

Analyzing data from Spellbee 2002

Ari Bader-Natal

April 1, 2003

Goal and assessment

- **Goal:** *Design a game structure which motivates peers to create appropriate challenges for each other.*
- **Assessment:** How can we best recognize and measure success? Begin by more formally specifying the goal and its ambiguously-defined terms.

Define and validate

- How do we measure *challenge difficulty*? why is this valid?
- How do we measure *student ability*? why is this valid?
- How do we measure *problem appropriateness*? why is this valid?

Questions

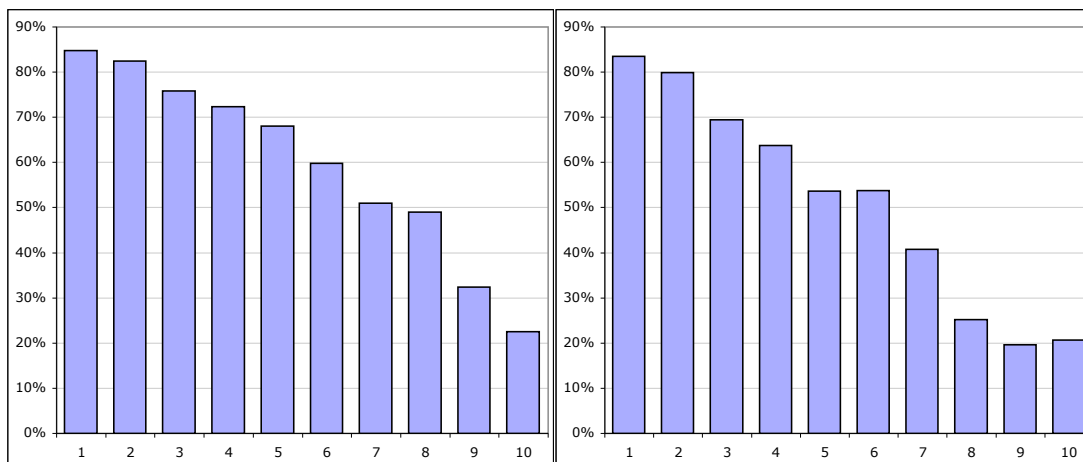
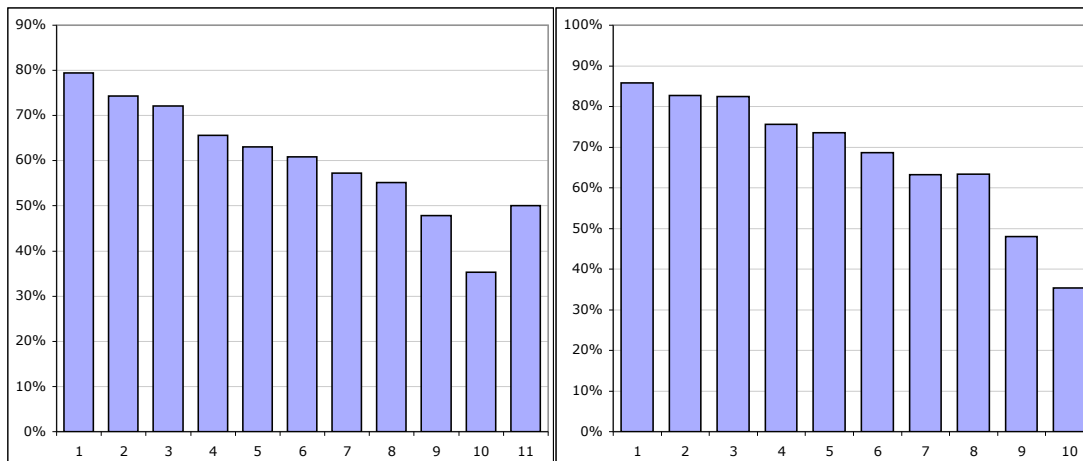
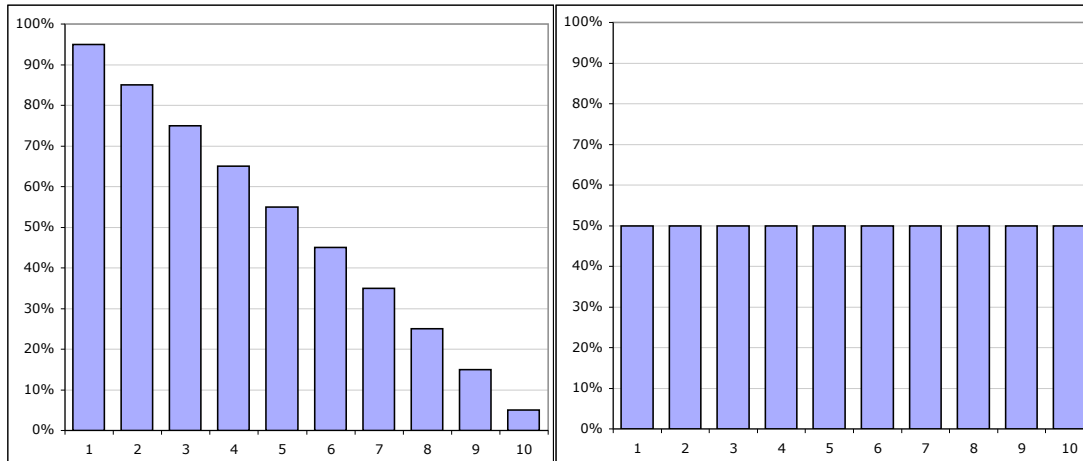
- Does our collected data support these decisions? (word difficulty, player level, and problem appropriateness metrics)
- Do students respond to the various game matrices?
- How does the choice of metrics affect play and analysis?
- How does the distribution of selected challenge difficulties vary over time and by protocol?
- Do we find that a student's ability increases over time?
- Do we find that players collude?

