Digital Tutoring and Accelerating Expertise in Information Technology: Crossing the 2-Sigma Threshold and Beyond

National Initiative for Cybersecurity Education
17 September 2013

J. D. Fletcher
Institute for Defense Analyses

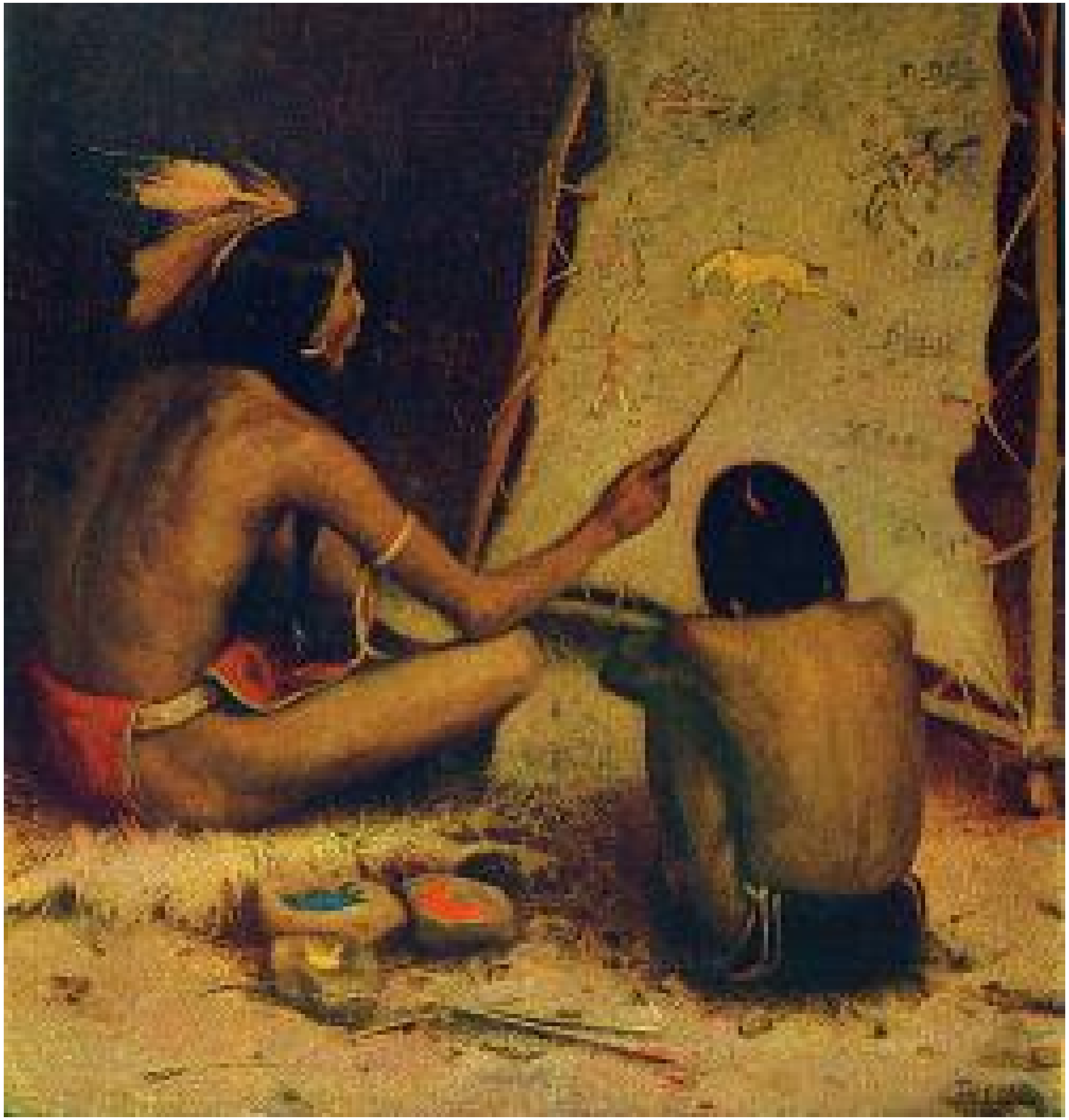
fletcher@ida.org

Topics

- Why tutoring?
  - Why computers?
- Can computers teach?
- About digital tutoring
- Do People Learn from Digital Tutors?
- The DARPA Digital Tutor
- Into the Future: ADL?
- Finale

Why (One-on-One) Tutoring?

The Last 50,000 years (or so) of Human Training and Education



Individualization:
From Yue-zheng (4th C. BC), to Quintilian (1st C. AD), to …

"The principal consequence of individual differences is that every general law of teaching has to be applied with consideration of the particular person ... responses to any stimulus … will vary with individual capacities, interests, and previous experience."

E. L. Thorndike (1906)

Empirical Basis for Individualization

"Whilst part of what we perceive comes through our senses from the object before us, another part (and it may be the larger part) always comes out of our mind"

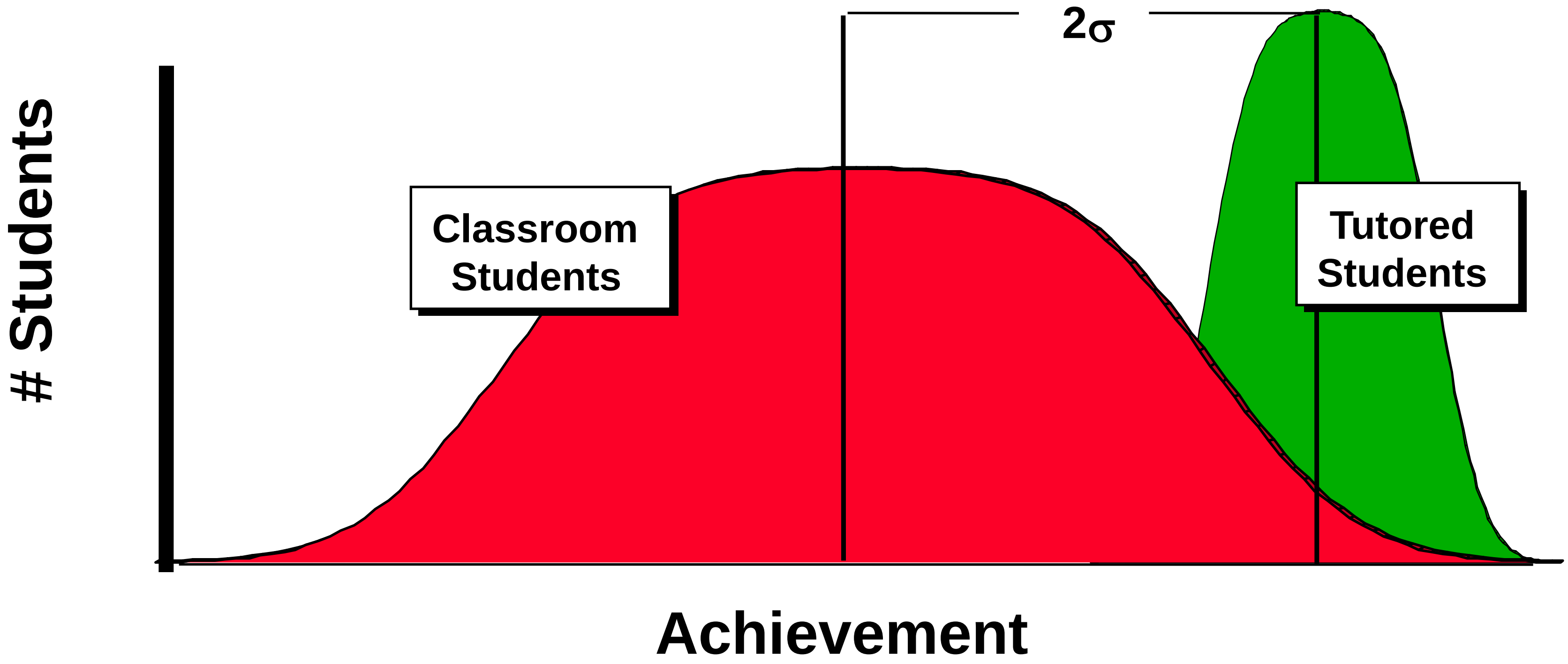
William James (1890)

Empirical Basis for Individualization

"The central assertion is that seeing, hearing, and remembering are all acts of *construction,* which may make more or less use of stimulus information depending on circumstances."

