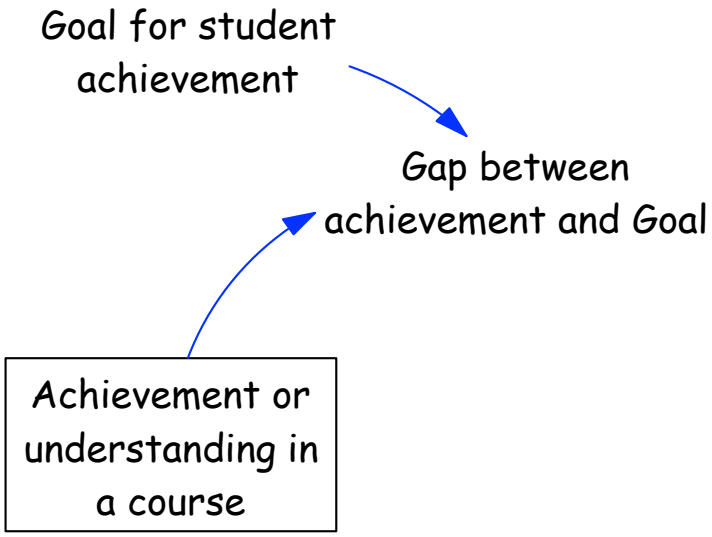
Fixed and Sliding Goals in Education?

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

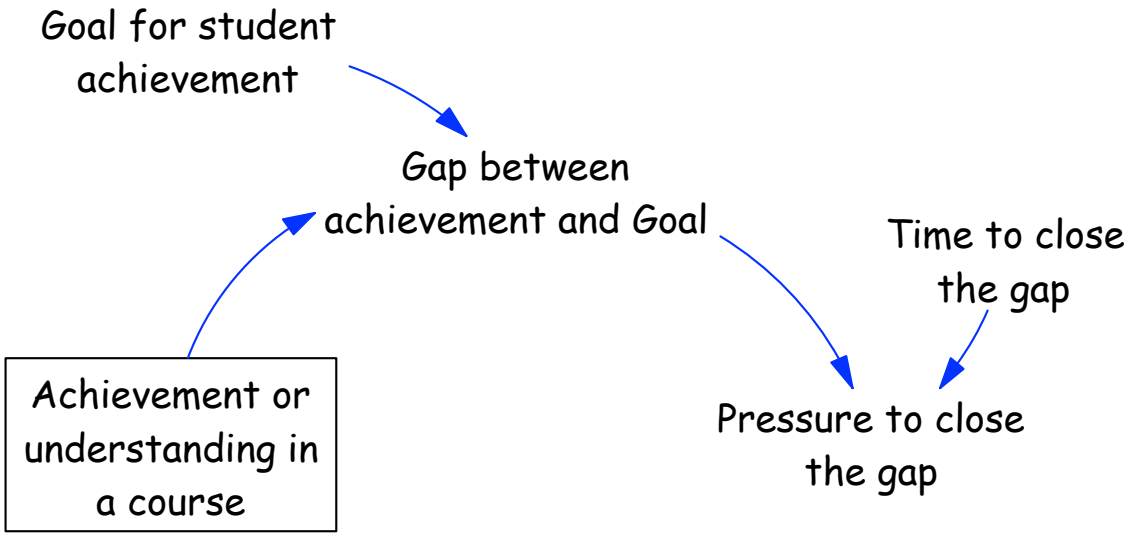
with a surprising, instructive error

Striving to Reach an Achievement Goal

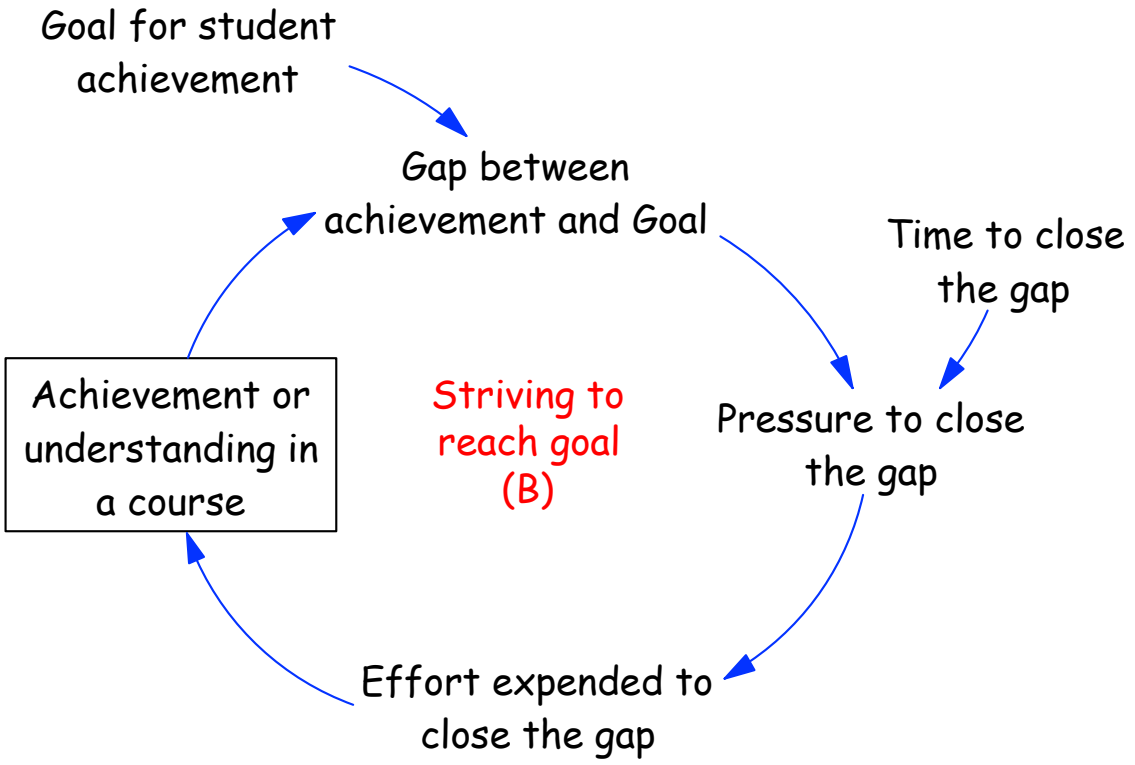


An adaptation of the sliding goals structure in Forrester (1968), "Market Growth as Influenced by Capital Investment"