Word Difficulty metric

- **The options:** *lowa* numbers for ~5000 words for grades 2-8, and “*scrabble score*” for any possible word.
- **Informal validity check:** Spellbee players attempted to spell several thousand words. Compare their success to the word difficulty metrics.

Comparing observed student behavior to various difficulty metrics

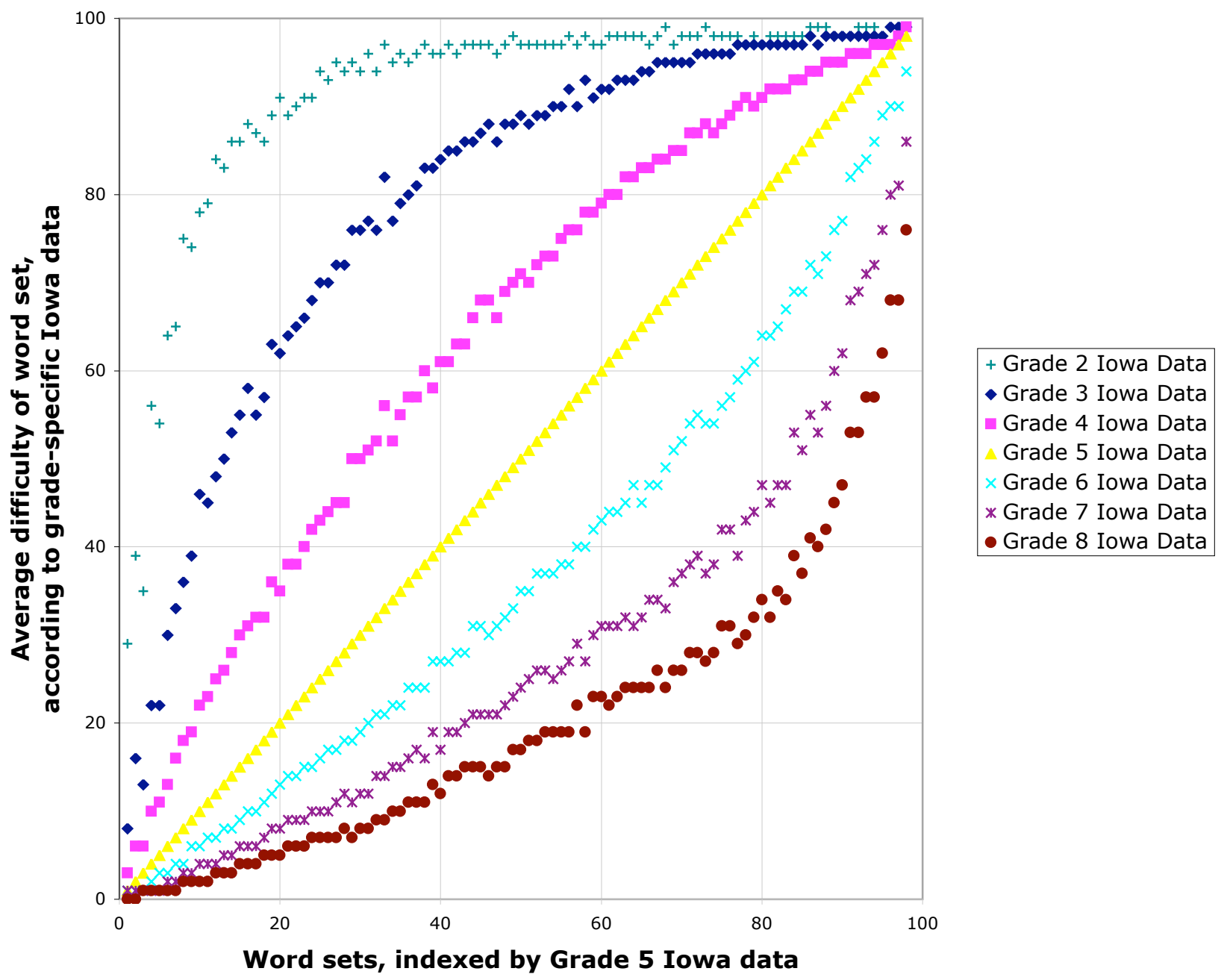
Percentage of words in this set answered successfully by Spellbee students



