Fixed and Sliding Goals in Education?

A tiny example of a famous systems insight, applied to education,  

*with a surprising, instructive error*
Striving to Reach an Achievement Goal

An adaptation of the sliding goals structure in Forrester (1968), “Market Growth as Influenced by Capital Investment”
Striving to Reach an Achievement Goal

- Goal for student achievement
- Gap between achievement and Goal
- Achievement or understanding in a course
- Time to close the gap
- Pressure to close the gap
Striving to Reach an Achievement Goal

- Goal for student achievement
- Gap between achievement and Goal
- Striving to reach goal (B)
- Pressure to close the gap
- Time to close the gap
- Effort expended to close the gap
Striving to Reach an Achievement Goal

- **Goal for student achievement**
- **Gap between achievement and Goal**
- **Achievement or understanding in a course**
- **Striving to reach goal (B)**
- **Pressure to close the gap**
- **Time to close the gap**
- **Effort expended to close the gap**
- **Pressure to devote time and effort elsewhere**
Striving (But Falling a Bit Short)

Pressure to close the gap : sliding goals
Goal for student achievement : sliding goals
Achievement or understanding in a course : sliding goals
Where Does the Goal Come From?

Forrester’s insightful question

Goal for student achievement

Gap between achievement and Goal

Achievement or understanding in a course

Striving to reach goal (B)

Pressure to close the gap

Effort expended to close the gap

Time to close the gap

Pressure to devote time and effort elsewhere
Where Does the Goal Come From?

Forrester’s insightful answer

Goal for student achievement

Sliding goal (R)

Gap between achievement and Goal

Achievement or understanding in a course

Effort expended to close the gap

Pressure to devote time and effort elsewhere

Time to close the gap

Pressure to close the gap

Striving to reach goal (B)
A Flexible Goal Can Slide

Time (Week)

Pressure to close the gap : TAG 12
Goal for student achievement : TAG 12
Achievement or understanding in a course : TAG 12
Sliding Goals in Sophistication of School Texts?
School Texts Have Gotten Simpler Over Time

• The most difficult readers were generally published before 1918. By modern standards, Professor McGuffy’s pre- and post-Civil War readers were very difficult.

• Average sentence length of 1963-91 books was shorter than that of 1945-1962 books.

• Mean length dropped from 20 to 14 words, “the equivalent of dropping one or two clauses from every sentence”

• Wording of schoolbooks after 1963 for 8th graders was as simple as that in books used by 5th graders before 1963.

• Wording of 12th grade texts after 1963 was simpler than the wording of 7th grade texts before 1963.

• Today’s mean sixth, seventh, and eighth grade readers are simpler than fifth grade readers were before World War II.
Could Declining Sophistication of Texts Account for Declining SAT Verbal Scores?

Figure 1: The SAT Verbal Time Series
Sliding Goals in School Texts?

- Sophistication of school texts
- Gap between capabilities and goal
- Ability of students to handle text sophistication
- Striving to reach goal
- Pressure to close the gap
- Effort expended to close the gap
- Pressures on students to devote time and effort elsewhere
- Sliding goal

G. P. Richardson, System Dynamics Conference
St. Gallen, Switzerland, July 2012

Rockefeller College of Public Affairs and Policy
University at Albany
A Formal Model to Fit to the Data

Sophistication of texts

Gap between capabilities and goal

SAT verbal

Striving to reach goal (B)

Effort expended to close the gap

Pressure to close the gap

Sliding goal (R)

Pressure to devote time and effort elsewhere

Time to adjust goal

Time to close the gap

Time to act on gap

Pressure to devote time and effort elsewhere
Historical and Simulated SAT Verbal Scores

Sliding_goals Fit

SAT verbal : sliding SAT goal2
SAT verbal : HistoricalSAT
A Remarkable Fit!

- Fitted simulation matches slope, shape, and curvature
- Visually very convincing
- $R^2 = 0.956$

- Very tempting evidence in support of the hypothesis of long-term sliding goals in school texts and verbal achievement
But We Have Big Problems Here!