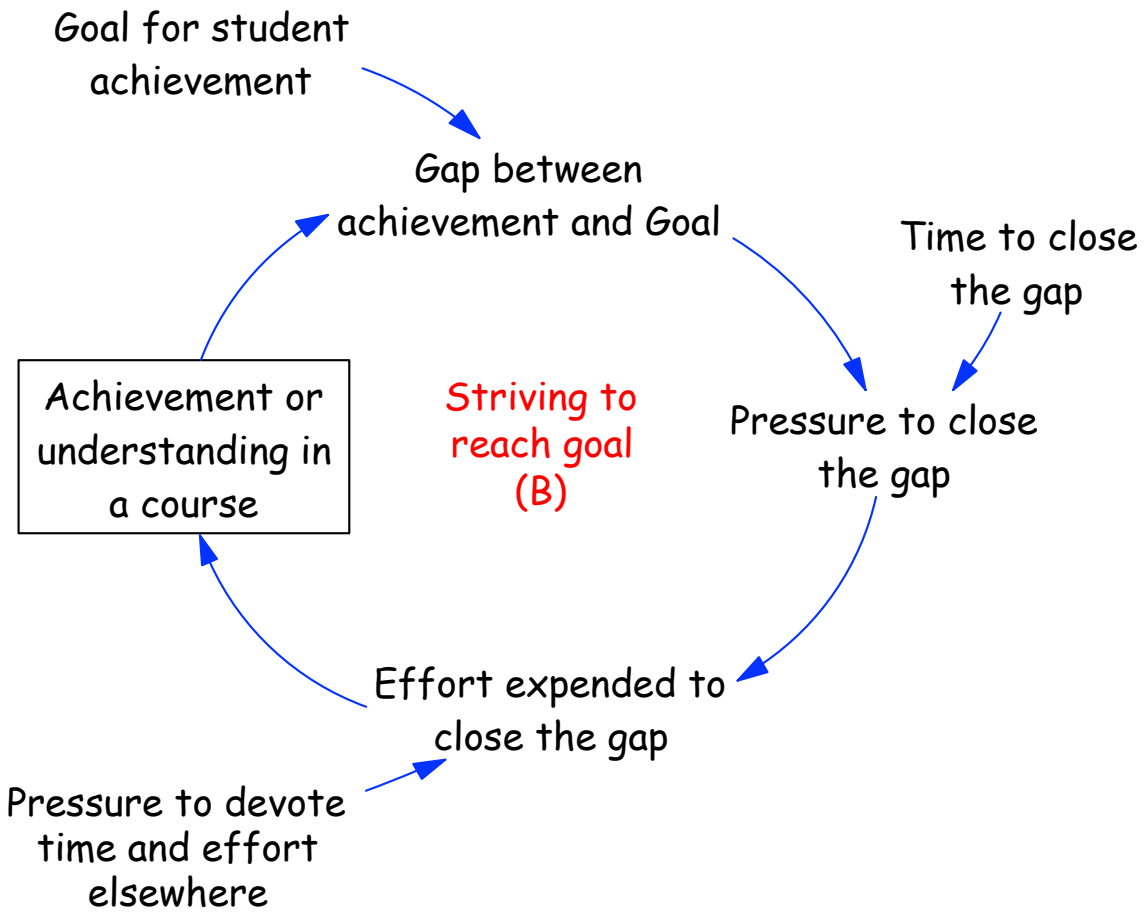
Striving to Reach an Achievement Goal



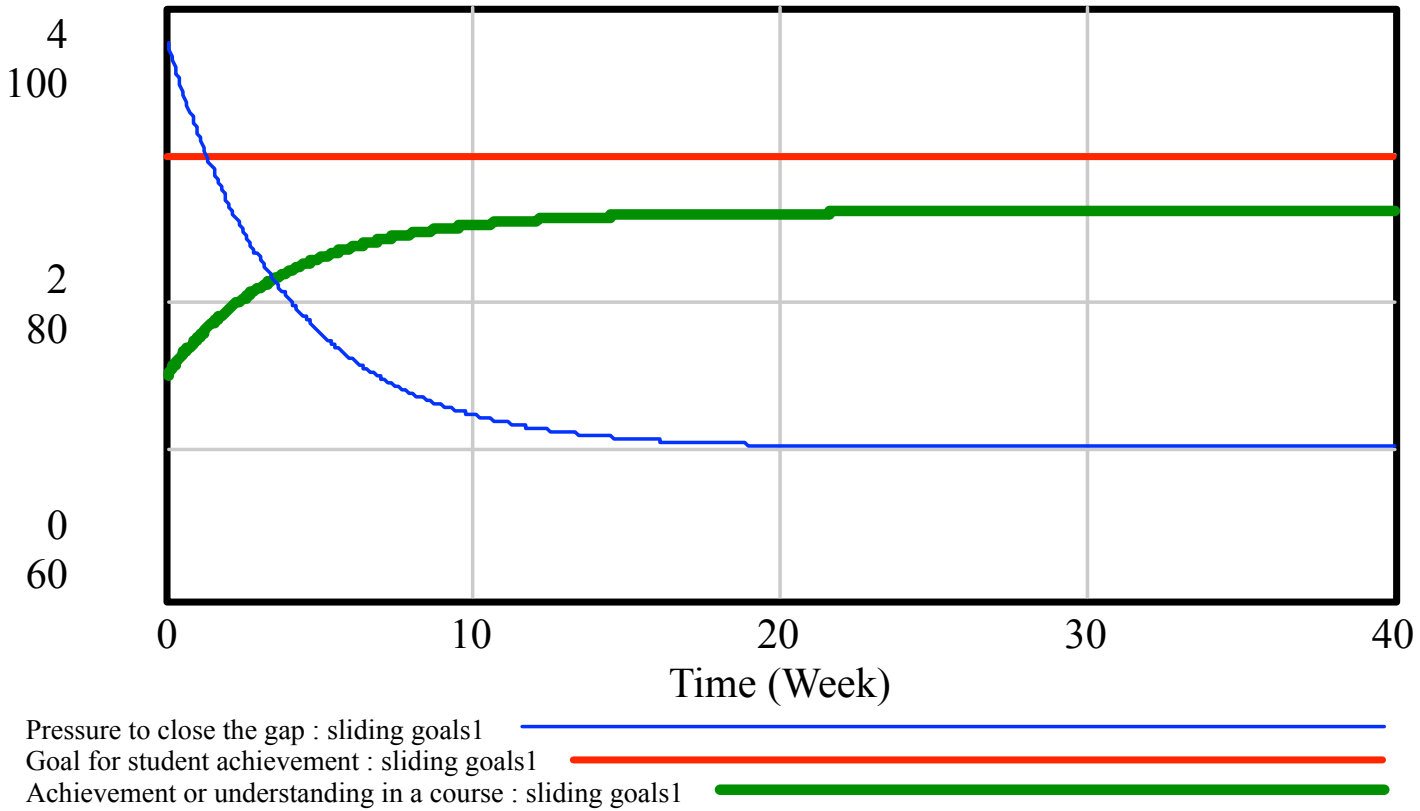
Striving to Reach an Achievement Goal



Striving to Reach an Achievement Goal

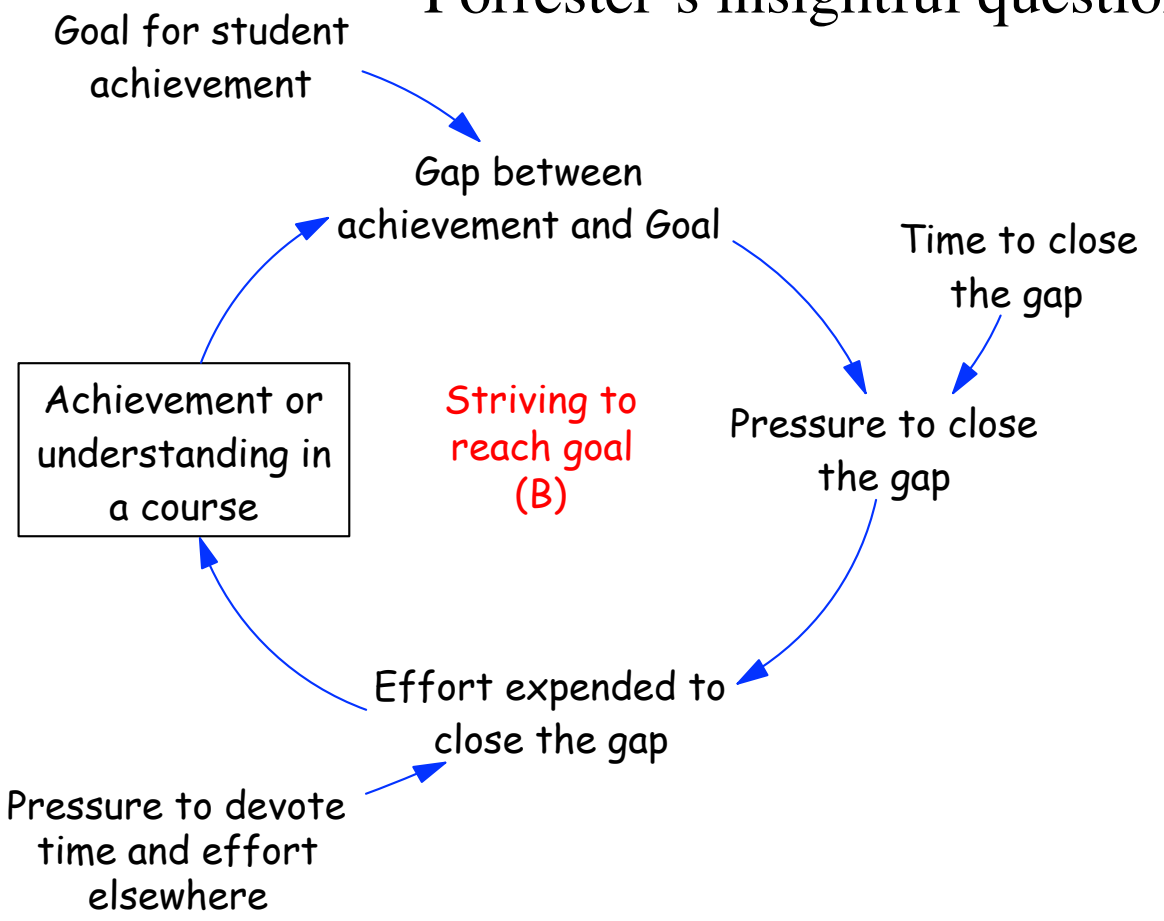


Striving (But Falling a Bit Short)



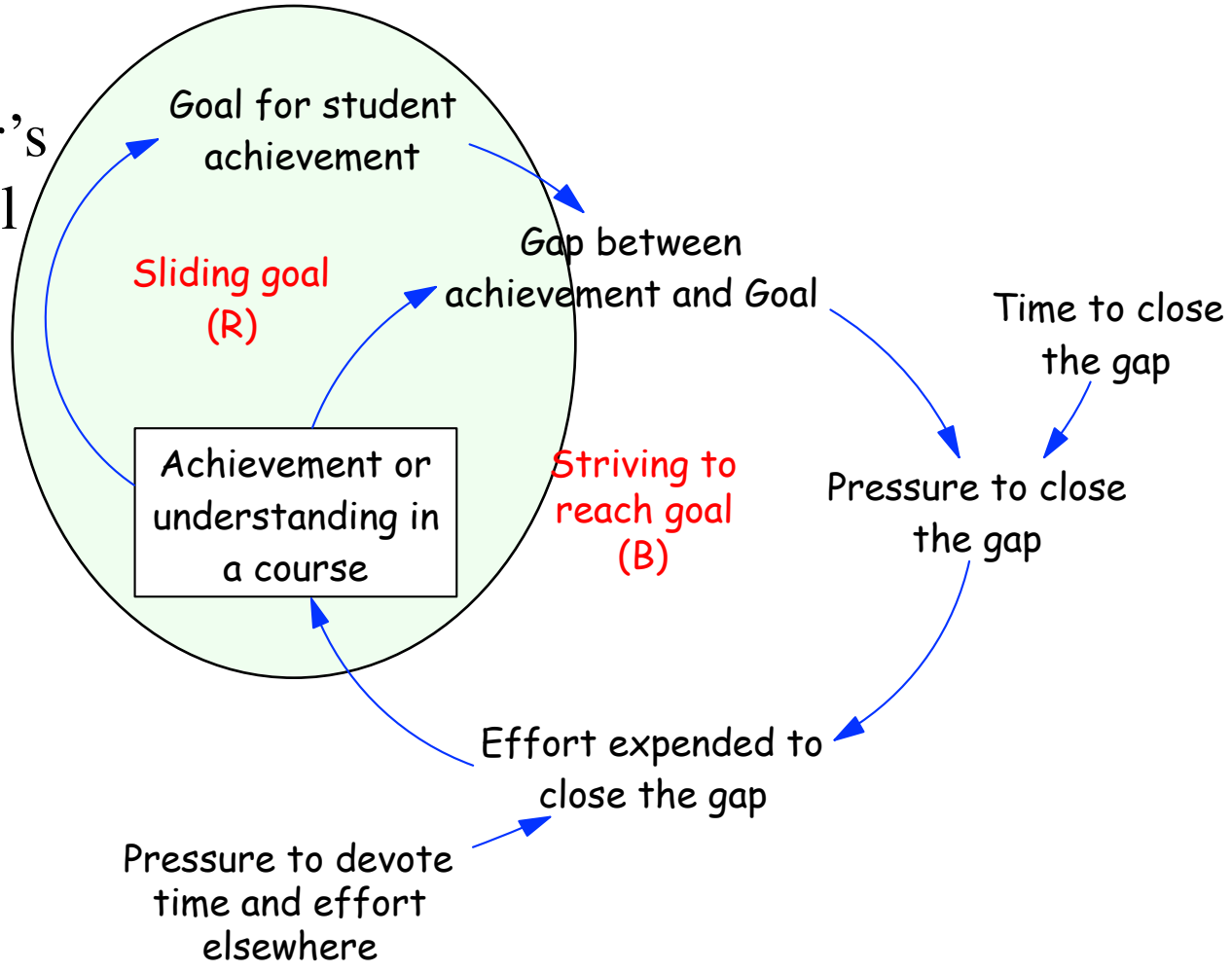
Where Does the Goal Come From?