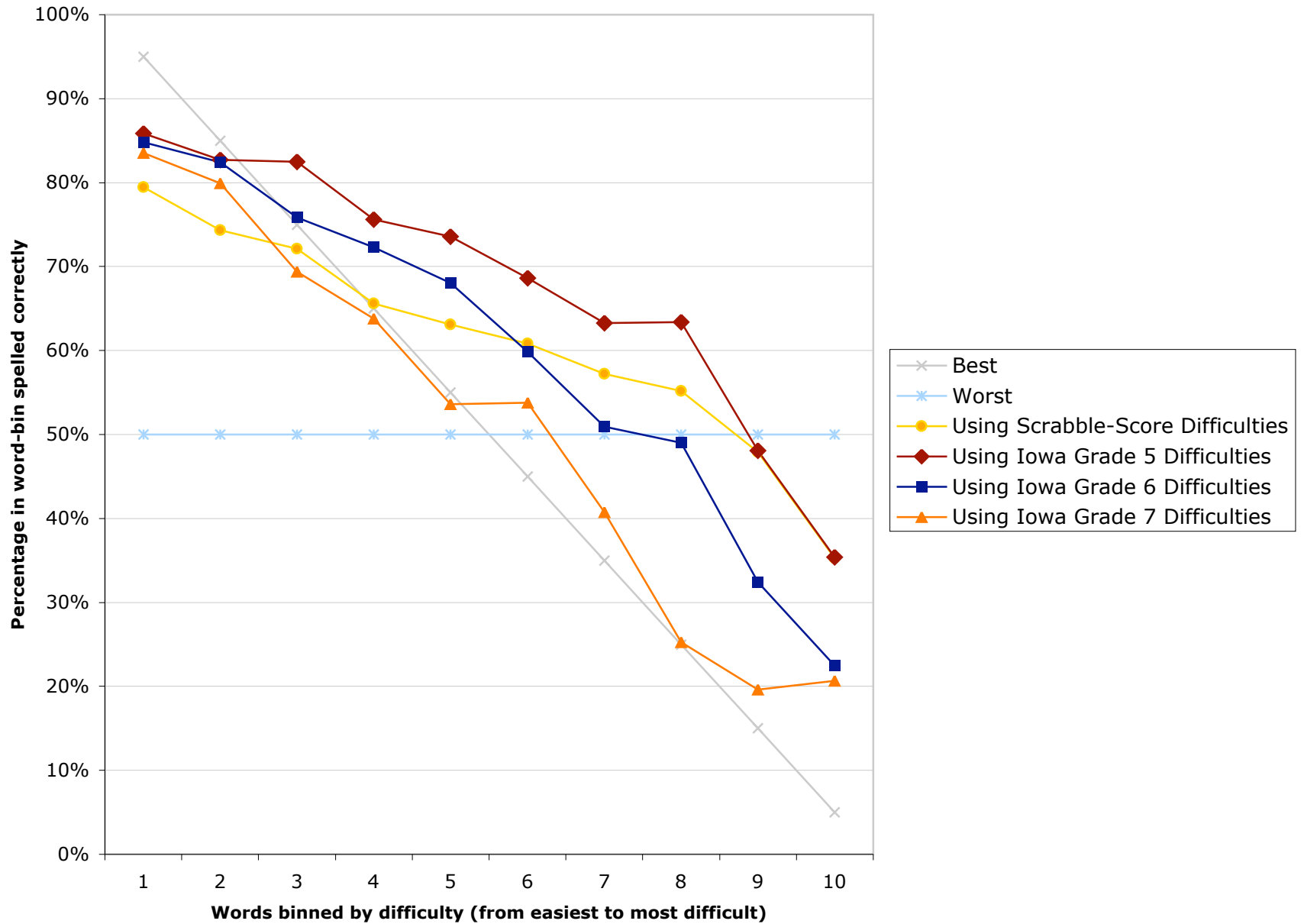
iowa6

iowa7

Word difficulties by grade, relative to Grade 5 Iowa data



Student spelling accuracy vs. word difficulty (measured using various Iowa datasets)

