Carnegie Mellon

When the Rubber Meets the Road: Lessons from the In-School Adventures of an Automated Reading Tutor that Listens

Jack Mostow and Joseph Beck

Project LISTEN, Carnegie Mellon University

www.cs.cmu.edu/~listen

Funding: National Science Foundation

Carnegie Mellon

Outline

1. Ideal
2. Reality
3. Usage
4. Efficacy
5. Conclusion

Project LISTEN's Reading Tutor

John Rubin (2002). The Sounds of Speech (Show 3). On Reading Rockets (Public Television series commissioned by U.S. Department of Education). www.readingrockets.org. Washington, DC: WETA.

Carnegie Mellon

Project LISTEN's Reading Tutor (video)

Reading Tutor's continuous assessment

The Reading Tutor uses continuous assessment to:

- Adjust level of stories chosen and help given
- Report progress measures that teachers want

Sources of information:

- Clicking for help
- Latency before word
  - Initial encounter of **muttered**:
    - I'll have to mop up all this (5630) **muttered** Dennis to himself but how
  - 5 weeks later:
    - Dennis (110) **muttered** oh I forgot to ask him for the money
- Comprehension questions
  - Multiple-choice fill-in-the-blank
  - Automatic generation, scoring, and instant feedback

Scaling up the Reading Tutor, 1996→2003

Deployment

- Sites: 1 school → 9 (diverse!) schools
- Students: N=8 → N>800 (including control groups)
- Grade levels: grade 3 → grades K-4+
- Computers: ours → school-owned (Windows 2000/XP)
- Installation: manual → InstallShield/clone
- Configuration: standalone → client-server + web-based reports

Supervision

- Setting: individual pullout → classroom, lab, specialist room
- User training: individual → automated
- Assessment/leveling: none → automated

Carnegie Mellon

Traditional instruction tries to move whole class together

Carnegie Mellon

Technology can free students to progress at their own pace

Carnegie Mellon

Evaluate against alternatives!

Gains from pre- to post-test

But teachers help too.

So compare to control(s)!

Carnegie Mellon

Results of Pre- to Post-Test Evaluations: Mounting Evidence of Superior Gains

1996, grade 3, N=6 lowest readers: gained 2 years in 8 mos.

1998, gr. 2-5, N=63: outgained classmates in comprehension

1999, gr. 2-3, N=131: vocabulary gains rivaled human tutors

2000, gr. 1-4, N=178: outgained independent practice

2001, gr. 1-4, N ~ 600: room gains correlated with usage

2002, gr. K-4, N ~ 600: still analyzing data

2003, gr. 1-3, N ~ 800: studies starting at 8 schools

See www.cs.cmu.edu/~listen for publications, effect sizes, …

Carnegie Mellon

"The history of AI is littered with the corpses of promising ideas" [A. Newell]