Forrester's insightful question

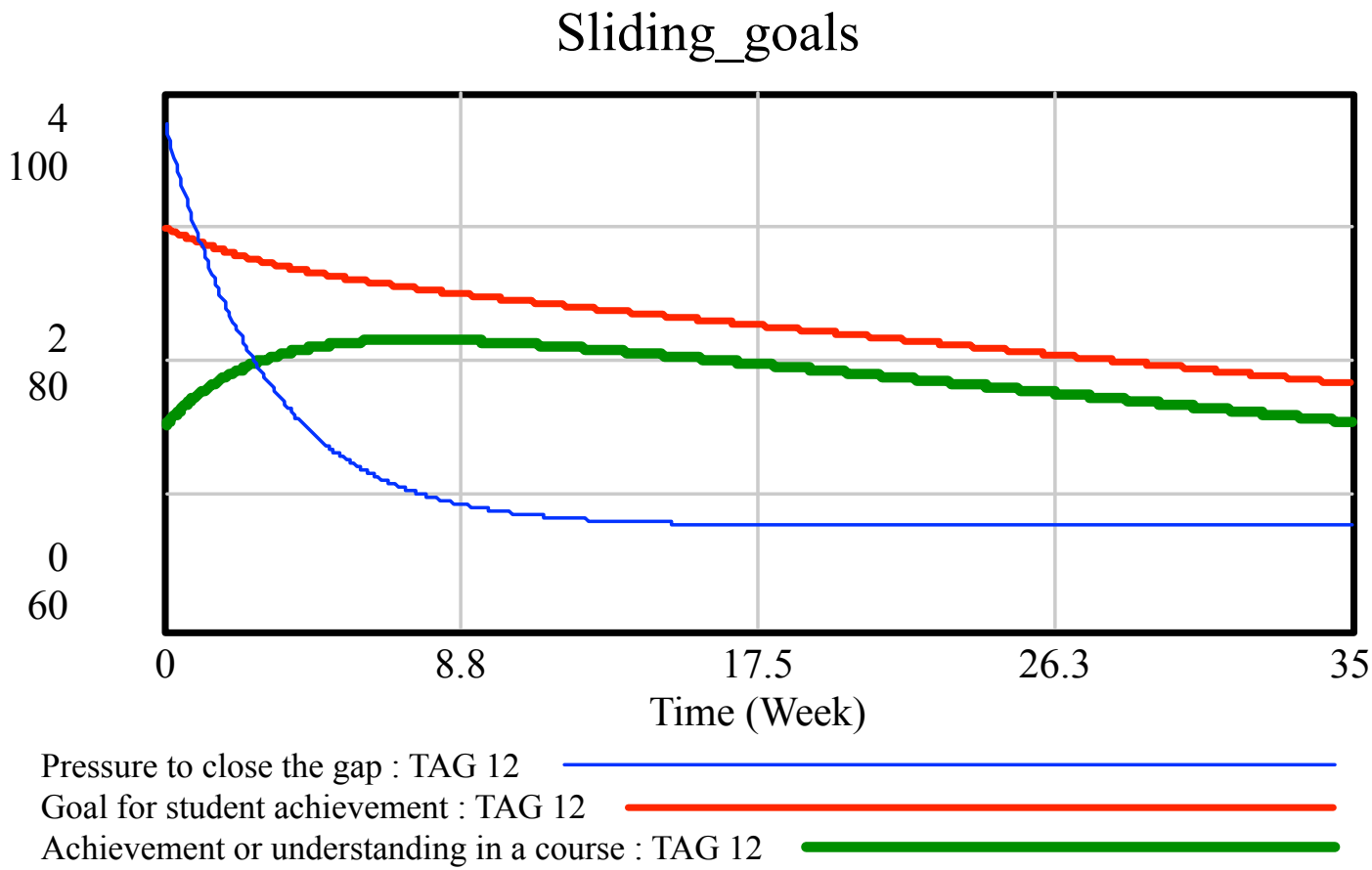


Where Does the Goal Come From?

Forrester's insightful answer



A Flexible Goal Can Slide



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?