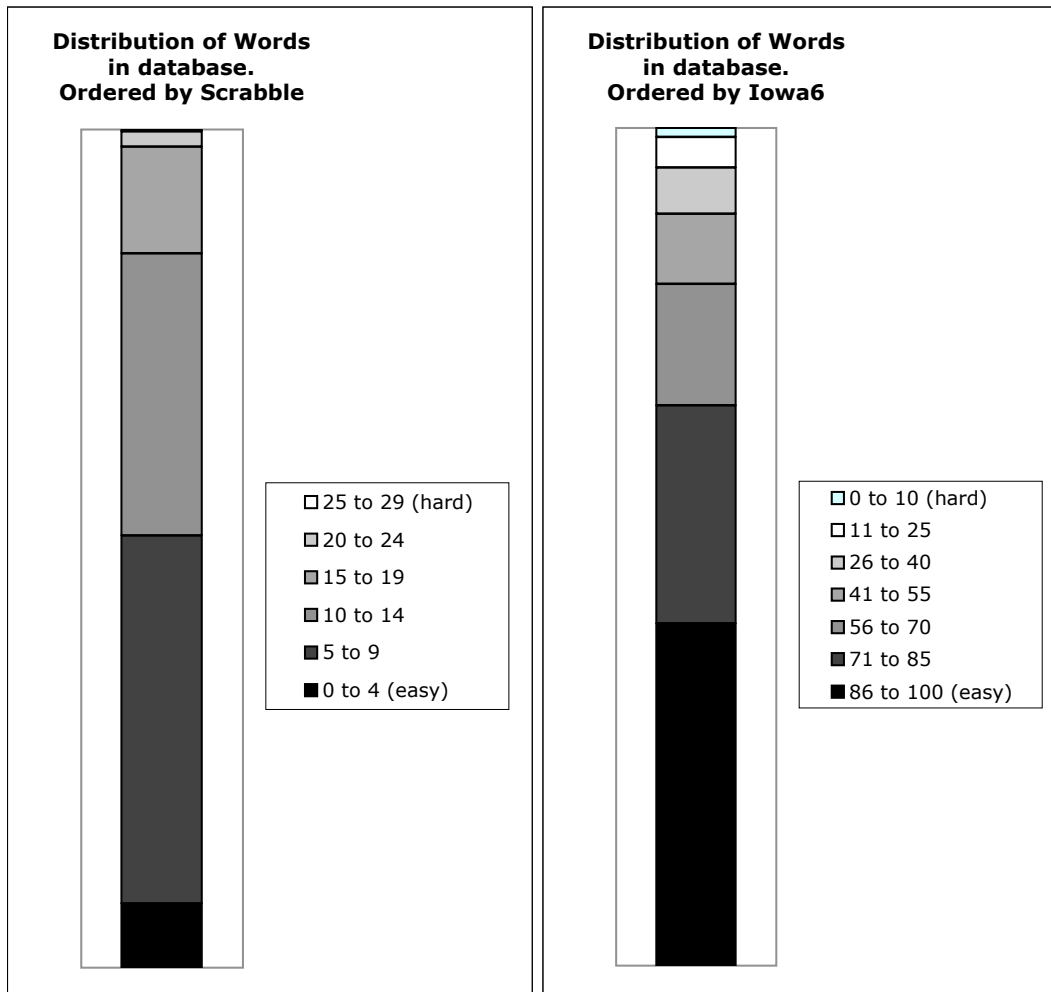


Changes in challenge difficulty over time

- Does this vary by protocol? Do students “learn” the protocol?
- Protocol Refresher:
 - Protocol 6: Reward success (regardless of difficulty)
 - Protocol 7: Reward success on hard problems and failure on easy problems
 - Protocol 8: Reward based on difficulty (regardless of right/wrong)
 - Protocol 9: Reward hard problems and double-reward failure on easy
- View changes in distribution of challenges over time...
 -