Ulric Neisser (1967)

Classroom vrs One-on-One Tutoring



(Bloom, 1984)

We can’t afford a human tutor for every learner, but we may be able to afford a computer, or a cell phone, or ...

Why Is Tutoring So Effective?

- Individualization
- Interactivity/Immersion

Individualization & Classroom Instruction: Pace

- Ratio of time needed to build words from letters in kindergarten -- 13:1 (Suppes, 1964)
- Ratios of time needed to learn in grade 5 -- 3:1 and 5:1 (Gettinger & White, 1980)
- Ratios of time needed by hearing impaired and Native American students to reach mathematics objectives -- 4:1 (Suppes, Fletcher, & Zanotti, 1975, 1976)

Interactivity/Immersion/"Flow"
Classroom Instruction & Tutoring

Number of Questions Asked Per Hour

	Traditional Classroom/Hr	Tutored Session/Hr
Student	0.1	20-30
Instructor	3	120-150

(Graesser & Person, 1994)

Why Computers?

“Individualization is an educational imperative and an economic impossibility.” (Michael Scriven, 1975)

Enter the Computer: A Third Revolution in Learning?

- Revolution #1: Writing
  Content of learning made available anytime, anywhere

- Revolution #2: Books
  Affordable content of learning made available anytime,anywhere

- Revolution #3(?): Technology
  Affordable content and the interactions of learning made available anytime, anywhere

- **On-demand learning is the common thread.**
- **We are returning to learning dialogues/conversations.**

Primordial Beginnings (1960s)

- **Illinois – PLATO**
  let every Asimov arise (and use Plasma panels)

- **MITRE/Texas/BYU – TICCIT**
  from ISD to computer system design

- **Stanford – Curriculum**
  Drill and Practice and information structures (!)

- **BBN & MIT - ICAI**
  Feuerzeig -- mixed-initiative dialogue and information structures

Also: US Air Force, IBM Labs, Columbia, Penn State

Individualizing with Paper: Keller's PSI

Instruction via modules

Pre-Test → Diagnose & Assess → Study Guide → Post-test

Also:
Postlethwait's Audio-Tutorial Approach
Klausmeier's Individually Guided Education
Bloom's Learning for Mastery
And others …

Individualizing with Paper: Crowder's Intrinsic Programming

In the multiplication 3 X 4 = 12,
the number 12 is called a ______.

A. Factor [Branch to remedial X1]
B. Quotient [Branch to remedial X2]
C. Product [Reinforce, go to next]
D. Power [Branch to remedial X3]

Individualizing with computers: PSI (Keller) & Intrinsic Programming (Crowder)

From previous module

Pre-test

Pass?

Yes, on to the next module

No, go to this module

In the multiplication 3 X 4 = 12, the number 12 is called a _______.

A. Factor [Branch to remedial X1]
B. Quotient [Branch to remedial X2]
C. Product [Reinforce, go to next]
D. Power [Branch to remedial X3]

(An Example Intrinsic Programming Item)

Post-test

Other Approaches: E.g., Suppes' Strands Approach

An excruciatingly detailed analysis of the subject matter, broken up into areas of activities (strands) followed by …

- **Instruction based on student history and level of achievement**
- **Acceleration where appropriate**
- **Repeated practice where necessary**
- **A daily profile reporting each student's progress to the student and classroom teacher**

Mathematics K-7 Strands

Strand	Content	Grade level
NUM	Number concepts	1.0-7.9
HAD	Horizontal addition	1.0-3.9
HSU	Horizontal subtraction	1.0-3.4
VAD	Vertical addition	l.0-5.9
VSU	Vertical subtraction	1.5-5.9
EQN	Equations	1.5-7.9
MEA	Measurement	l.5-7.9
HMU	Horizontal multiplication	2.5-5.4
LAW	Laws of arithmetic	3.0-7.9
VMU	Vertical multiplication	3.5-7.9
DIV	Division	3.5-7.9
FRA	Fractions	3.5-7.9
DEC	Decimals	4.0-7.9
NEG	Negative numbers	6.0-7.9

Allocating Time and Effort to the Strands

Strand		Half year													
		1.0	1.5	2.0	2.5	3.0	3.5	4.0	4:5	5.0	5.5	6.0	6.5	7.5	7.5
NUM	PT	50	24	24	17	10	5	7	7	8	11	14	10	15	15
	PP	36	18	16	12	10	4	8	8	10	14	20	10	19	19
HAD	PT	26	21	21	9	14	9								
	PP	32	28	26	10	14	8								
HSU	PT	14	10	16	9	4									
	PP	18	14	16	10	4									
VAD	PT	10	10.	9	19	19	7	8	2	3	1				
	PP	14	12	12	22	20	6	10	2	4	2				
VSU	PT		9	8	15	22	10	13	3	3	1				
	PP		12	12	18	20	8	10	2	4	2				
EQN	PT		17	12	16	17	14	17	7	5	7	7	8	15	15
	PP		10	10	12	16	12	20	8	8	12	12	10	19	19

Note.--PT = proportion of time; P P = proportion of problems.

Allocating Time Among Students: Fun with Regression Equations

Linear:

$$E(O_i) = b_0 + b_1 P_i + b_2 T_i$$

Linear with interaction:

$$E(O_i) = b_0 + b_1 P_i + b_2 T_i + b_3 P_i T_i$$

Cobb-Douglas:

$$E(\ln O_i) = b_0 + b_1 \ln P_i + b_2 \ln T_i$$

Exponential:

$$E(\ln O_i) = b_0 + b_1 \ln P_i + b_2 \ln T_i + b_3 (\ln T_i)^2 + b_4 (\ln T_i)^3$$

(O = outcome; P = pretreatment measure; T = time)

Can Computers Teach?