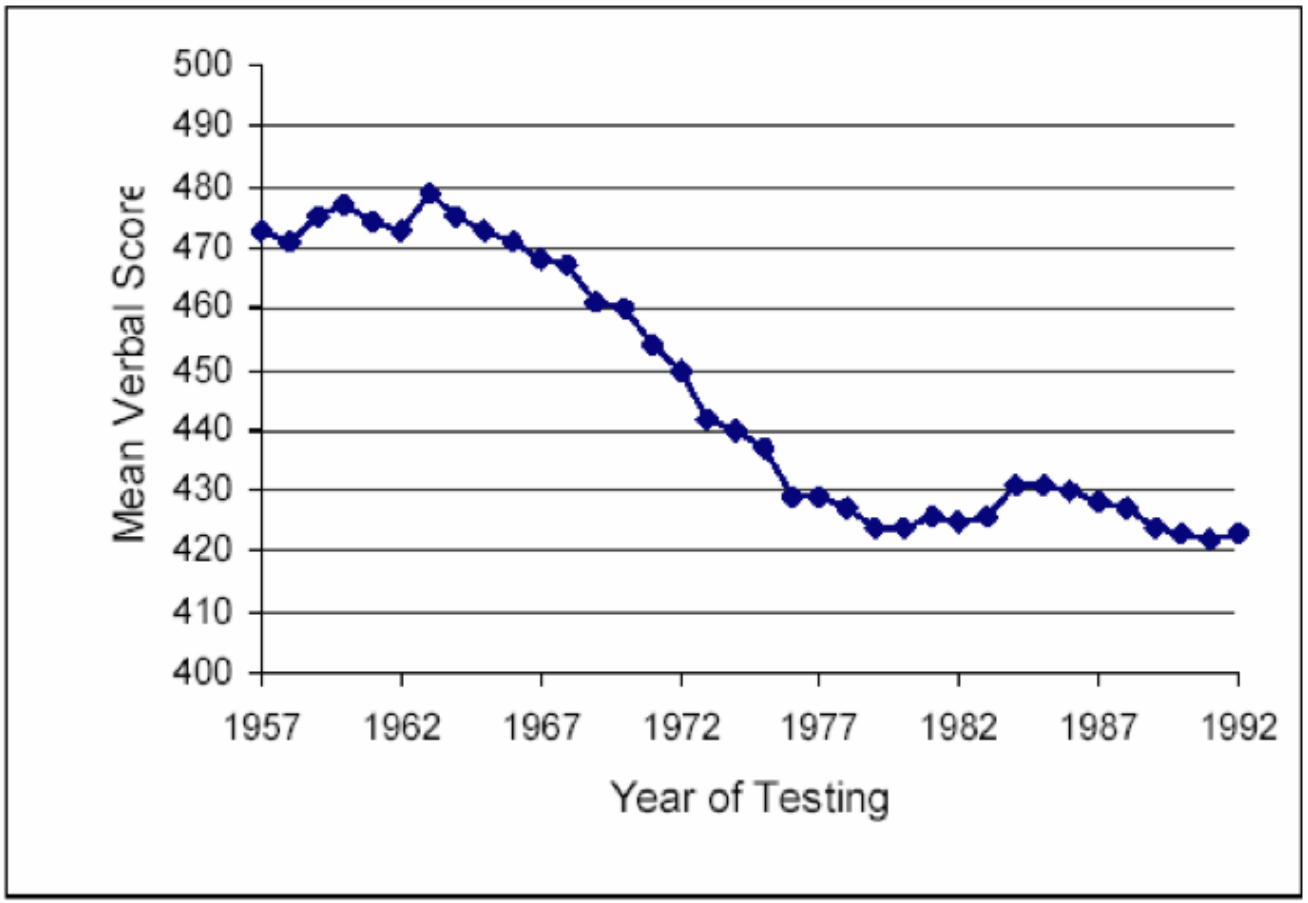
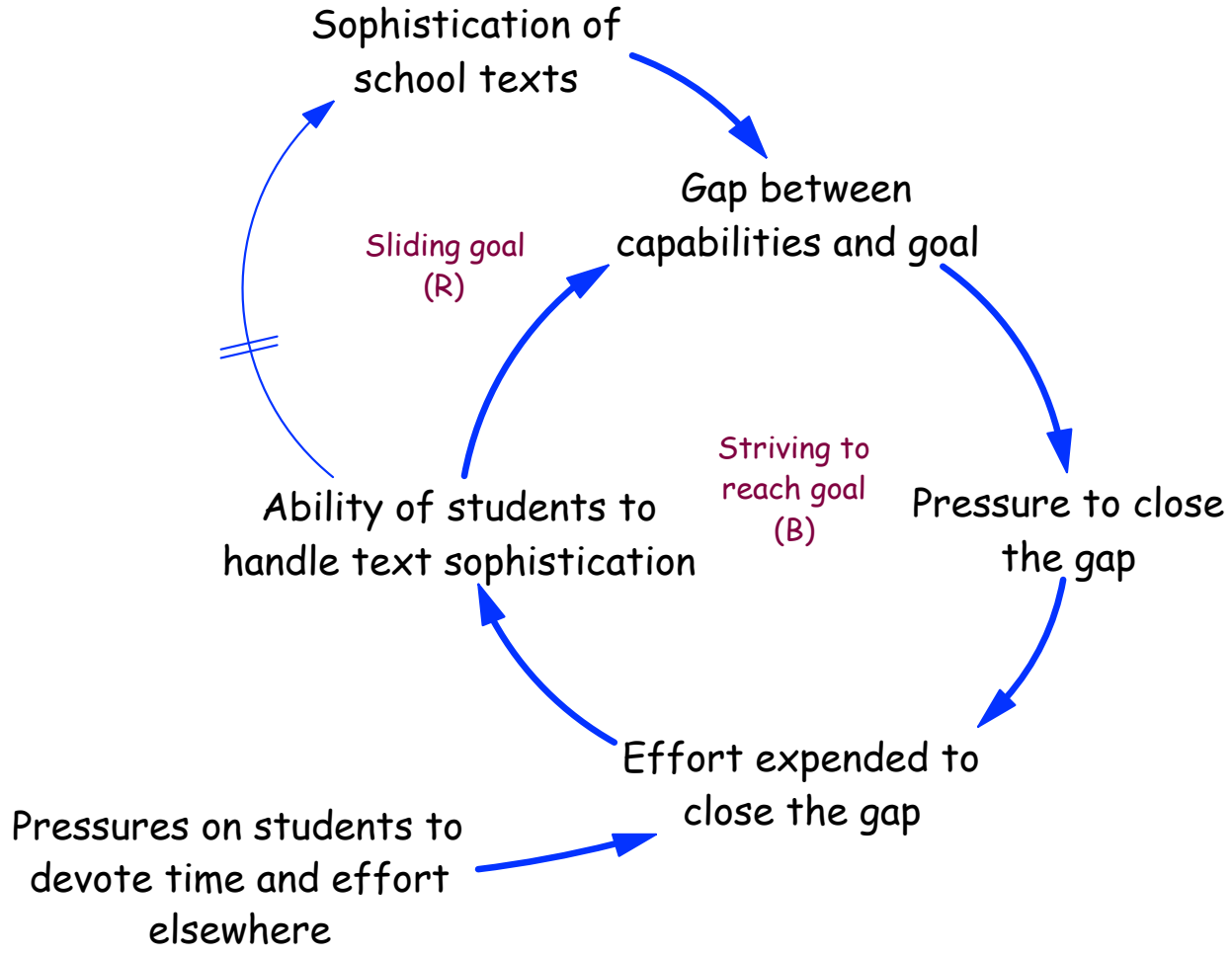
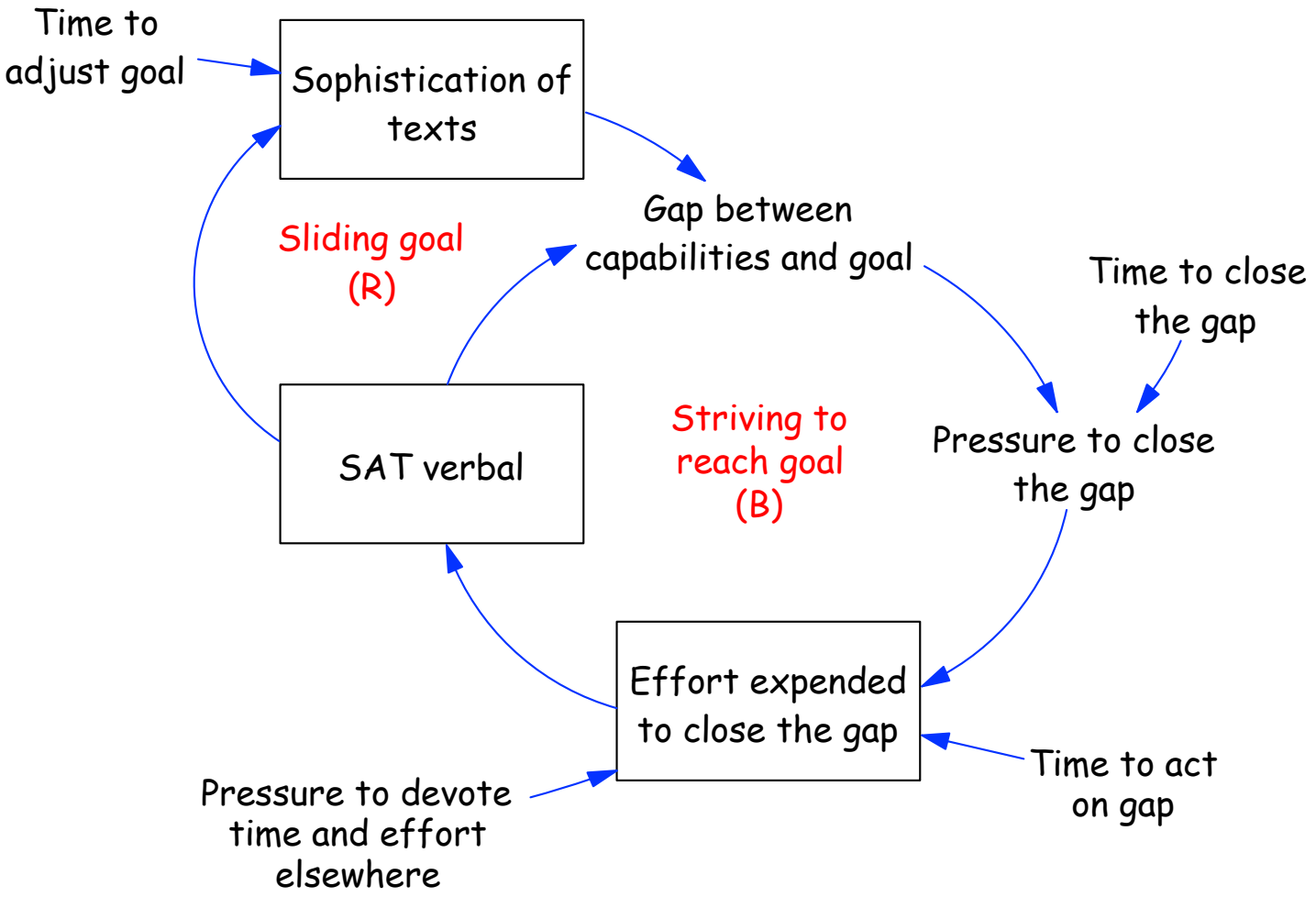


Figure 1: The SAT Verbal Time Series

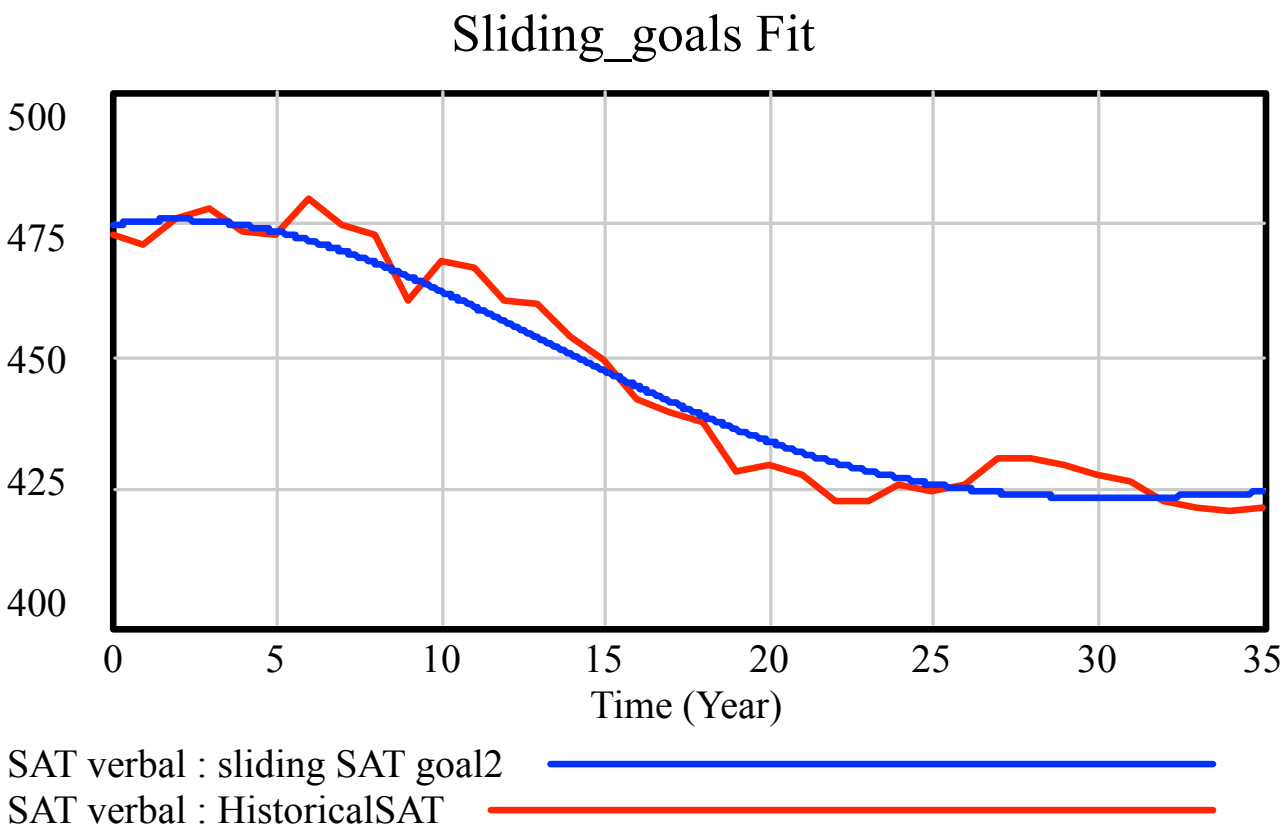
Sliding Goals in School Texts?



A Formal Model to Fit to the Data

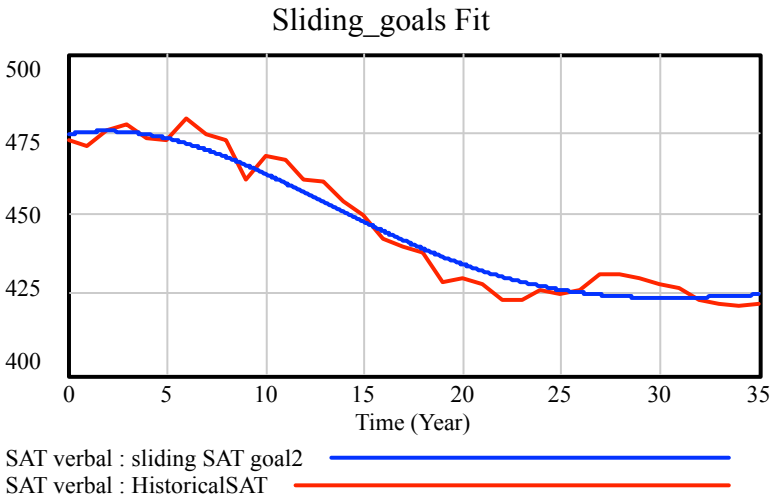


Historical and Simulated SAT Verbal Scores



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

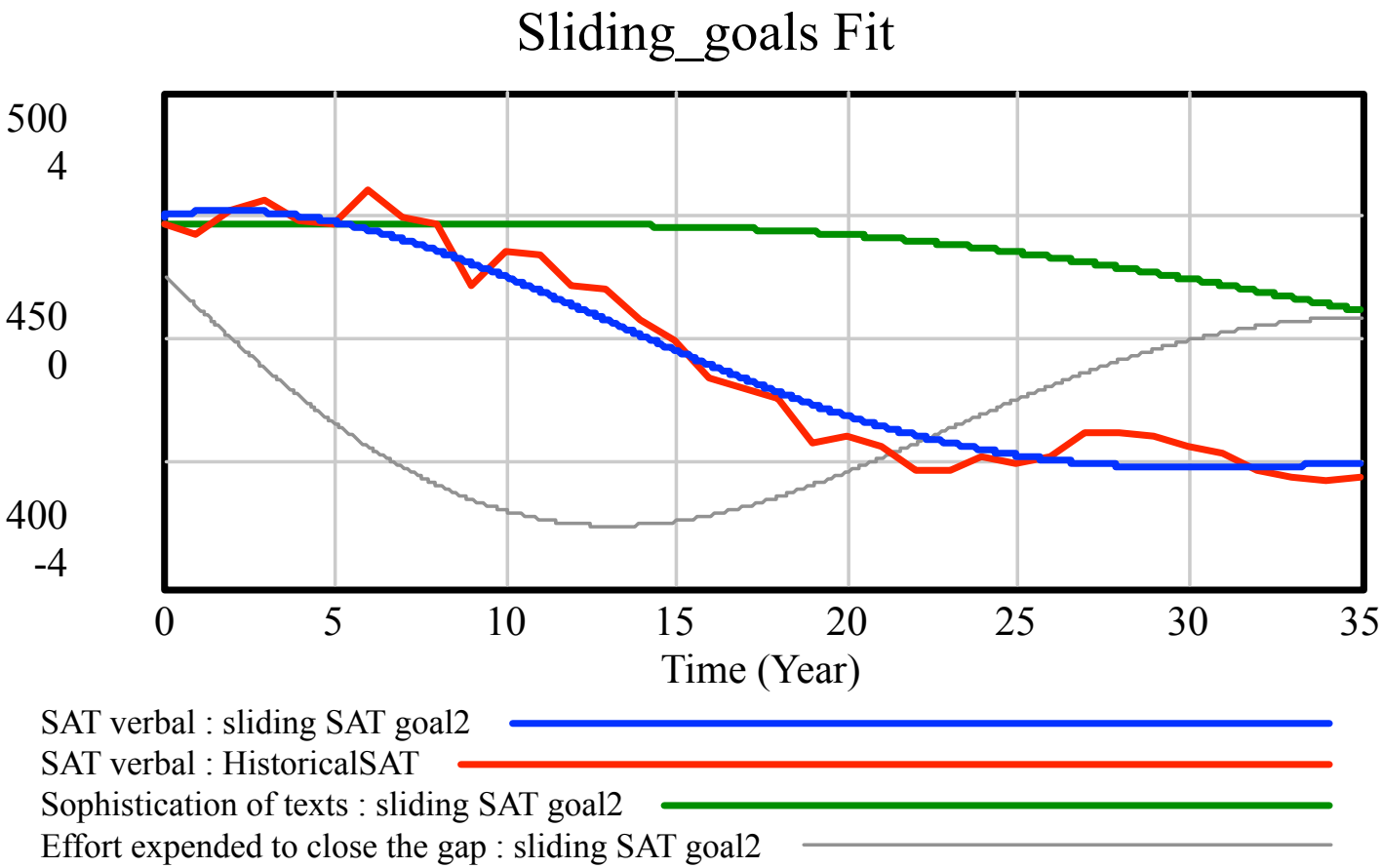
Optimal Parameters for SAT Fit

- 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?

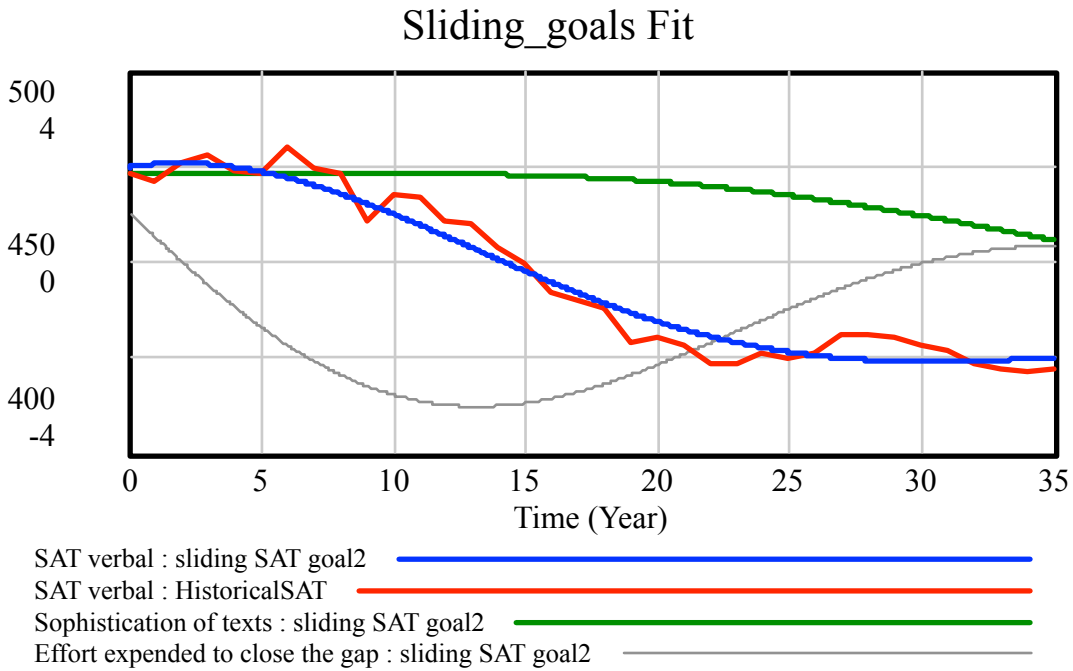
- Time to adjust goal = 38.0654 (years)
  - 38 years (almost two generations) to adjust the sophistication of texts seems reasonable.
- Time to act on gap = 17.2979 (years)
  - 17 years (a bit more than the twelve years of K-12 education) to change student tendencies seems reasonable.
- Pressure to devote 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



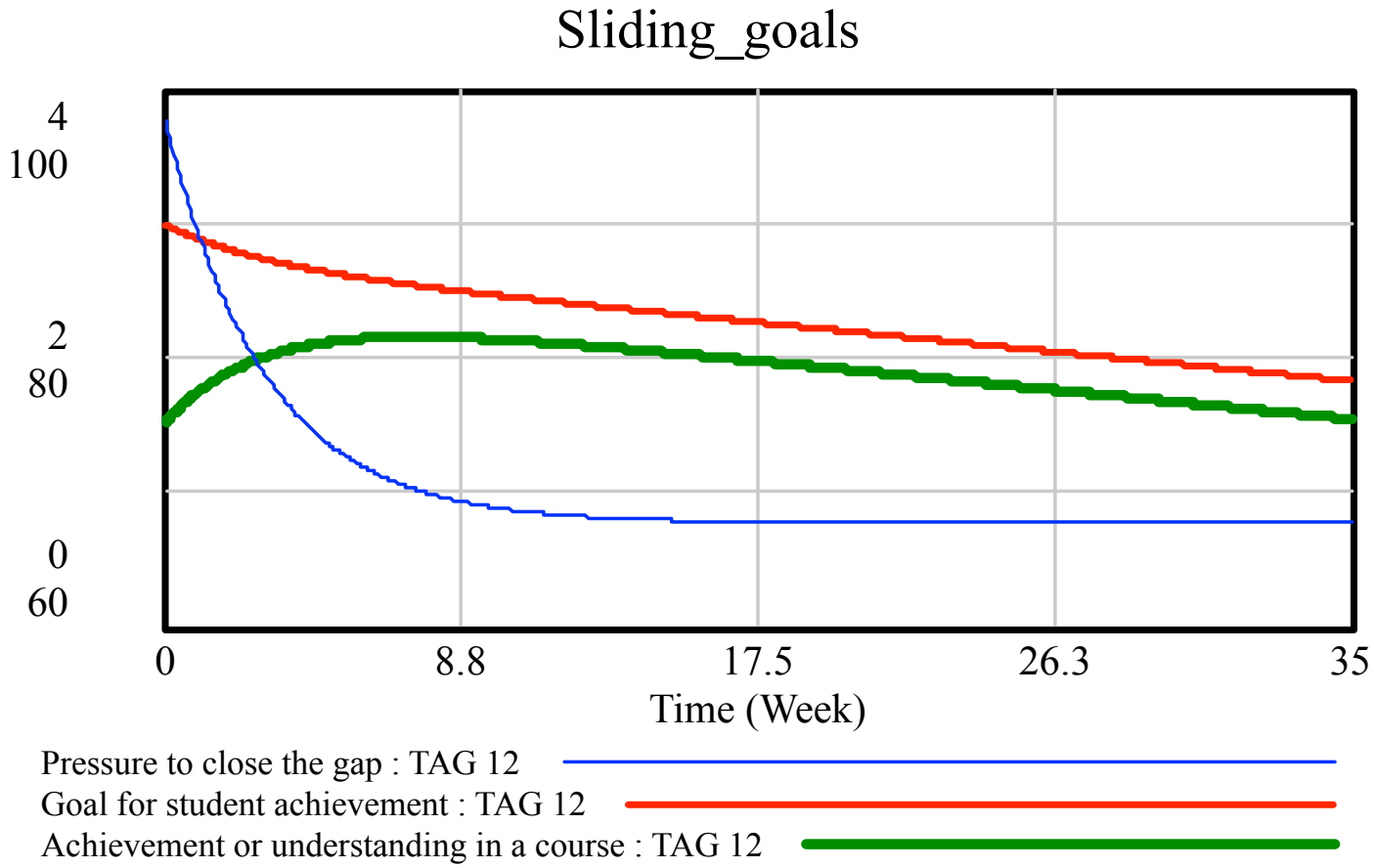
What do We See?

- *Effort* (grey) is appropriate to achieve the fit.
- But *Sophistication of Texts* (green) does not follow the path of SAT scores!



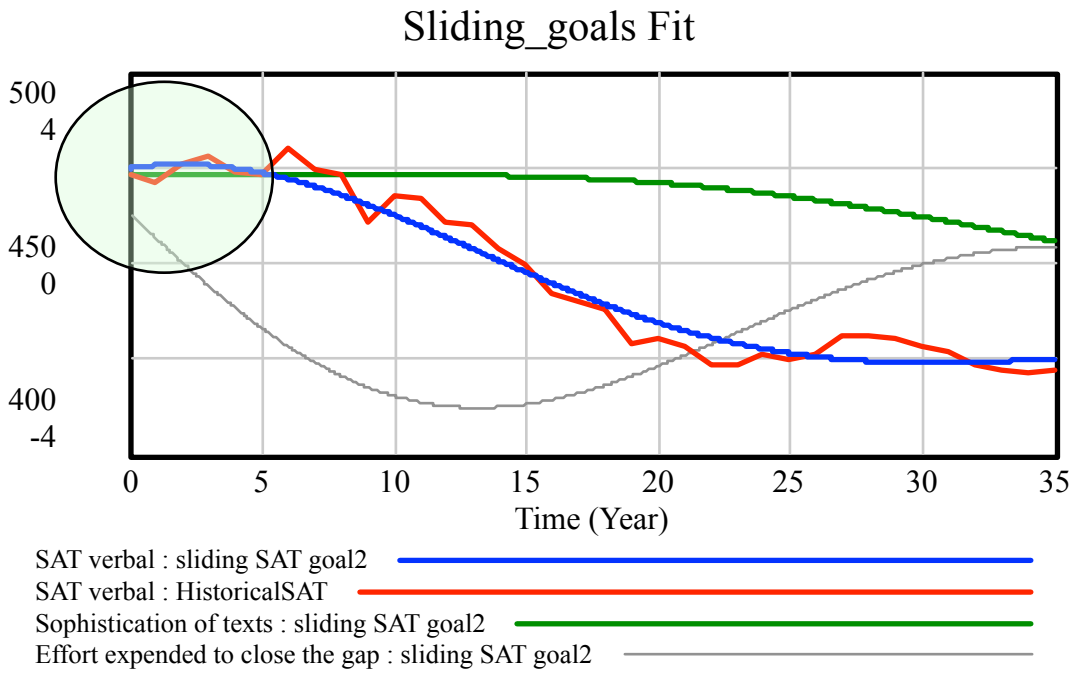
- Exerts upward pressure on SAT scores throughout the run.

Variable and Goal ought to Slide Together, as in the Student Achievement example:



Observations and Hypotheses

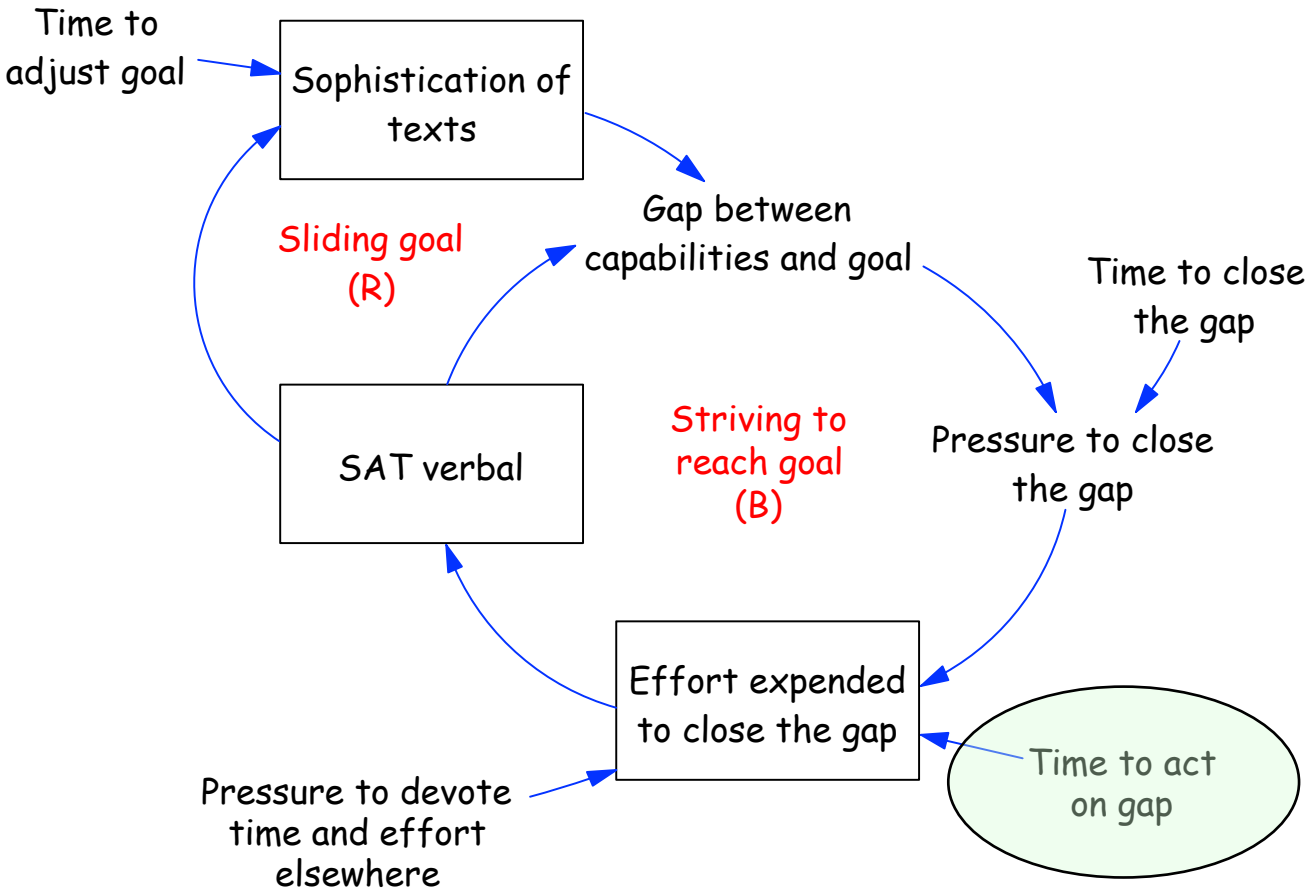
- Initial values of the stocks were selected by eye.
  - They were accidentally wonderful.



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

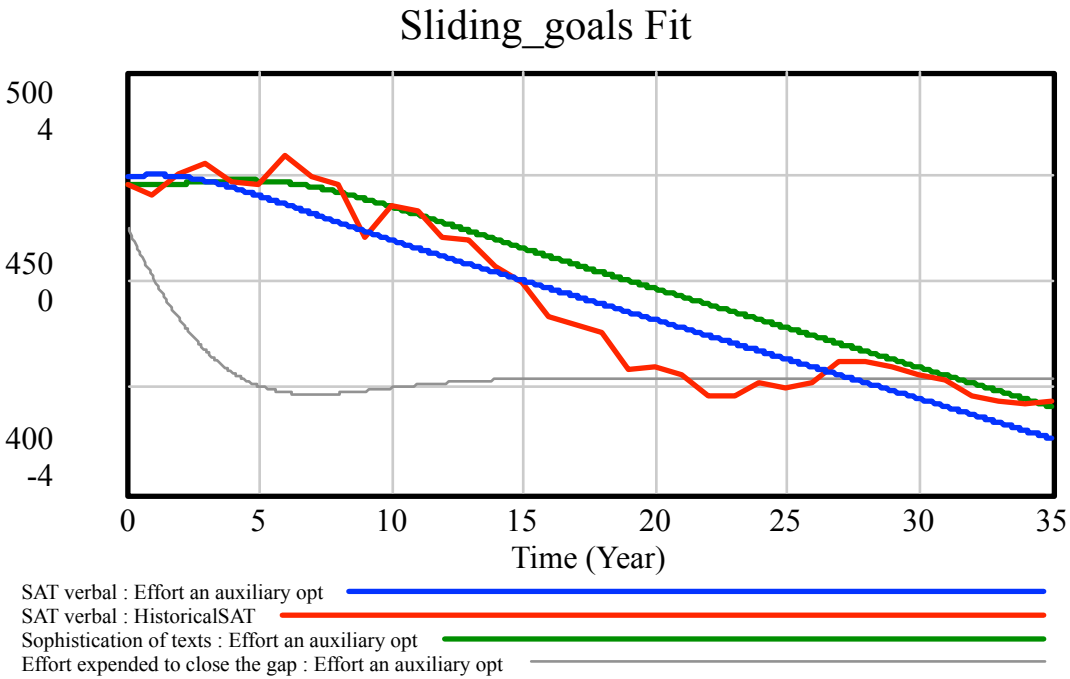


Revised Optimization Results, with Time to Close the Gap constant at 5

- 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?

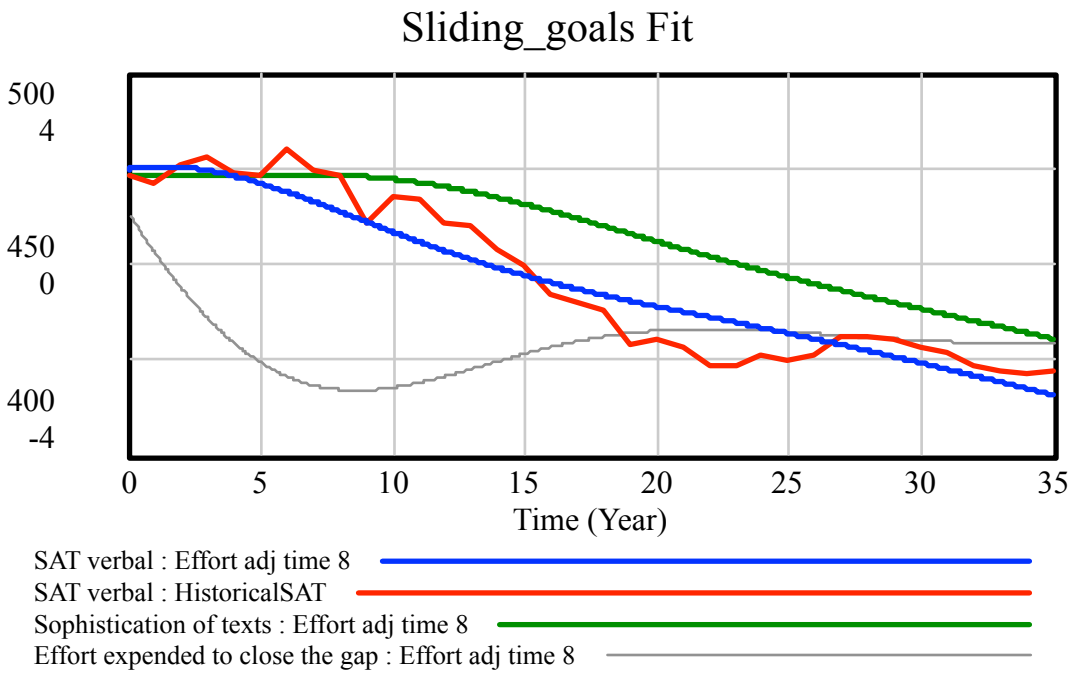
- *Effort* (grey) settles appropriately to a constant -2.
- *Sophistication of Texts* (green) now follows SAT scores (blue)



- 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

- Marilyn Jager Adams, Advancing our Students' Language and Literacy, The Challenge of Complex Texts, *The American Educator* **34,4** (winter 2010-11)
- Hayes, Wolfer & Wolfe (1996), Schoolbook Simplification and Its Relation to the Decline in SAT-Verbal Scores, *American Educational Research Journal 33 (2): 498-508*.
  - http://www.soc.cornell.edu/hayes-lexical-analysis/schoolbooks/Papers/HayesWolferAndWolf1996.pdf

LEX Scores of School Readers, Grades 1 - 8

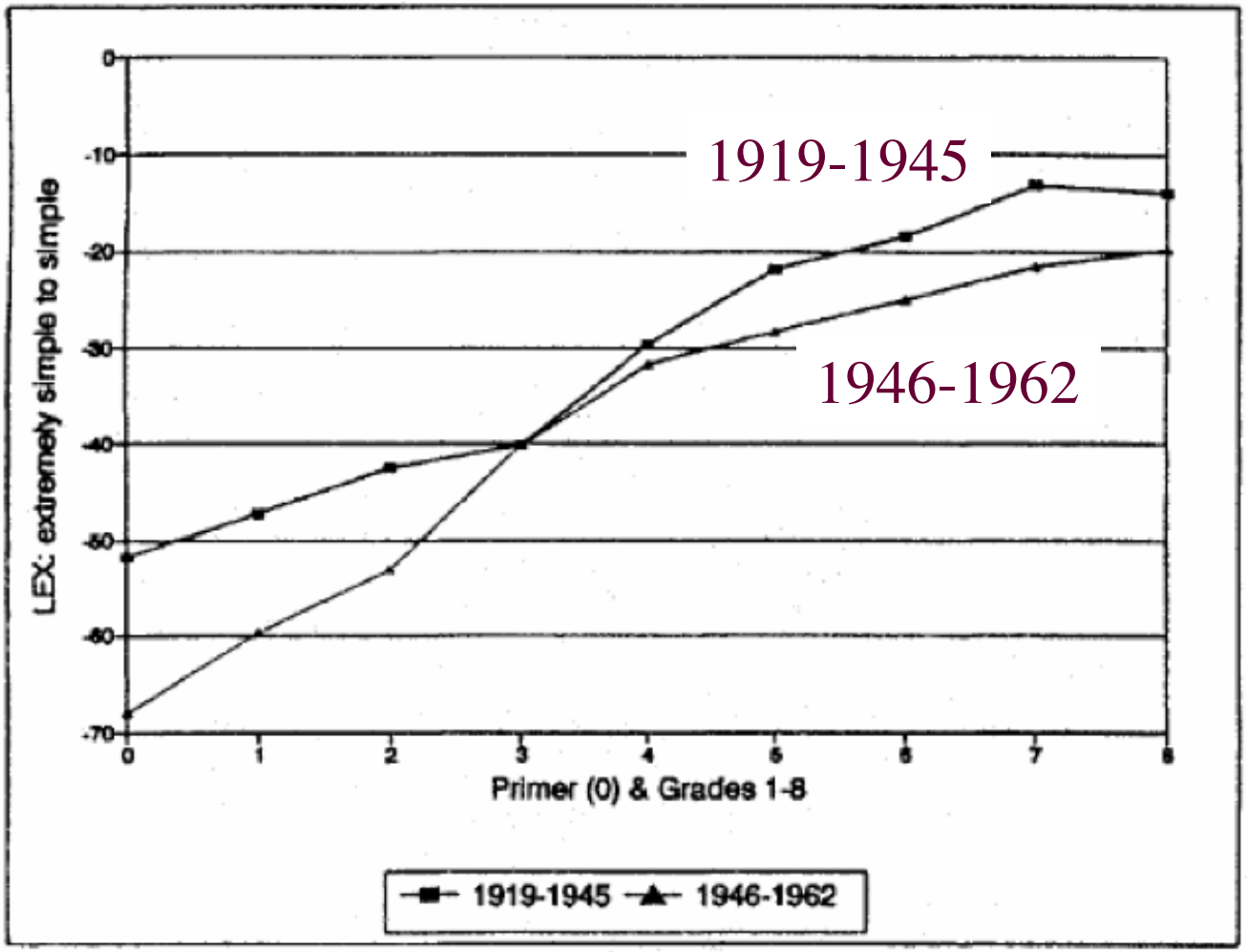


Figure 3. Mean LEX levels for school readers: 1919-1945, 1946-1962

LEX Scores of School Readers in Two Time Periods, Grades 1 - 8

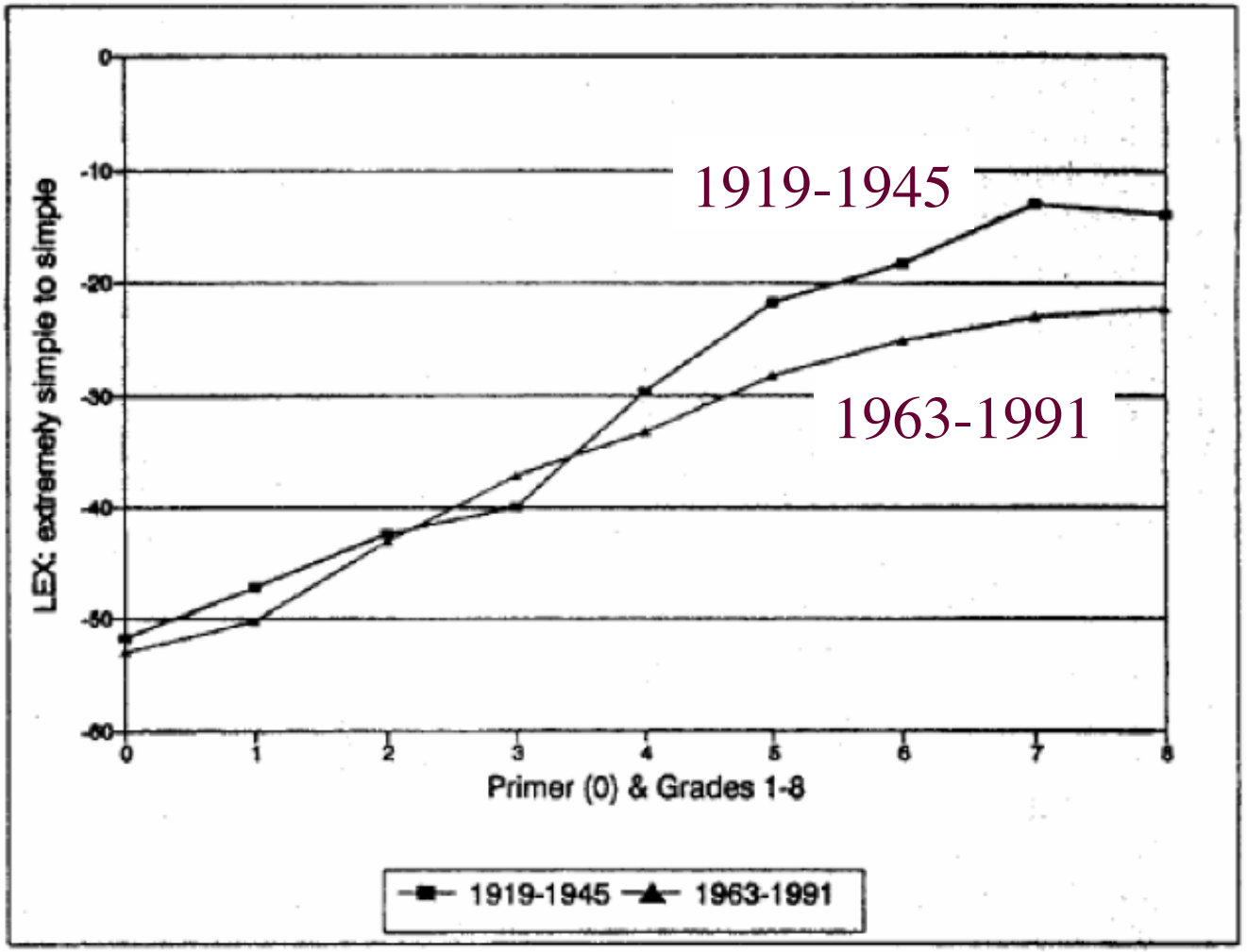


Figure 4. Mean LEX levels for school readers: 1919-1945, 1963-1991

Text Source	Date/N	LEX
Nature — an article on transhydrogenese	1960	58.6
New England Journal of Medicine — articles	1990	33.3
Scientific American — articles	1991	14.3
Popular Science — articles	1994	4.6
Time — articles	1994	1.6
Newspapers: English, N = 61 International	1665-1994	0.0
National Geographic — articles	1984	-0.6
Sports Illustrated — articles	1994	-10.3
Adult books — fiction, USA	N = 34	-15.8
The funnies — in newspapers	1982	-21.6
Nancy Drew mystery series	N = 69	-23.4
Comic books — GB & USA	N = 37	-23.7
Children's books age 10 — 14, GB	N = 261	-24.3
TV — cartoon shows	N = 26	-28.6
Children's books age 9 —12, USA	N = 94	-29.0
TV — reruns — popular with children	N = 33	-35.3
TV — primetime shows	N = 44	-36.4
Preschool books read to children	N = 31	-37.0
Mother's talk to children, age 5	N = 32	-45.8
Dairy farmer talking to his cows	1988	-56.0
Pre-primer — Scott, Foresman	1956	-80.5

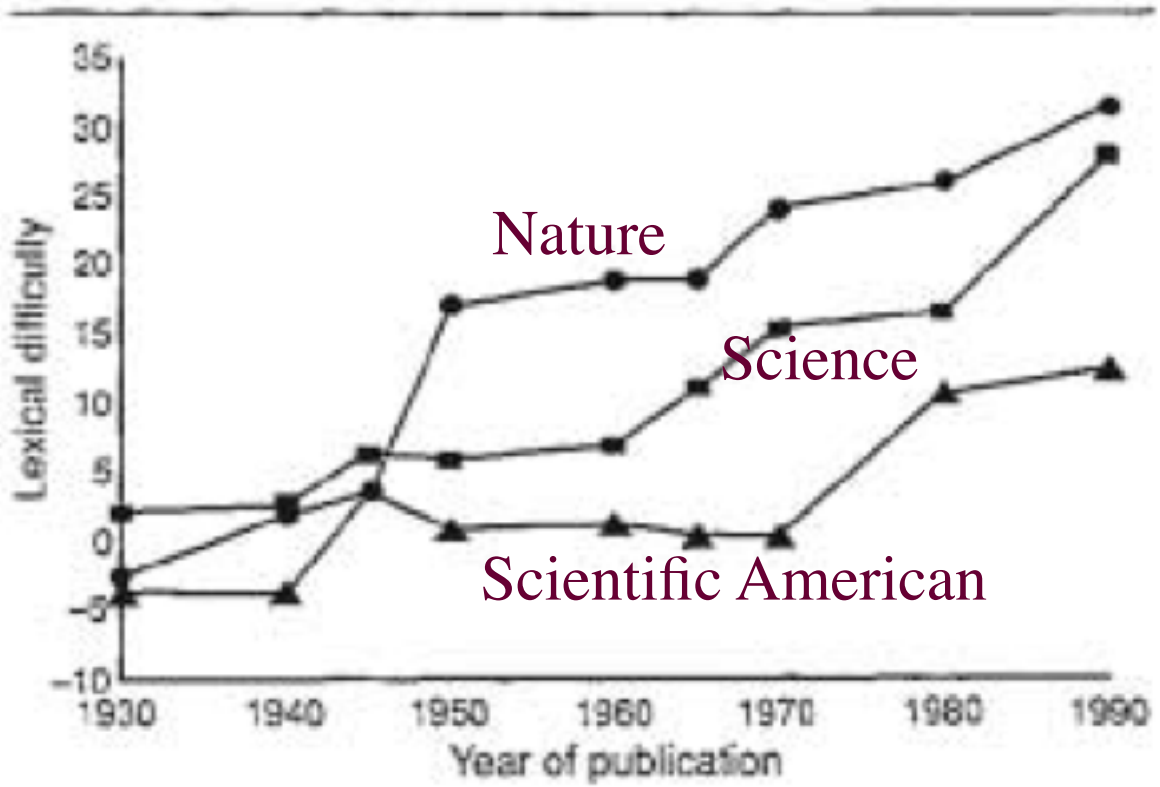
Milestones in Declining Difficulty of Texts

- 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 heir 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

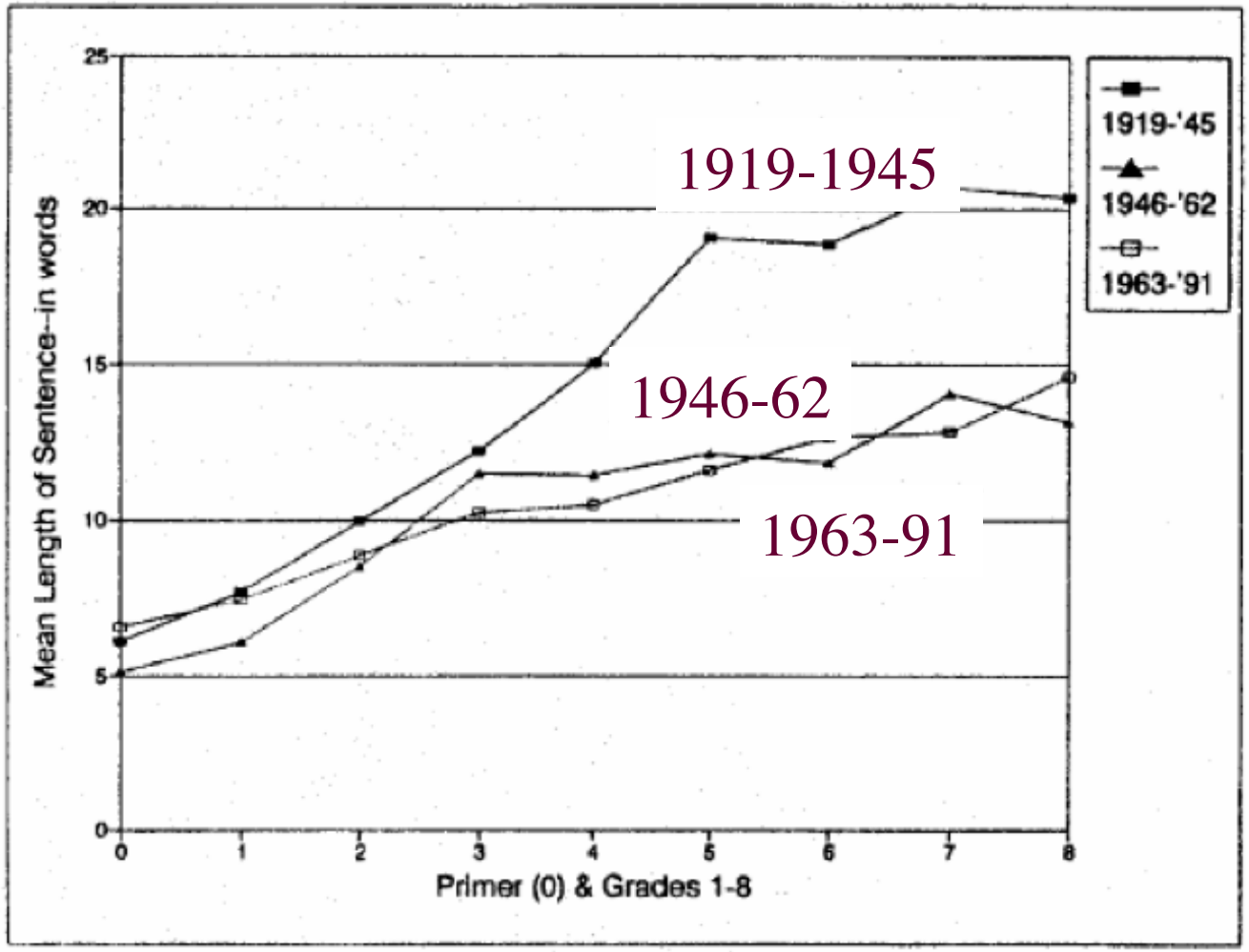


Figure 5. Mean length of sentence in school readers: 1919-1991