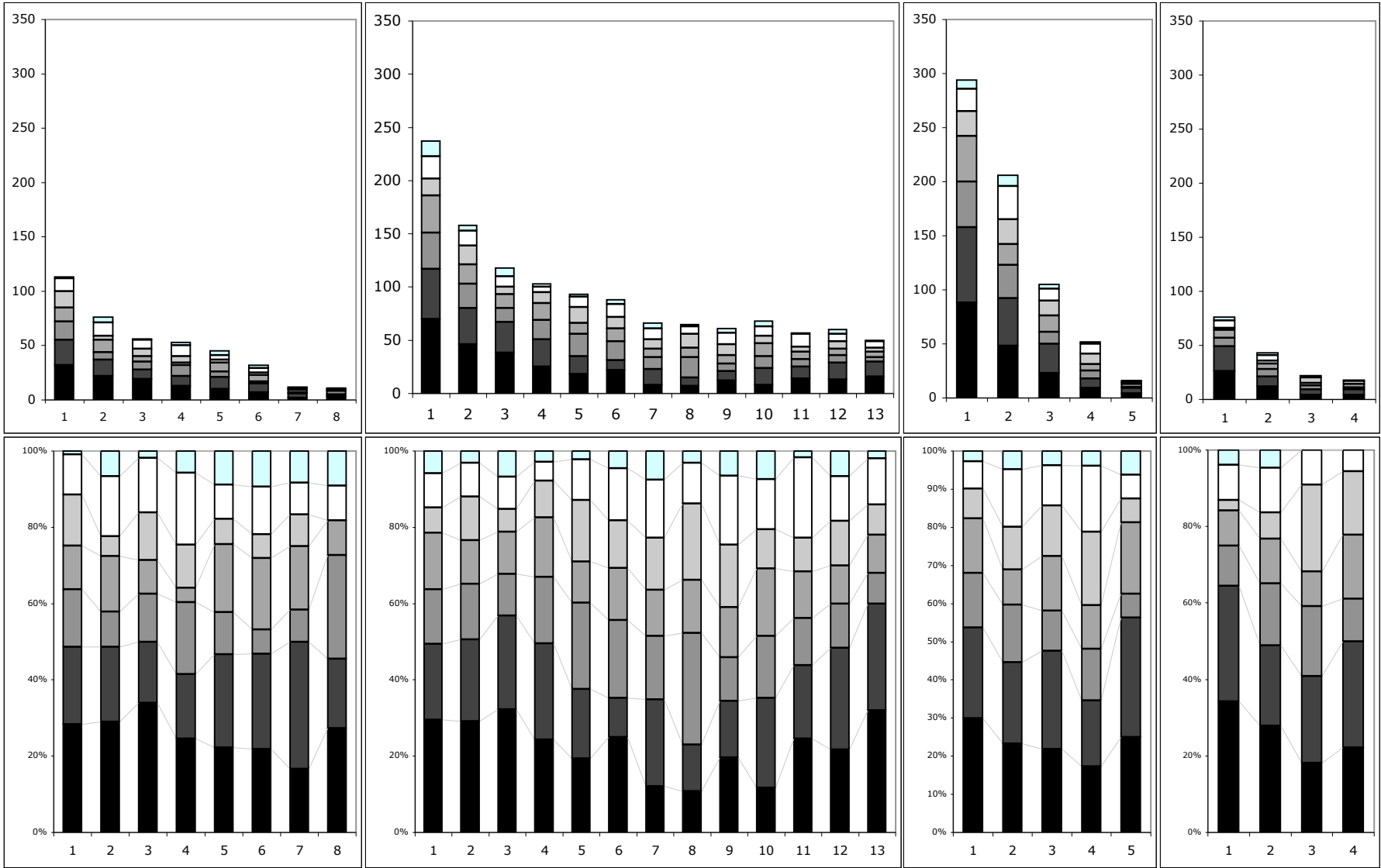
Biases

Words options were randomly selected from the database,
but both metrics view this database as a biased source of words.



- The database itself is a biased source of words, according to the various metrics.