An Aside -- Effect Size

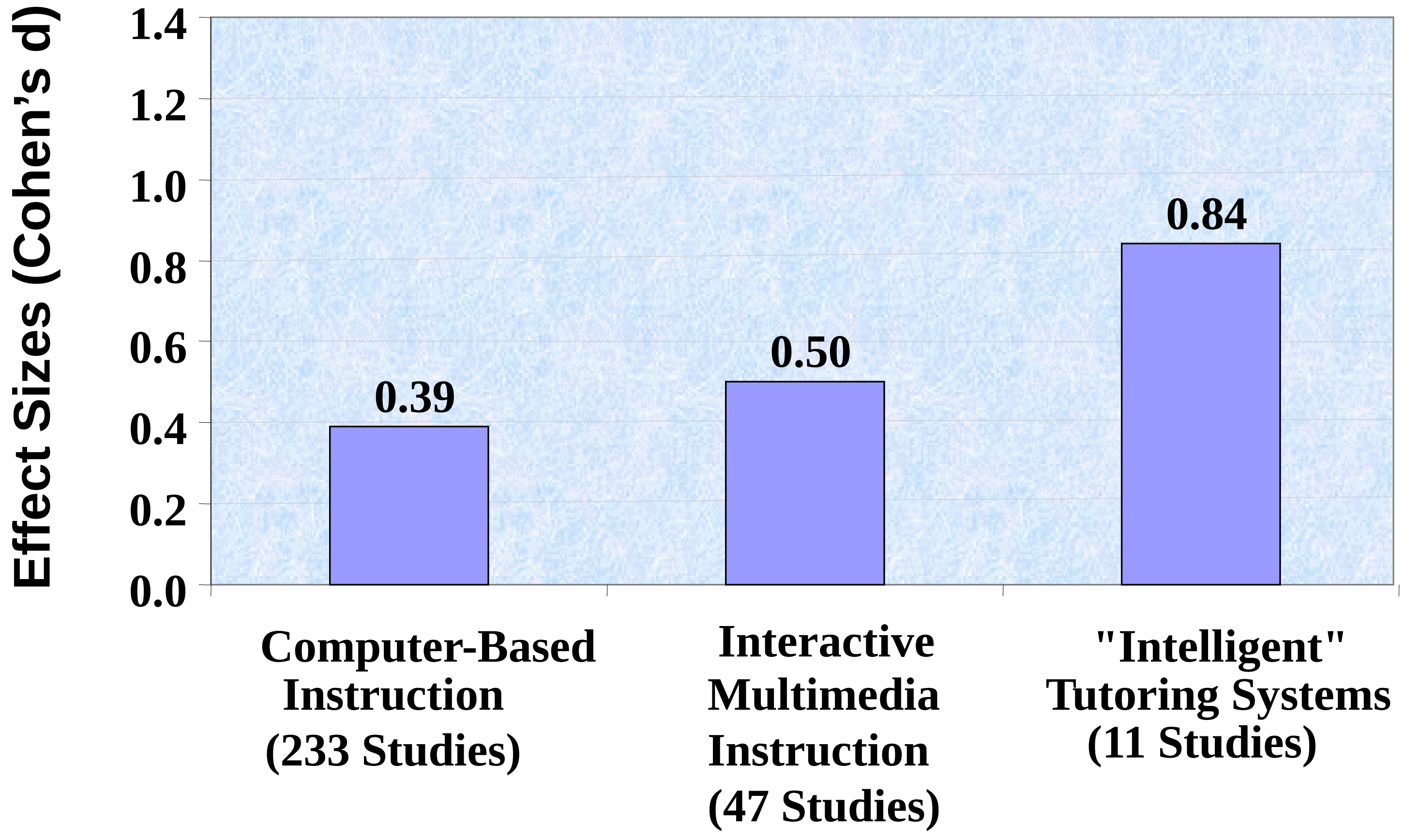
A descriptive (not inferential) statistic commonly used to estimate the magnitude of an effect (e.g., experimental treatment).

Cohen's d =

$$\frac{\text{Mean Group 1} - \text{Mean of Group 2}}{\text{“Pooled” Standard Deviation}}$$

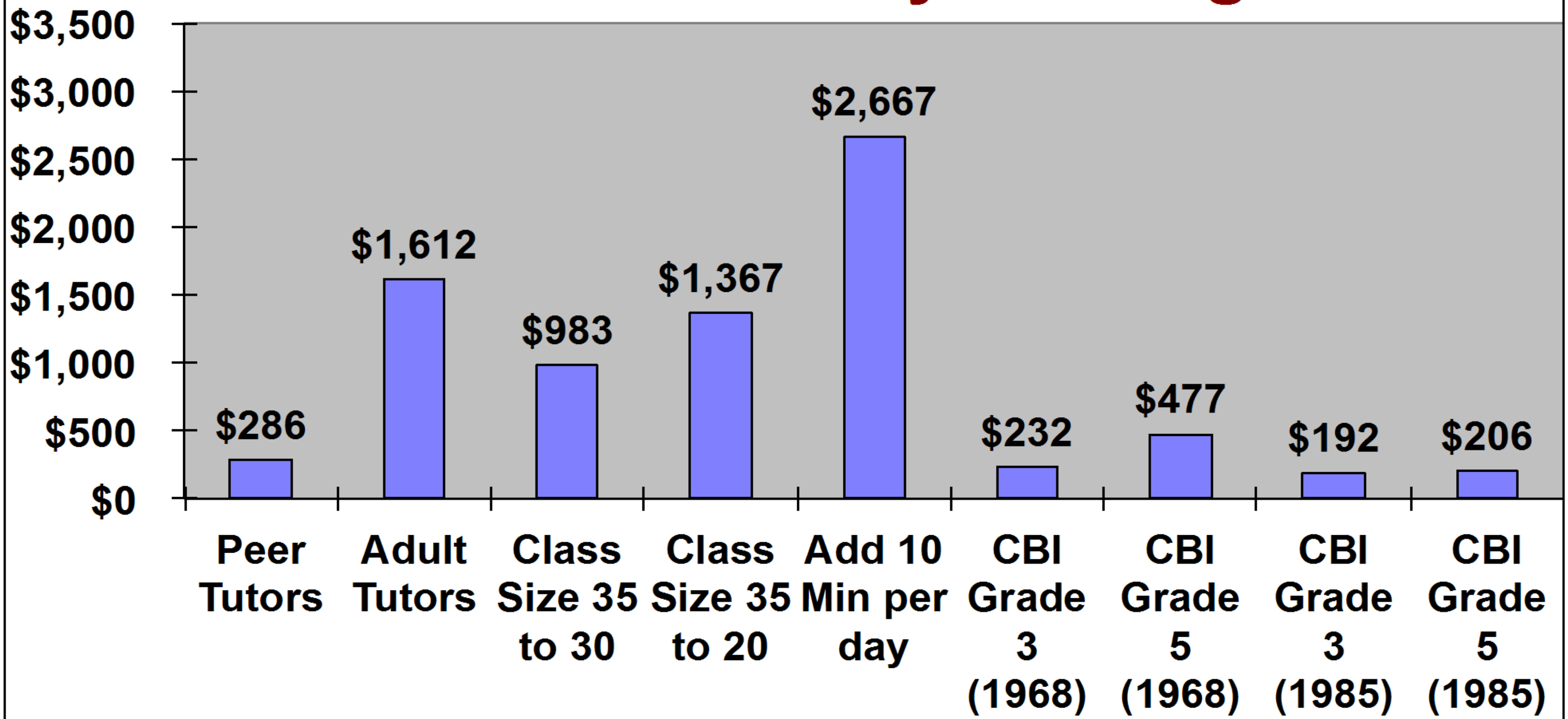
d < 0.25	Negligible
> 0.25 to 0.40	Small
> 0.40 to 0.60	Moderate
> 0.60 to 0.80	Large
d > 0.80	Very Large