- The fit to data is very persuasive,
- But grossly misleading.
- The fit is actually mostly *accidental*.
- The optimization routines used were robust,
- But the conclusions are wrong.
- Let’s investigate.
Strategies to Uncover Flaws

• Check optimization assumptions
• Check optimization results
  • Are fitted values at extremes, suggesting best fit requires larger search intervals?
  • Are fitted parameters reasonable?
• Plot everything
Optimal Parameters for SAT Fit

• To avoid local optima, multiple starts with random initial parameter values
• Maximum payoff found at:
  • Pressure to devote time and effort elsewhere = 8.0001
  • Time to adjust goal = 38.0654
  • Time to act on gap = 17.2979
• Simulations = 30465
• Optimizations = 129
• Pass = 3
• Payoff = -711.039
Are Fitted Values Reasonable?

- Time to adjust goal = 38.0654 (years)
  - 38 years (almost two generations) to adjust the sophistication of texts seems reasonable.

- Time to act on gap = 17.2979 (years)
  - 17 years (a bit more than the twelve years of K-12 education) to change student tendencies seems reasonable.

- Pressure to devote time and effort elsewhere = 8.0001
  - Units in terms of “SAT equivalents”, not “operational” units
  - But 8 such units per year not obviously a problem
Plotting all the Stocks

Sliding_goals Fit

SAT verbal : sliding SAT goal2
SAT verbal : HistoricalSAT
Sophistication of texts : sliding SAT goal2
Effort expended to close the gap : sliding SAT goal2
What do We See?

• **Effort** (grey) is appropriate to achieve the fit.

• But **Sophistication of Texts** (green) does not follow the path of SAT scores!

• Exerts upward pressure on SAT scores throughout the run.
Variable and Goal ought to Slide Together, as in the Student Achievement example:

![Sliding_goals](chart)

- Pressure to close the gap: TAG 12
- Goal for student achievement: TAG 12
- Achievement or understanding in a course: TAG 12
Observations and Hypotheses

- Initial values of the stocks were selected by eye.
- They were accidentally wonderful.
Observations and Hypotheses

• Initial values of the stocks were selected by eye.
  • They were accidentally wonderful.
  • (An optimization involving them verified that.)
• Making Effort a stock (a smooth) may have distorted the fit.
  • The S-shaped undulations come from the second-order goal-seeking loop.
  • Maybe making it first order, or close to it, would shed some light.
Making *Effort* respond quickly, Setting *Time to Act on the Gap* to 5

- Time to adjust goal
- Sophistication of texts
  - Sliding goal (R)
  - Gap between capabilities and goal
- SAT verbal
  - Effort expended to close the gap
  - Pressure to devote time and effort elsewhere
- Pressure to close the gap
- Striving to reach goal (B)
  - Time to act on gap
  - Time to close the gap
- Time to adjust goal
Revised Optimization Results, with Time to Close the Gap constant at 5

- Maximum payoff found at:
  - Pressure to devote time and effort elsewhere = 3.74661
  - Time to adjust goal = 3.96484
- Simulations = 17162
- Optimizations = 148
- Pass = 3
- Payoff = -7434.32
What do We See?

- **Effort** (grey) settles appropriately to a constant -2.

- **Sophistication of Texts** (green) now follows SAT scores (blue)

- But the pretty S-Shaped fit goes away, as do the reasonable parameter values
A Possible Compromise: Time to Close Gap = 8

- **Effort** (grey) settles more slowly to a constant -2.
- Slight S-shape in SAT scores (blue) and a hint of one in **Sophistication of Texts** (green)
- But the fit is still visually far less convincing. No longer “obviously” insightful
Conclusions

• Handle optimizations with care!
• Plot everything.
• Check fitted parameters for ...
  • Plausible values
  • Fitted values at extremes of feasible intervals (a red flag)
• Rethink the dynamic hypothesis in light of optimization results.
Ultimate Conclusion Here

• Investigate Sliding Goals in School Texts and SAT verbal scores with ...
  • A richer, more detailed sequence of models
  • More of the actual processes involved in writing and choosing school texts
  • More of the actual processes involved in developing SAT tests.
  • And more hypotheses for the observed SAT dynamics.
• That is, do a serious study of the Dynamics of School Texts and Student Verbal Abilities
Appendices

Data on Declining SAT Verbal Scores and Declining Sophistication of Texts
Sources


LEX Scores of School Readers, Grades 1 - 8

Figure 3. Mean LEX levels for school readers: 1919-1945, 1946-1962
LEX Scores of School Readers in Two Time Periods, Grades 1 - 8

Figure 4. Mean LEX levels for school readers: 1919-1945, 1963-1991
<table>
<thead>
<tr>
<th>Text Source</th>
<th>Date/N</th>
<th>LEX</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Nature</em> — an article on transhydrogenases</td>
<td>1960</td>
<td>58.6</td>
</tr>
<tr>
<td><em>New England Journal of Medicine</em> — articles</td>
<td>1990</td>
<td>33.3</td>
</tr>
<tr>
<td><em>Scientific American</em> — articles</td>
<td>1991</td>
<td>14.3</td>
</tr>
<tr>
<td><em>Popular Science</em> — articles</td>
<td>1994</td>
<td>4.6</td>
</tr>
<tr>
<td><em>Time</em> — articles</td>
<td>1994</td>
<td>1.6</td>
</tr>
<tr>
<td>Newspapers: English, N = 61 International</td>
<td>1665-1994</td>
<td>0.0</td>
</tr>
<tr>
<td><em>National Geographic</em> — articles</td>
<td>1984</td>
<td>-0.6</td>
</tr>
<tr>
<td><em>Sports Illustrated</em> — articles</td>
<td>1994</td>
<td>-10.3</td>
</tr>
<tr>
<td>Adult books — fiction, USA</td>
<td>N = 34</td>
<td>-15.8</td>
</tr>
<tr>
<td>The funnies — in newspapers</td>
<td>1982</td>
<td>-21.6</td>
</tr>
<tr>
<td>Nancy Drew mystery series</td>
<td>N = 69</td>
<td>-23.4</td>
</tr>
<tr>
<td>Comic books — GB &amp; USA</td>
<td>N = 37</td>
<td>-23.7</td>
</tr>
<tr>
<td>Children’s books age 10 — 14, GB</td>
<td>N = 261</td>
<td>-24.3</td>
</tr>
<tr>
<td>TV — cartoon shows</td>
<td>N = 26</td>
<td>-28.6</td>
</tr>
<tr>
<td>Children’s books age 9 — 12, USA</td>
<td>N = 94</td>
<td>-29.0</td>
</tr>
<tr>
<td>TV — reruns — popular with children</td>
<td>N = 33</td>
<td>-35.3</td>
</tr>
<tr>
<td>TV — primetime shows</td>
<td>N = 44</td>
<td>-36.4</td>
</tr>
<tr>
<td>Preschool books read to children</td>
<td>N = 31</td>
<td>-37.0</td>
</tr>
<tr>
<td>Mother’s talk to children, age 5</td>
<td>N = 32</td>
<td>-45.8</td>
</tr>
<tr>
<td>Dairy farmer talking to his cows</td>
<td>1988</td>
<td>-56.0</td>
</tr>
<tr>
<td>Pre-primer — Scott, Foresman</td>
<td>1956</td>
<td>-80.5</td>
</tr>
</tbody>
</table>
Milestones in Declining Difficulty of Texts

- Between 1860 and 1991, American publishers produced readers for the same grade at widely divergent LEX levels (e.g., in 1968, first grade readers were available between -68 and -31).
- The most difficult readers were generally published before 1918. By modern standards, Professor McGuffy’s pre- and post- Civil War readers were very difficult.
- After World War I, mean reader LEX levels for all grades were generally simplified.
- After World War II, mean levels of readers for all grades but third became even simpler. These were the books used by the Baby Boomers and successive cohorts.
Milestones in Declining Difficulty of Texts

- School publishers in Great Britain did not simplify their first grade readers after World War II, implying that there was no compelling educational reason for American publishers to further simplify their readers.
- Today’s mean sixth, seventh, and eighth grade readers are simpler than fifth grade readers were before World War II.
- Sentences were also shortened, from about 20 words before World War II to about 14 words now in Grades 4—8.
But Science Journals Have Been Getting Tougher
Mean Sentence Length of Texts

Figure 5. Mean length of sentence in school readers: 1919-1991