Changes in challenge difficulty distributions over time - measured using Iowa6



Protocol 6

Protocol 7

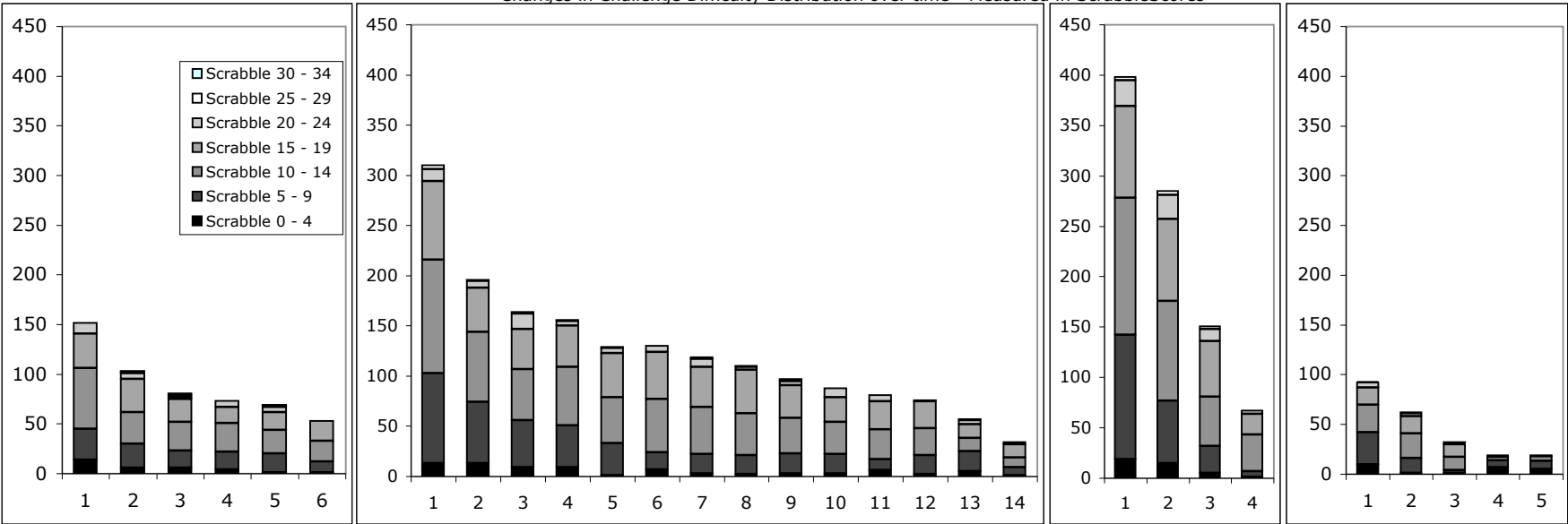
Protocol 8

Protocol 9

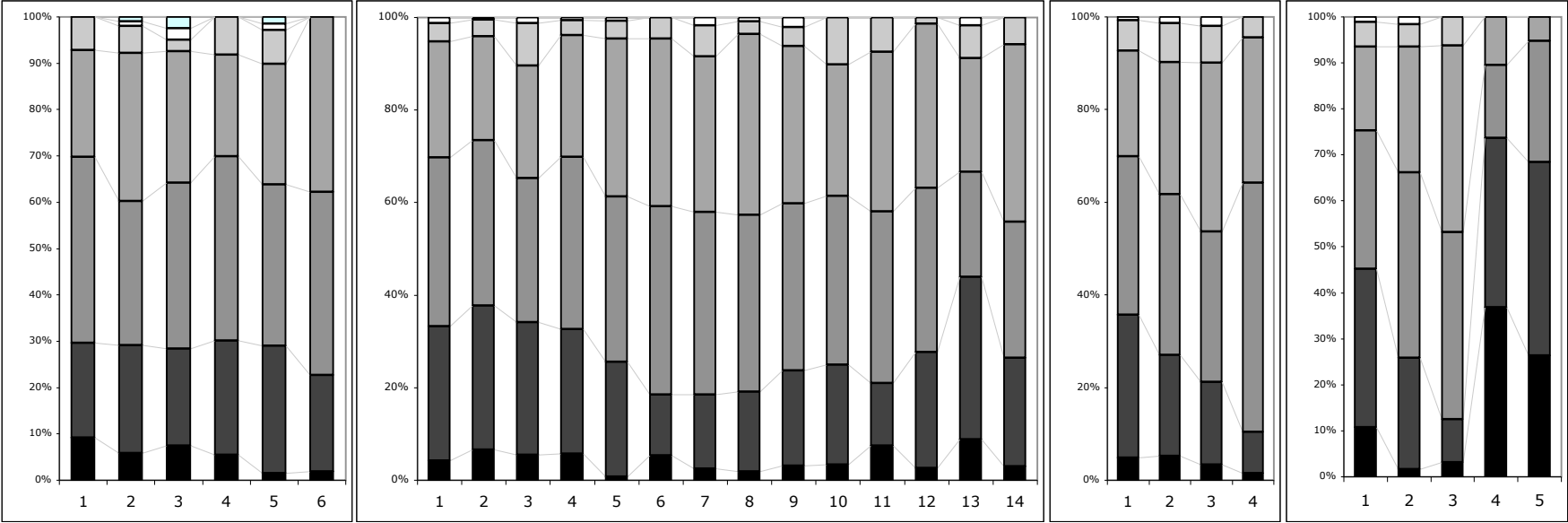
Difficulty values drawn from Iowa data - Grade 6 column

Changes in Challenge Difficulty Distribution over time - Measured in ScrabbleScores

Word Count



Distributions of above data



Protocol 6

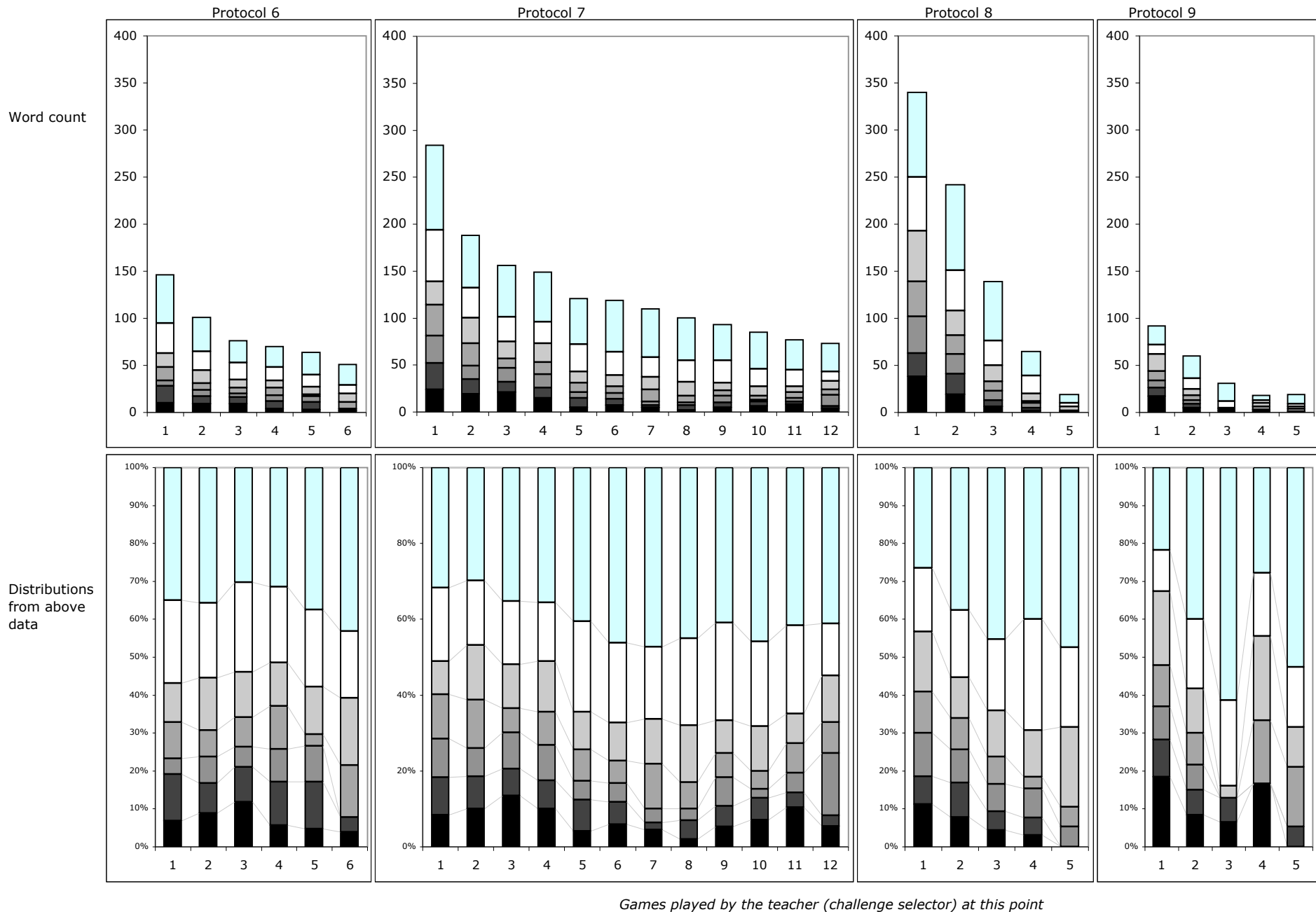
Protocol 7

Protocol 8

Protocol 9

Games Played by the teacher (challenge selector) at this point

Changes in Challenge Difficulty over time - Measured by relative selection difficulty

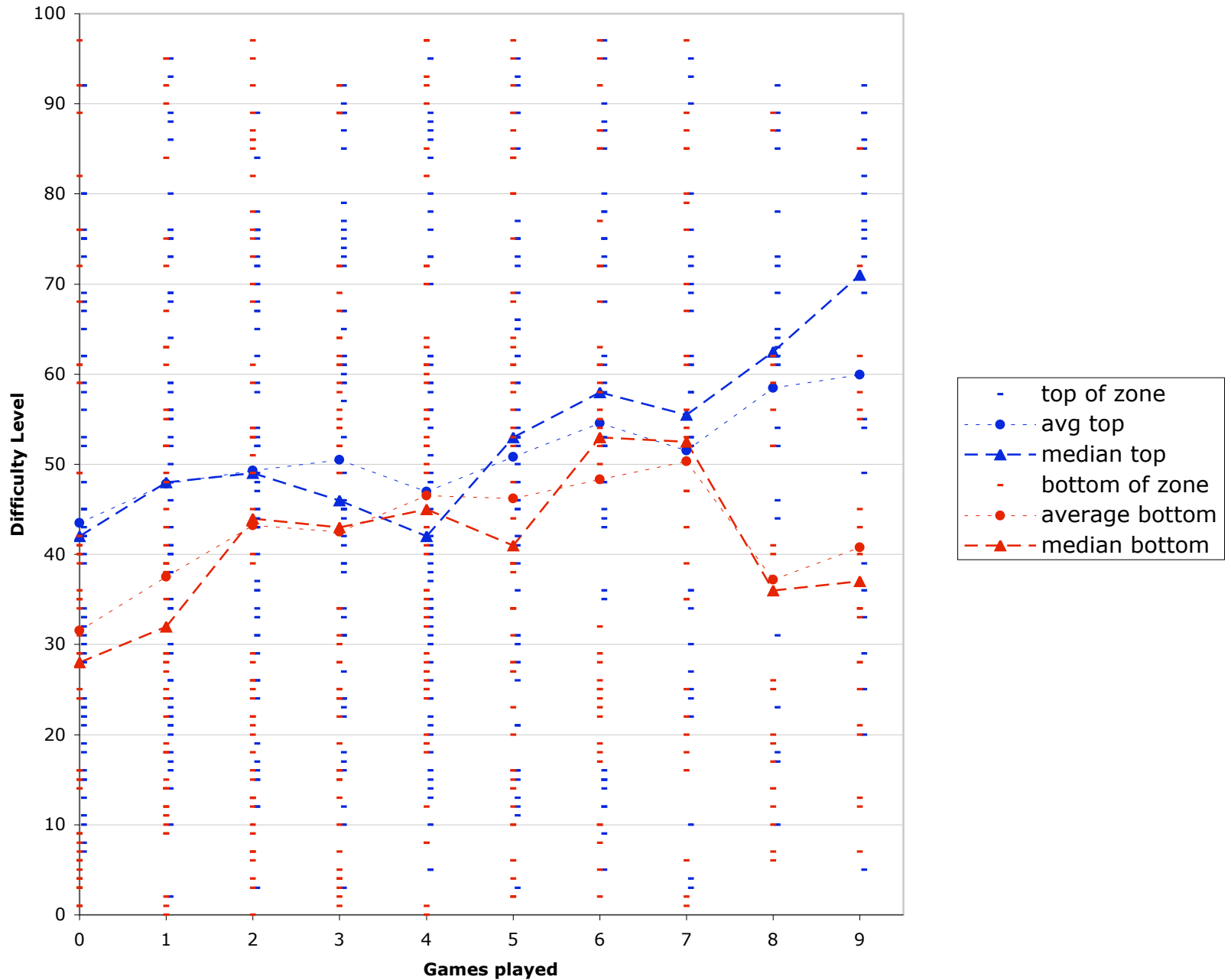


Each word/challenge selected by the teacher is a choice of one of seven possible options. In these graphs, a word's difficulty is measured relative to the other options presented.