Learning Gains with Early (pre-1992) Computer-Assisted Instruction



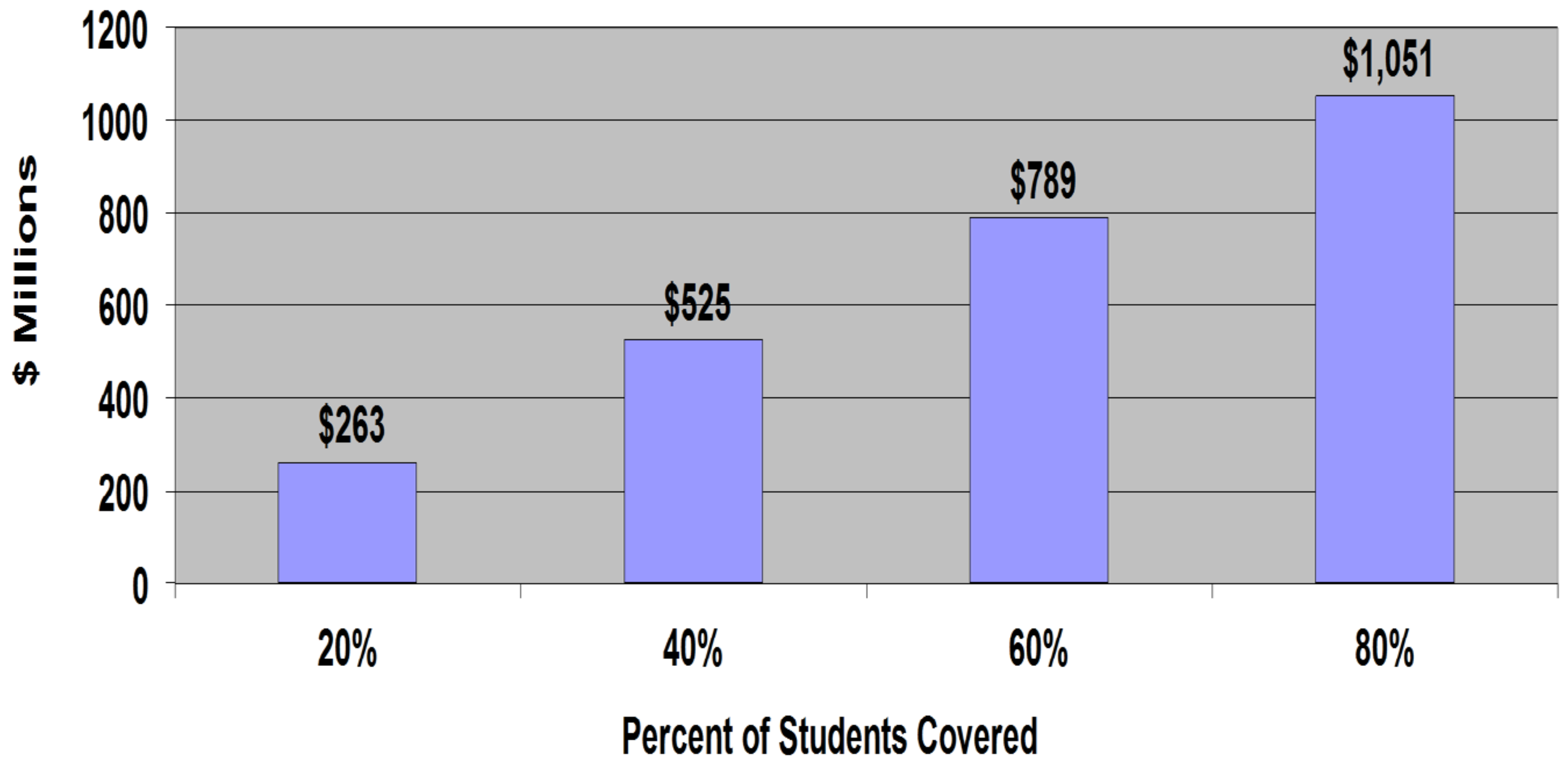
(Fletcher, 1997)

Costs to Increase Mathematics Achievement by One Sigma



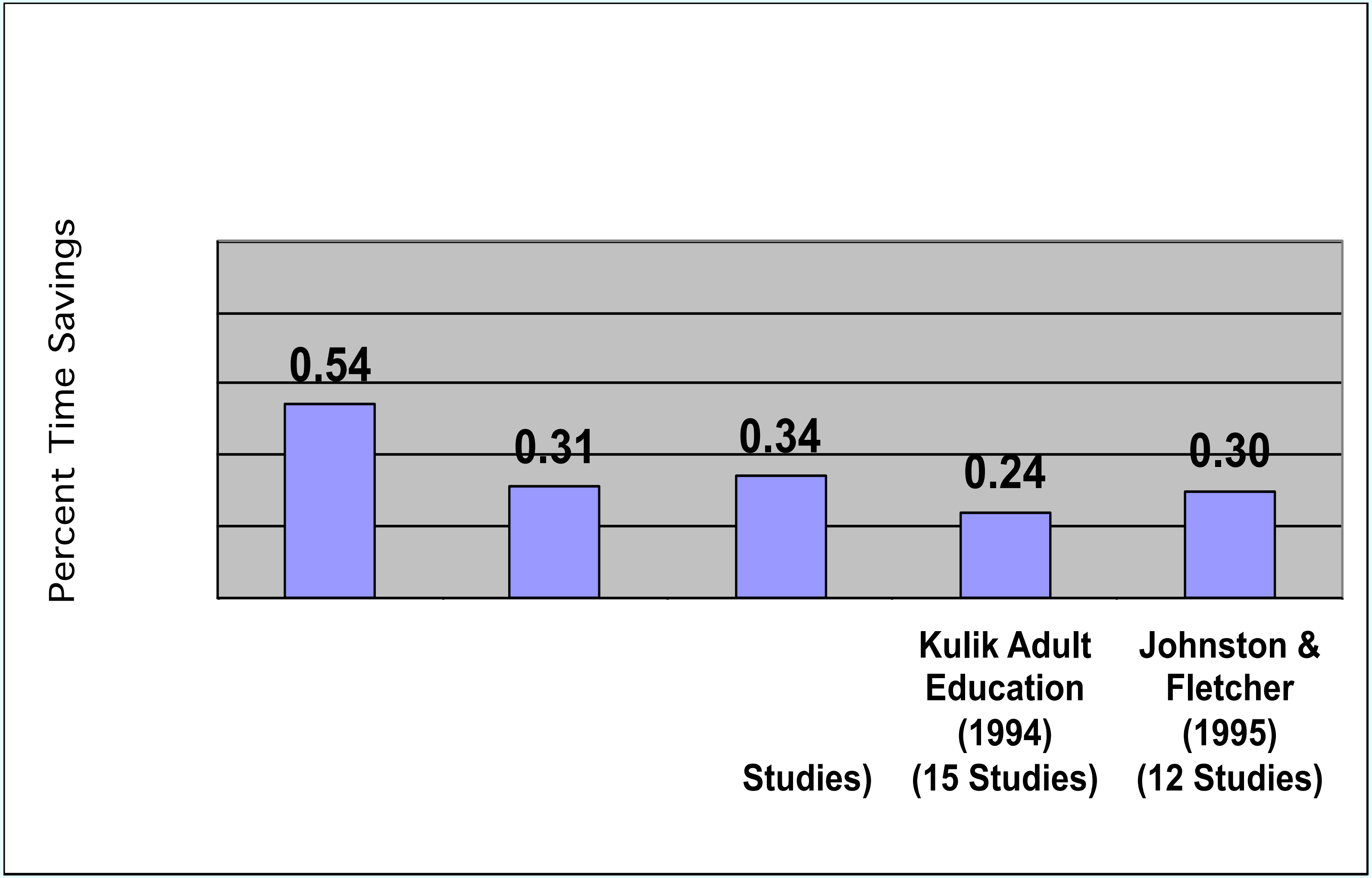
(Fletcher, Hawley, & Piele, 1990)