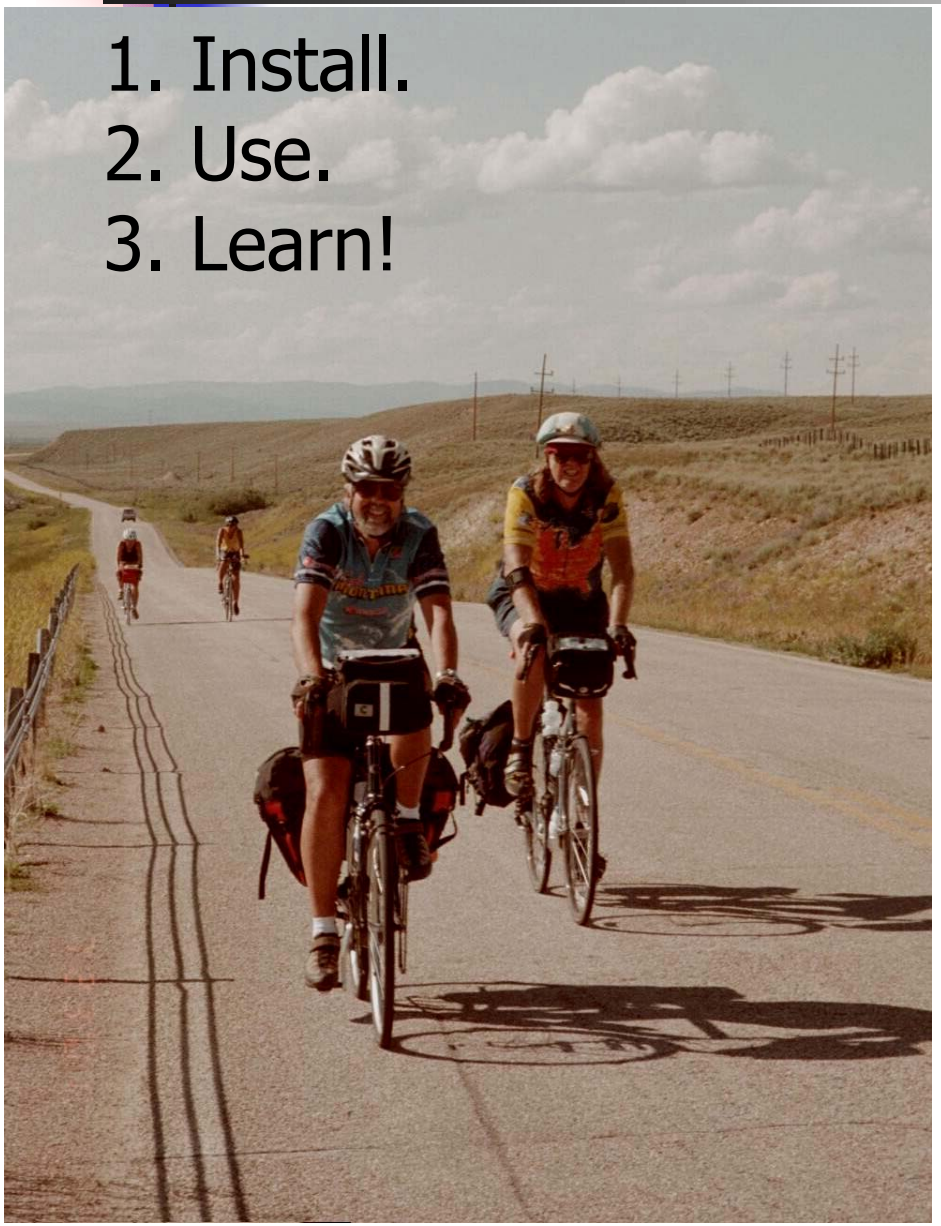
Carnegie Mellon

From the teacher's perspective

- Technology as burden
- Shared resource makes scheduling more difficult

Carnegie Mellon

The Ideal:



The Reality:

Carnegie Mellon

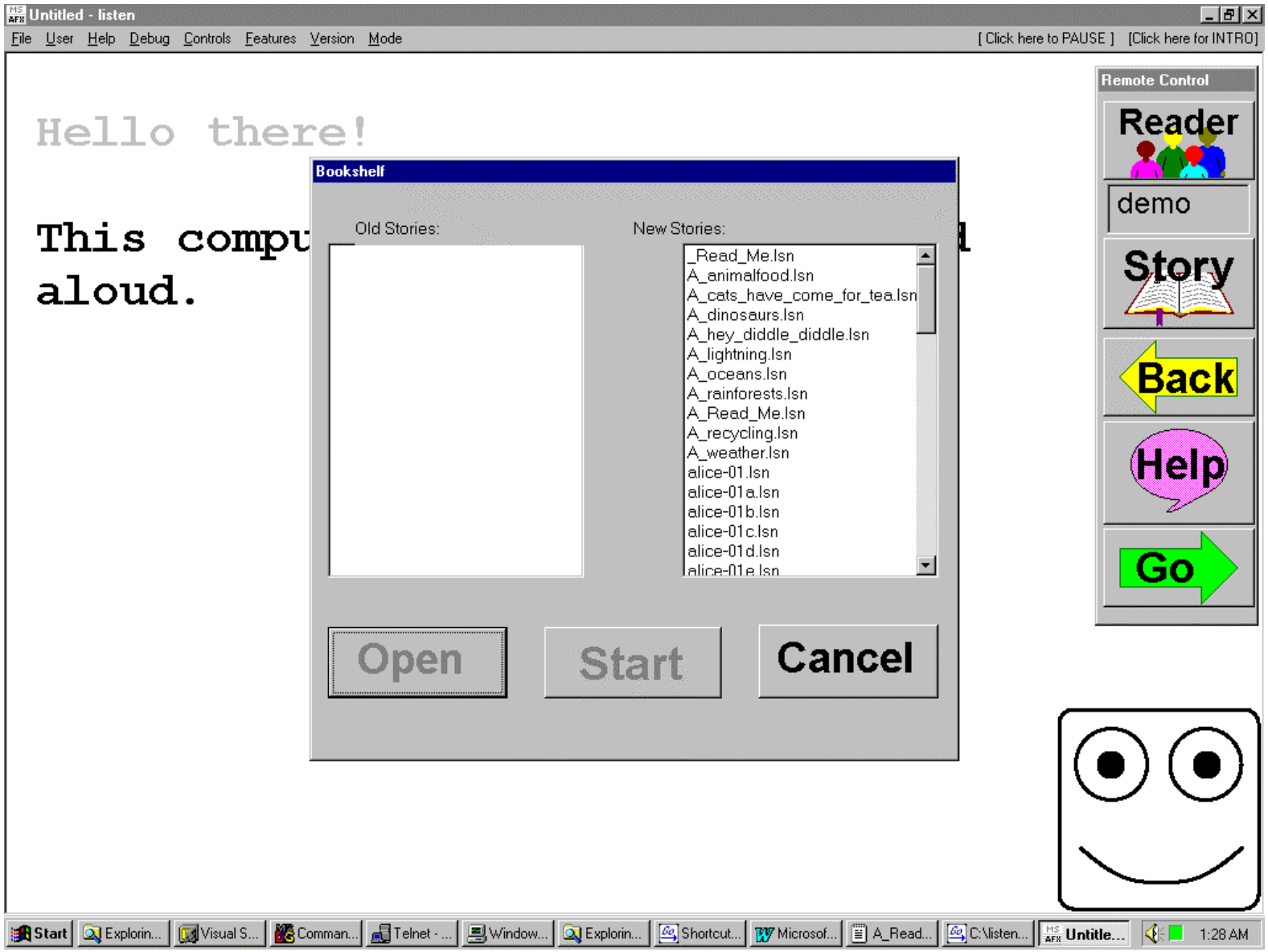
Why design iteration must field-test: Features revealed in new settings

Crash!

Score!

Escape!

Riot!



Carnegie Mellon

Usage: how much student uses Tutor

What might influence usage (directly or indirectly)?

- Student: attitude, attendance
- Tutor: reliability, usability, reports
- Teacher: schedule, attitude, organization
- Setting: classroom? lab? specialist? resource room?
- School: policy, schedule, supportiveness
- Support: training, repair time

How can we measure such influences?

- Observer effects: teachers put kids on when we visit.
- So instrument: Reading Tutor sends back data nightly.

Carnegie Mellon

2002-2003 data by setting and grade

Most students were in grades 1-3:

Setting:	K	1	2	3	4	5	6
Classroom	14	**52**	**73**	**40**	20		
Lab		**72**	**66**	**40**			
Resource		**12**	**3**				
Specialist		**2**	**4**		2	2	7

(cell values = # of students)

Carnegie Mellon

Reading Tutor in a classroom setting

Carnegie Mellon

Reading Tutor in a lab setting

Carnegie Mellon

How 2002-2003 usage varied by setting

Frequency

- How often a student uses the Reading Tutor (% of possible days)

Duration

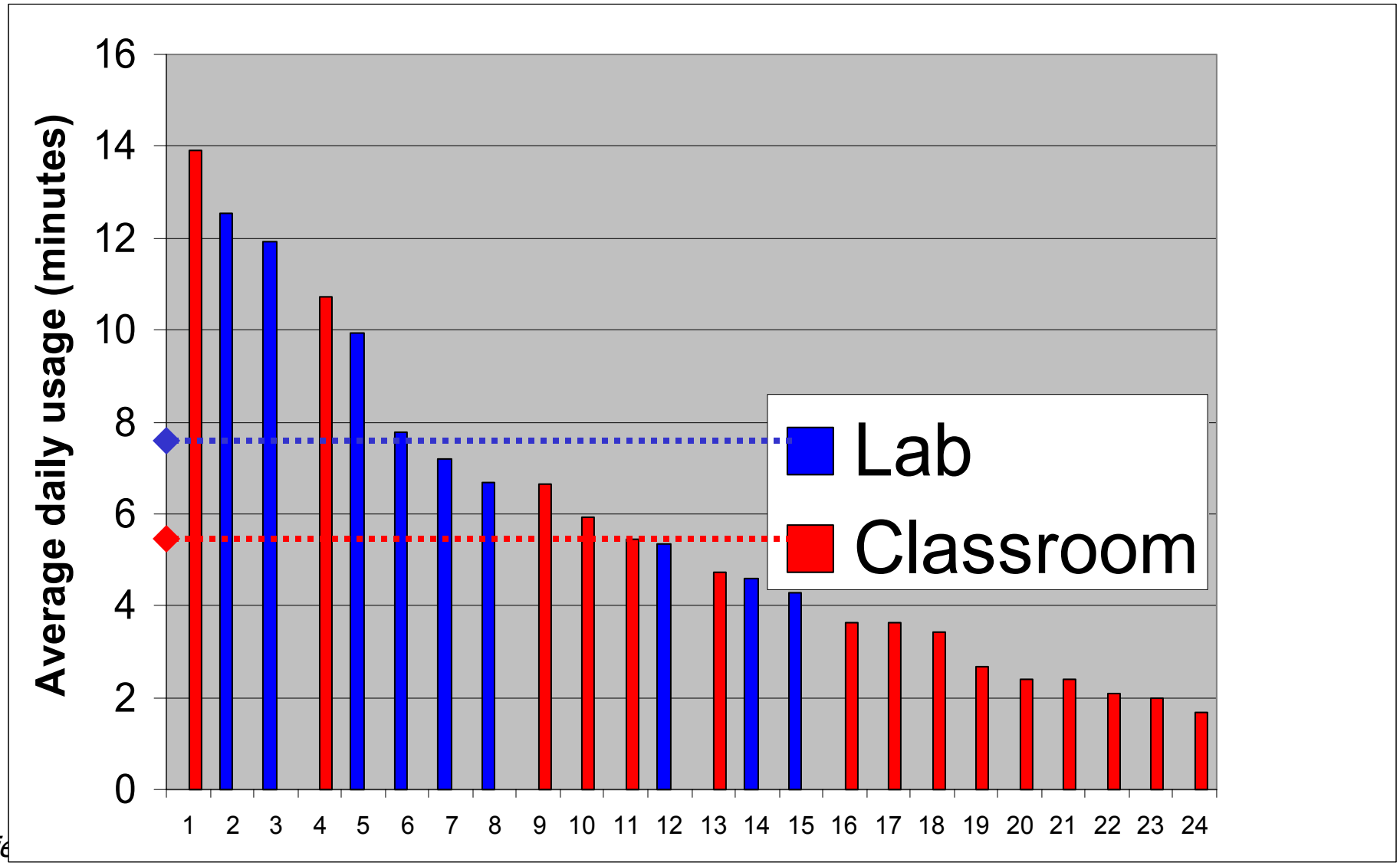
- How long a student's average session lasts (minutes)

Setting:	Lab	Class	Specialist	Resource
Frequency	40.2% >	30.1% >	16.7%	10.0%
Duration	19.2 >	15.1	13.5	12.7

(> indicates statistically significant difference;
results adjusted to control for differences in grade and ability)

Carnegie Mellon

2002-2003 usage: lab > classroom -- but top classrooms > average lab



Summary of 2002-2003 usage analysis

Setting had huge effect

- Labs averaged higher than all but top teachers
- Specialists liked the Tutor but saw kids rarely

Teacher had strongest influence on usage

- Accounted for almost all variance in frequency
- Accounted for over half of variance in duration

#students/computer affected classroom usage

- Correlated -0.4 with frequency and duration

Efficacy: gain per hour on Tutor

What influences efficacy?

- What the Reading Tutor does
- What the student does

Carnegie Mellon

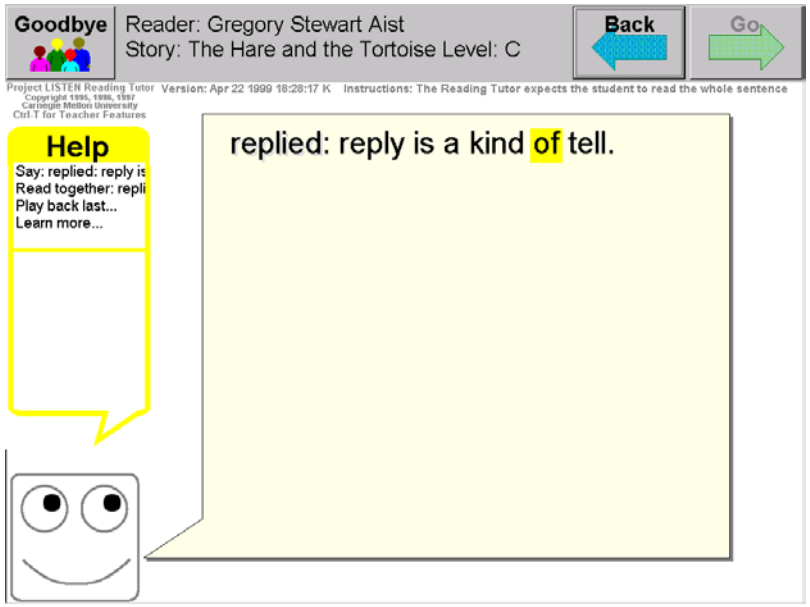
How to trace effects of tutoring? Find signature of tutoring on student.