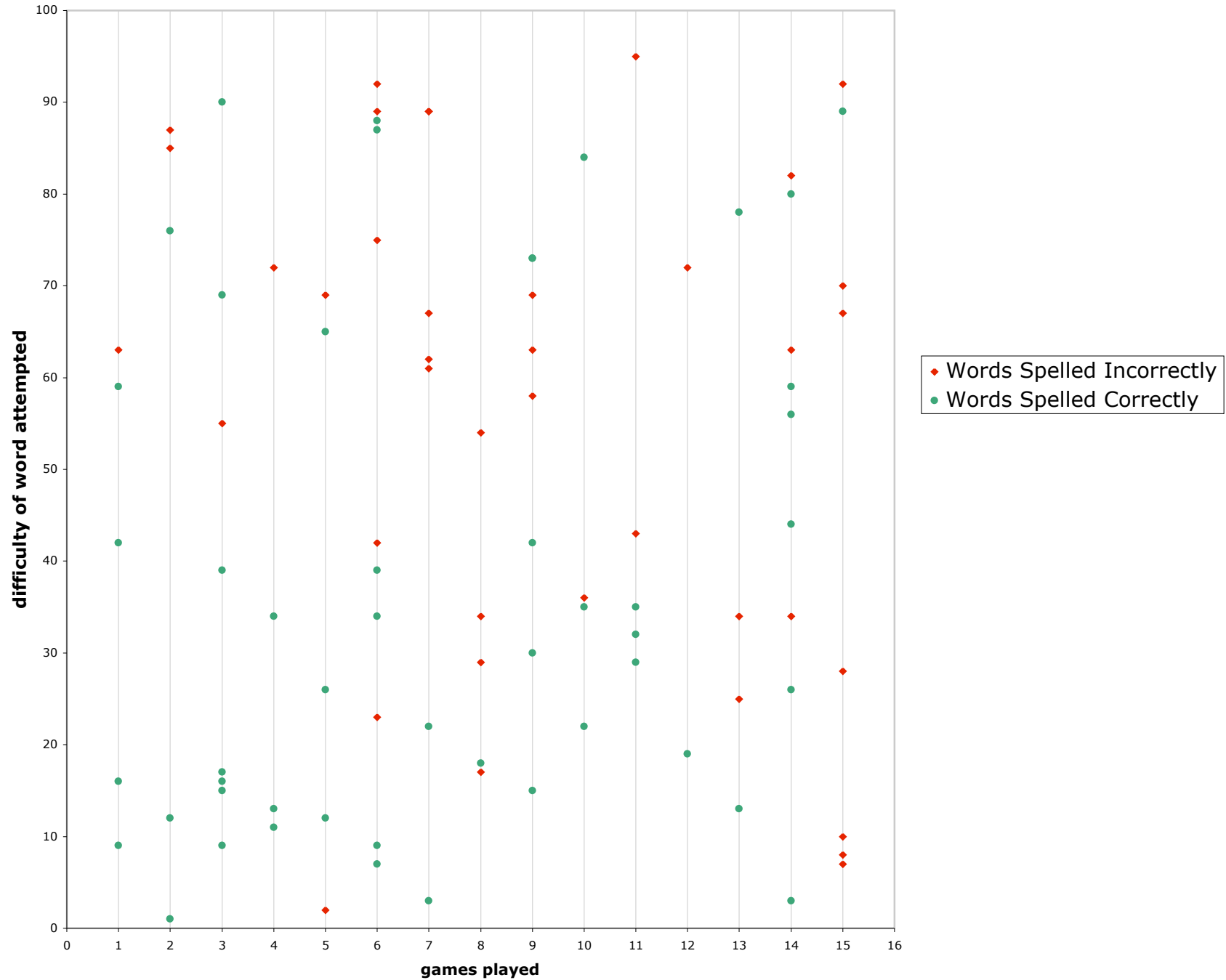
The bottom sections (black) indicate the selected word was the easiest of the seven, and the top sections (blue) indicate that the selected word was the most difficult of the seven.

Relative rankings were done based on ScrabbleScore, as teacher payoffs were based on this.

Aggregate changes in zone boundaries over the first ten games of play



History of Fred as Student



What's Next

- Moving towards analyzing *individual* games, and *individual* student histories. (Effects like collusion will likely occur in individual cases but are hard to spot in aggregate data.)