Savings in DoD Skill Training With 30 Percent Reduced Training Time



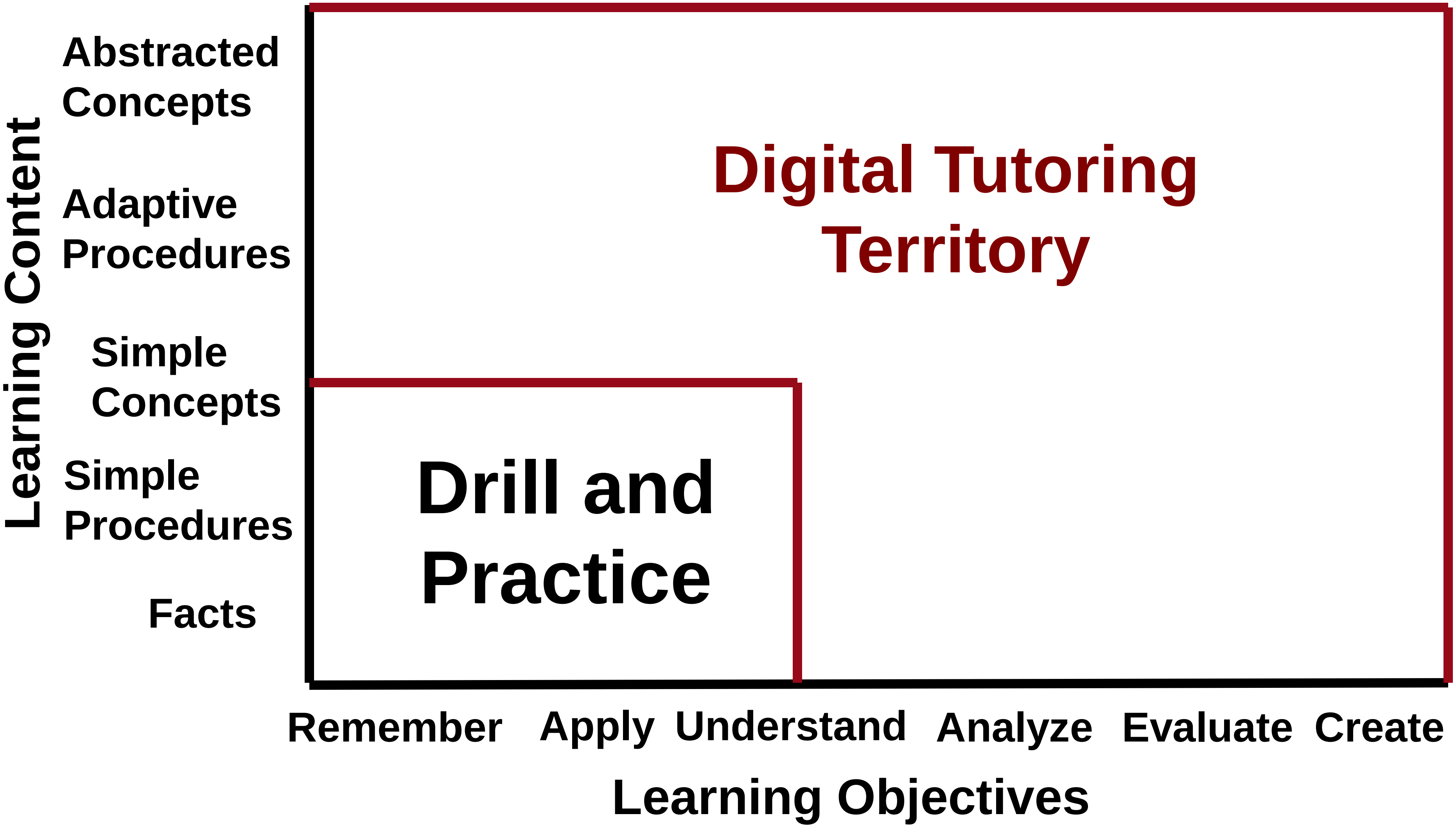
Model from Angier & Fletcher, 1992

Are 30 Percent Time Savings Achievable?



About Digital Tutoring

A Role for Digital Tutoring



(Framework courtesy of Anderson & Krathwohl, 2001)

Primordial ITS (nee ICAI) motivation

Two findings from early (1960s- early 1970s) research:

- **Computers can teach effectively.**
- **Anticipating and then dealing with all possible states of the learner is time-consuming and difficult (i.e., costly) -- and incomplete).**

A thought:

- **Perhaps computers can do some of this_ ‘authoring’ themselves, i.e., they might generate instruction:**
  - **In real time**
  - **On demand**
  - **As tutorial dialogue**

What Do We Mean by Digital Tutoring?

Two main differences (Carbonell, 1970):

- **Based on information structures rather than ad-hoc frames**
- **Mixed-initiative dialogue**

Requires:

- **Explicit models of the subject matter**
- **Dynamically evolving model of the learner**
- **Tutorial strategies, based on learning objectives and evolving state of the learner**

Can We Generate Dialogues? Mentor (1)

Mentor: Begin the physical examination.
Student: PHYSICAL EXAM

M: Fine, what in particular would you like to know?
S: RESPIRTORY RATE

M: It is common practice to begin an examination by checking the patient's general appearance.
S: GENERAL APPEARANCE

M: Acutely prostrated, lying in bed on his right side ….

-
-

S: LABORATORY REPORTS
M: Finish the physical exam first.

-
-

Tutoring in Mentor (2)

S: CONTACT WITH BIRDS
M: None, except the patient has a pet canary

S: PSITTACOSSIS
M: There's insufficient evidence for any conclusive diagnosis at this time. Show a little more restraint please.

S: PULSE?
M: 100 a minute. Occasional extra systole.

S: TEMPERATURE?
M: 103

S: PNEUMOCOCCAL PNEUMONIA.
M: Stop guessing and continue with the examination.

•

•

(Feurzeig, 1969)

A Dialogue from Sophie

Student: What is the base emitter voltage of the voltage limiting transistor?

SOPHIE: The base emitter voltage of Q5 is -.037 volts.

Student: What about for the current limiting transistor?

SOPHIE: The base emitter voltage of Q8 is .349 volts.

Student: What should it be?

SOPHIE: In a working circuit, the base emitter voltage of Q8 is .593 volts.

Student: Replace Q8.

SOPHIE: I am going to ask you some questions about how Q8 is faulted. Are any junctions shorted?

•

•

(Brown, Burton, & DeKleer, 1982)

Early ICAI/ITS Systems

MENTOR	Medical Diagnosis	Feurzeig, 1967
SCHOLAR	South America geography	Carbonell, 1970
WHY	Causes of rainfall	Stevens, et. al. 1982
SOPHIE	Electronic troubleshooting	Brown, et al., 1982
WEST	Arithmetic expressions	Burton & Brown, 1979
BUGGY	Subtraction	Brown & Burton, 1978
WUSOR	Logical relations	Goldstein, 1982
EXCHECK	Logic and set theory	Suppes, 1982
BIP	BASIC programming	Barr, et al., 1976
SPADE	LOGO programming	Miller, 1982
ALGEBRA	Algebra word problems	Lantz, et al., 1983
LMS	Algebraic procedures	Sleeman, 1982
QUADRATIC	Quadratic equations	O'Shea, 1982
GUIDON	Infectious diseases	Clancey, 1982
MENO	PASCAL programming	Soloway, et al., 1983
STEAMER	Steam propulsion (USN)	Williams, et al., 1981

Do People Learn from Digital Tutors?

Meta-Analyses

VanLehn (2011):

- **27 Evaluations**
  - **Effect size of 0.59 overall**
  - **Effect size of 0.76 for step-based tutoring**
  - **Effect size of 0.40 for substep-based tutoring**

Kulik/Fletcher (2012):

- **45 “Systems Evaluations”**
  - **Effect size of 0.63 overall**
  - **Effect size of 0.86 for 39 properly aligned studies**

Where Is Digital Tutoring Best Applied?

Reference	Effect Sizes for Deep Learning	Effect Sizes for Shallow Learning
Person, Bautista, Graesser, & Mathews (2001)	0.34	0.00
Graesser, Moreno, Marineau, Adcock, Olney, & Person (2003)	0.30	0.03
VanLehn, Lynch, Schulze, Shapiro, Shelby, Taylor, & Wintersgill (2005)	0.95	–0.08
Overall	0.62	–0.02

The DARPA Digital Tutor (DT)

The DARPA Challenge

16 weeks of tutoring to produce graduates who are superior in knowledge and practical skills to technicians with many years of experience.

Context for the Digital Tutor (DT)

- A product of DARPA's Training Superiority and Education Dominance programs (DARWARS, Ambush!, Tactical Language and Cultural Training, …)
- Focused on accelerating expertise
- Provides 16 weeks covering "A" school and some "C" school training for USN Information Systems Technology (IT) rating
- Approach is to capture procedures and practices of <u>expert</u> one-on-one tutors
  - Spiral curriculum focused on concepts
  - Hands-on work with IT systems

Basic Approach for the Digital Tutor

- **Borrows ideas from intelligent tutoring technology and constructivist notions, but aspires to be rigidly neither:**
  - **Its strategy is eclectic and pragmatic**
  - **Its validation is job performance**

- **Its approach is to:**
  - **Capture procedures and practices of subject matter experts who are <u>also</u> expert one-on-one tutors**
  - **Emphasize active (situated, authentic) problem solving to develop higher order concepts**

The DARPA IT Tutor

Why Information Technology?

- An operationally critical competency
- Current training in sore need of improvement (agreement across all echelons)
- An Incredibly Complex Task

Design Features: Strategies

- **Thorough front end analysis to determine objectives for expertise**
- **Modeled on human tutors who were expert in subject elements <u>and</u> 1-1 tutoring**
- **Spiral curriculum with focus on problem solving**
- **Not dependent on expert solutions**
- **Focus on conceptual understanding**
- **Use Drill and Practice to teach the basics**

Design Features: Tactics

- **No hints**
- **Never solve the problem for the learner**
- **Build on what the learner knows to resolve impasses**
- **Always question a successful solution**
- **Frequently question a successful step**

Five Assessments

- **July-August 2009 – IWAR 1 – 1 week DT + 15 weeks of human tutoring**
- **April 2010 – 4 weeks then available of the DT**
- **November 2010 – 8 weeks then available of the DT**
- **March-April 2012 – IWAR 2 – 16 weeks – First version of a fully completed DT**

NB:

- **Assessments focused on job performance**
- **The usual tests of statistical significance**
- **Effect size measured by Cohen's d**

What are we looking for?
Kirkpatrick's Four Levels of Evaluation

Level	Description	Evaluation
(Did we do things right?)		
1	Surveys	Impressions and opinions
2	Outcomes	Were the objectives achieved?
(Did we do the right things?)		
3	Transfer	Is the unit more effective?
4	Benefits	Is the enterprise more effective?

Phase 1 IWAR
(IDA 2010 Document D-4047)

Comparison:

- 1 week of DT + 15 weeks of 1:1 (<u>human</u>) tutoring (N = 12)

  & Fleet ITs with 4-18 years Navy IT experience (N = 12)

Measures:

- 139-item Written Knowledge Test
- Practical Troubleshooting Exercises (2.5 days)
- System Building Exercise (6 hours)
- (Also Dockside and Deployed observations)

Phase 1 IWAR

Results:

- **Written Knowledge Test**
  DT > Fleet ITs (d = 1.02)[a]
- **Written Knowledge Test 1-Week DT portion**
  DT > Fleet ITs (d = 1.73)[a]
- **Practical Troubleshooting Exercises**
  DT > Fleet ITs (99 vrs 79 solved)[a]
- **System Building Exercises:**
  Fleet > DT (113 vrs 84 objectives met)

[a] $(p < 0.01)$

April 2010
(IDA 2011 Document NS D4260)

Comparison:

- 4 weeks of then available DT (N = 20)
  & 16 week Integrated Learning Environment (ILE)
  CBT graduates (N = 31)
  & School Instructors (N = 10)

Measure:

152-item written knowledge test covering IT material

Results:

DT ~ ILE	(d = 2.81)[a]
DT ~ Instructors	(d = 1.26)[a]
Instructors ~ ILE	(d = 1.25)[a]

[a]($p < 0.01$)

November 2010
(IDA 2011 Document NS D4260)

Comparison:

- 8 weeks of the available DT (N = 20)
  & 16 week Integrated Learning Environment (ILE)
  CBT graduates (N = 18)
  & 19 week IT of the Future (IToF) graduates (N = 20)
  & School Instructors (N = 10)

Measures:

- 293-item written knowledge test covering DT material
- 4 hours practical trouble shooting exercises
- 2 hours packet tracing exercises
- Oral exams (about 30 minutes) of 7 DT and 6 IToF students

November 2010 Results

- **Written Knowledge test**
  - **DT & ILE** (d = 4.68)[a]
  - **DT & IToF** (d = 1.95)[a]
  - **DT & Instructors** (d = 1.35)[a]
- **Practical exercises**
  - **DT & IToF** (d = 1.90)[a]
- **Packet Tracing Exercises**
  - **DT & IToF** (d = 0.74)[a] (Un-Weighted)
  - **DT & IToF** (d = 1.00)[a] (Weighted)

[a]($p < 0.01$)

IWAR 2 -- March-April 2012 – Full DT (IDA 2012 Document D-4686)

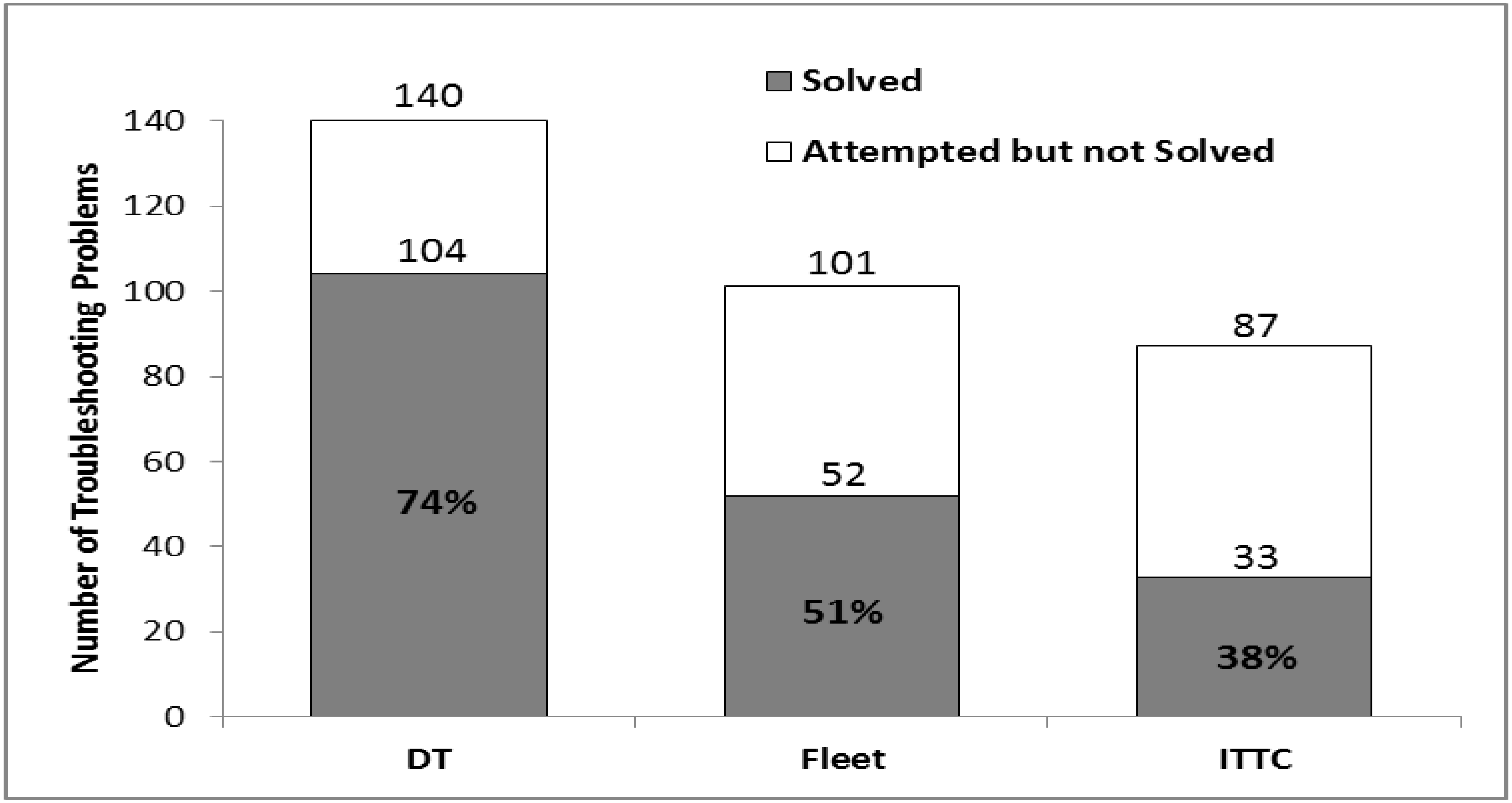
Comparison Groups:

- 16 weeks of the completed DT (N = 12)
- 35 weeks of IT Training Continuum (ITTC) (N = 12)
- Fleet ITs (N = 12) 9.6 Years average IT Experience

Measures:

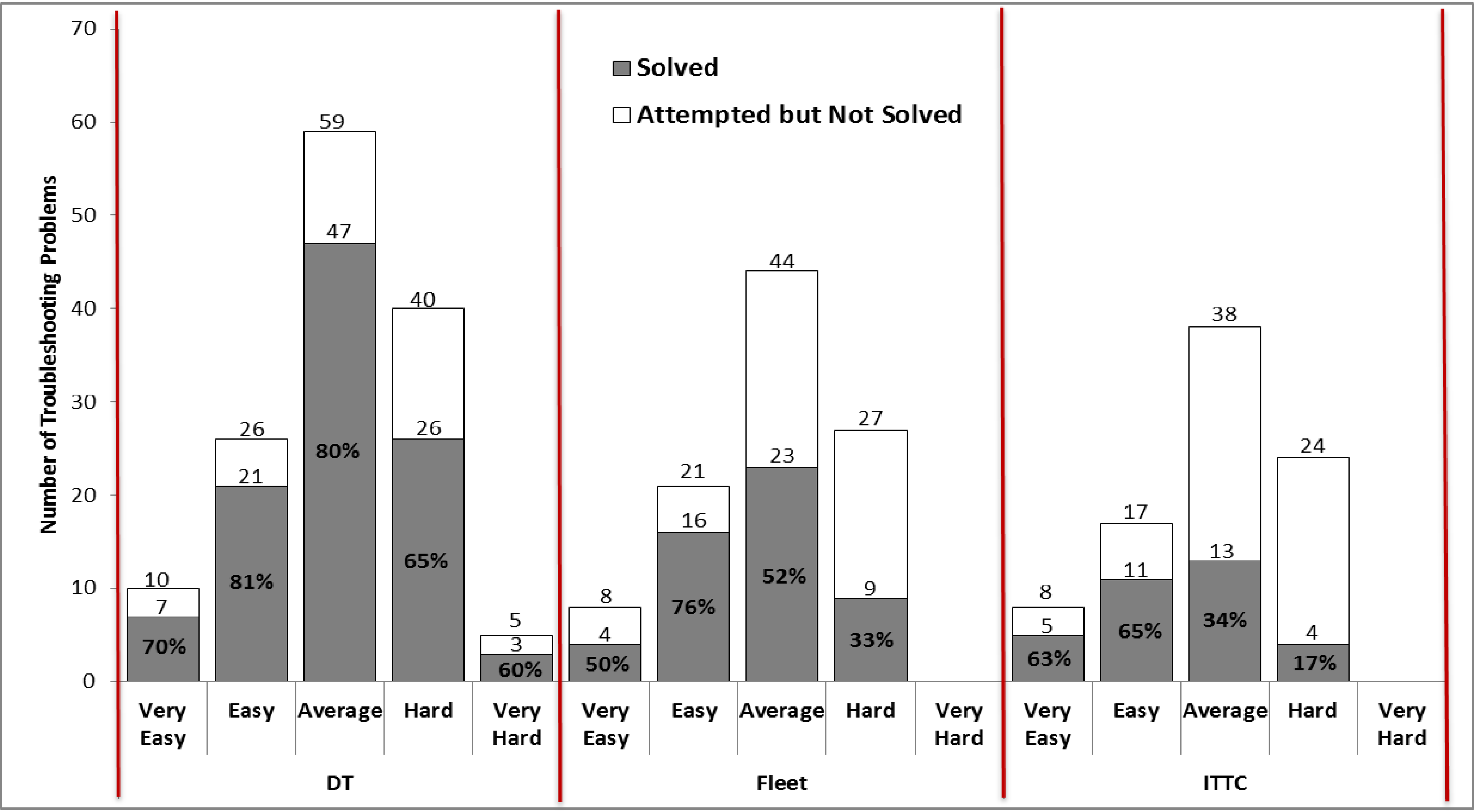
- 6 hours of problem solving (troubleshooting) exercises
- 272-item written knowledge test
- 3 hours of security exercises
- 6 hours of a system design and develop exercise
- 20-30 minute individual interviews

IWAR 2 Troubleshooting Exercises



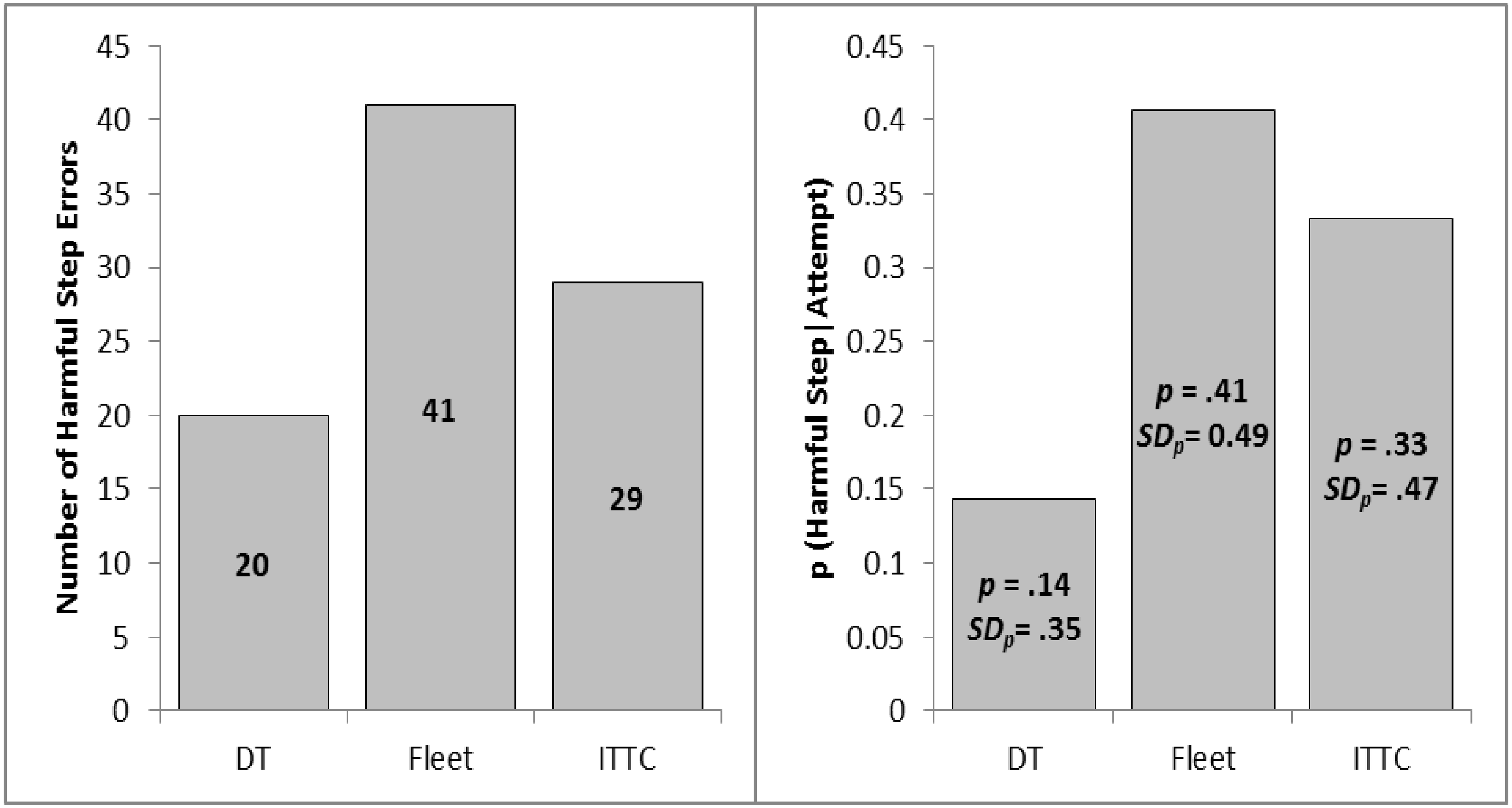
Solution Quality Scores		
	Occur by Chance	d
DT 3.78 (1.91) > Fleet 2.00 (2.26)	$p < 0.0001$	0.85
DT 3.78 (1.91) > ITTC 1.41 (2.09)	$p < 0.0001$	1.13

Troubleshooting Solutions and Difficulty



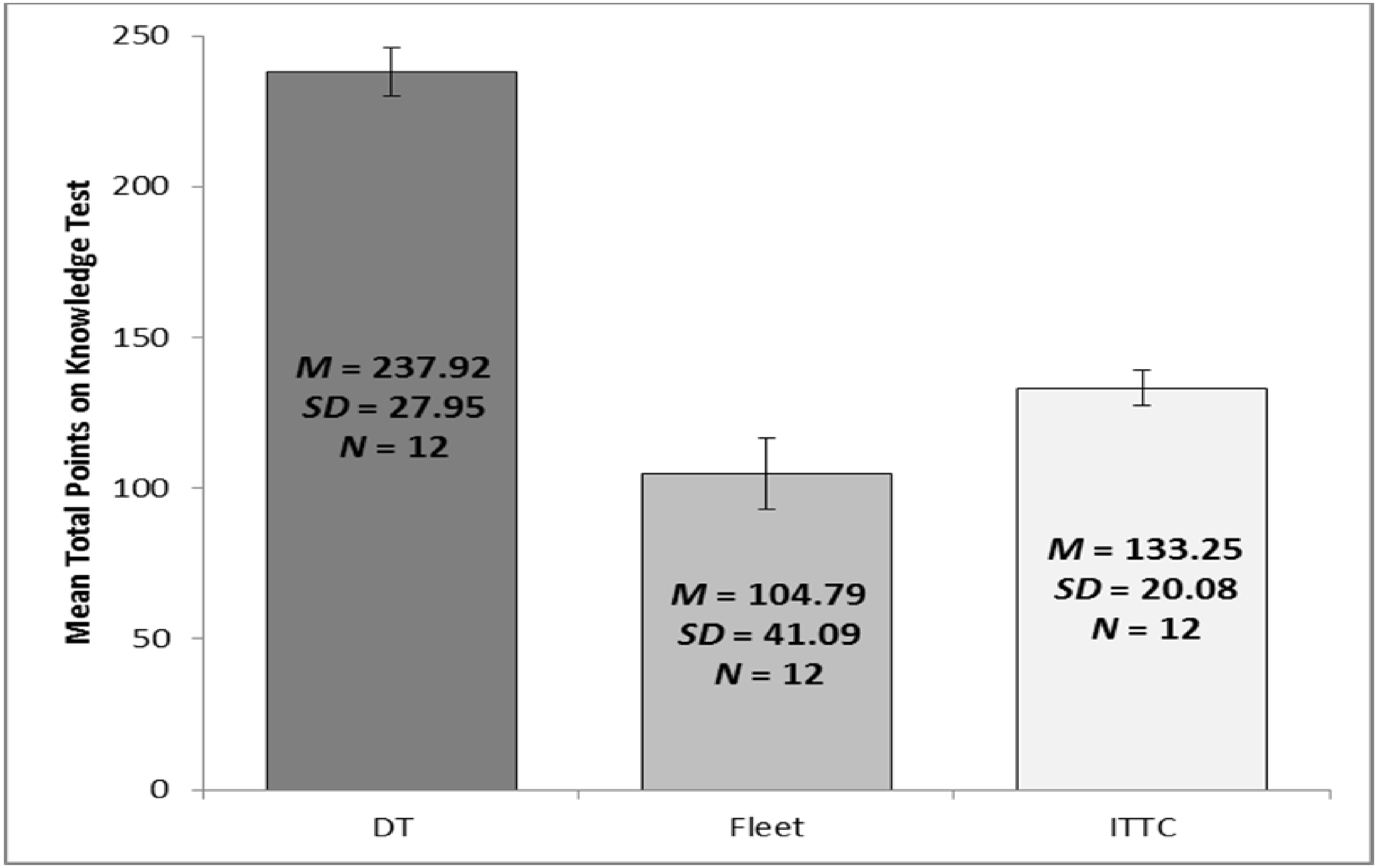
DT students attempted more problems at every difficulty level with greater probability of success.

Uncorrected Harmful Changes in Troubleshooting



Probability of Uncorrected Harmful Change		
	Occur by Chance	d
DT > Fleet	$p < 0.0001$	-0.61
DT > ITTC	$p < 0.01$	-0.44

Knowledge Test Scores



Knowledge Test Scores		
	Occur by Chance	d
DT > Fleet	$p < 0.0001$	4.30
DT > ITTC	$p < 0.0001$	3.38

IWAR 2 Results Summary DT & Fleet

Performance Measure	DT > Fleet?	Occur by Chance	Effect Size
Problem Solving (PS) Total	+	p < .0001	Large (0.85)
PS (Fewer) