frequency) of patterns to compute the information gain. The i-frequency is a normalized measure of how often individual students uses a particular pattern
- For each patter, the information gain will be computed only once which improved the computational complexity over the classifier wrapper method

**Case Study**

**Dataset:**
- The dataset comes from an experiment we ran with 98 middle school students who used Betty’s Brain system to learn thermoderegulation phenomena
- We used the sum of the students’ thermoregulation multiple choice and thermoderegulation short answer score gains from pre- to post-test, $S_{gain}$, to divide them into:

- Top 3 patterns as differentiators for both selection methods

**Results and Analysis**

**Future Work**

- We will perform more systematic analysis of the differences between groups using hypothesis testing methods
- In addition, we will use correlational analysis to study in more depth the relations between behaviors and performance
- We will also work towards using the patterns derived to detect learner behaviors online, and develop scaffolding mechanisms