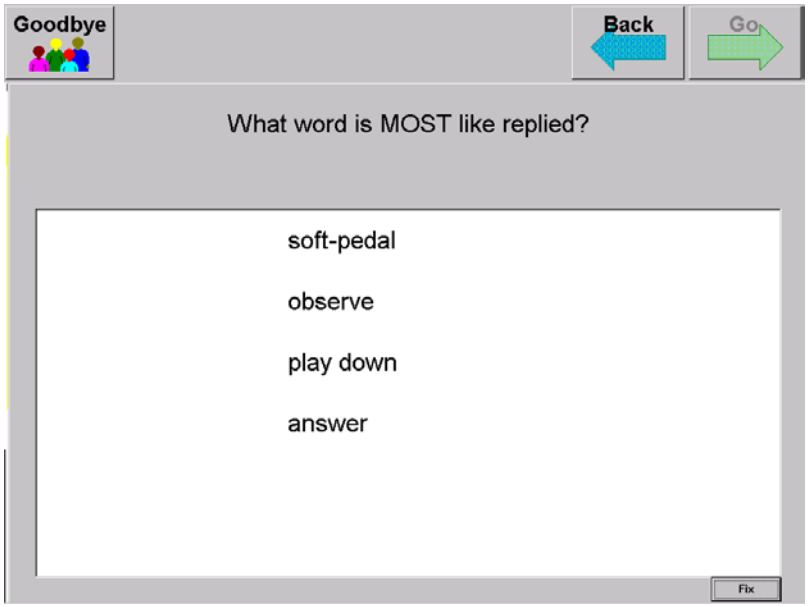
Carnegie Mellon

Experiment to trace effects of tutoring

Does explaining new vocabulary help more than just reading in context?

Randomly pick <u>some</u> new words to explain; later, test <u>each</u> new word.





Did kids do better on explained vs. unexplained words?

- Overall: no; 38% ≈ 36%, N = 3,171 trials [Aist 2001 PhD].
- Rare 1-sense words tested 1-2 days later: yes! 44%>>26%, N=189.

Carnegie Mellon

How to trace effects of student behavior? Relate time allocation to gains.

Relating time allocation to gains

Compute time allocation among actions

- Logging in, picking stories, reading, writing, waiting, …

Partial-correlate pre-to-post-test gains against % of time

- Control for pretest score differences among students

Fluency gains in 2000-2001 study:

- +0.42 partial correlation with % time spent reading
- –0.45 partial correlation with % time picking stories

Mostow, J., Aist, G., Beck, J., Chalasani, R., Cuneo, A., Jia, P., & Kadaru, K. (2002, June 5-7). A La Recherche du Temps Perdu, or As Time Goes By: Where does the time go in a Reading Tutor that listens? *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems (ITS'2002)*, Biarritz, France, 320-329.

Carnegie Mellon

Conclusion: Effectiveness = Usage × Efficacy

- Technology impact depends on context.
- The dependencies must be studied.
- Instrumentation can help.

Carnegie Mellon

The end…

Carnegie Mellon

More? See www.cs.cmu.edu/~listen

Carnegie Mellon

Questions?

Carnegie Mellon

What does classroom technology need?

- Electric power
- Student acceptance
- Teacher acceptance
- Tech support
- Student ratio
- Responsibility
- Administration support
- Community acceptance
- Critical mass
- Funding
- Affordability

Problems:

intrinsic vs. temporary

Aphorisms

What's in the box?

- SW? HW? Support? Assessment? …

A feature is something you can turn off.

A switch is something you can set wrong.

- Hide, don't delete.

The range of technical expertise among the students is greater than among the teacher.

Anything that can go, can go wrong.

- We are all idiots.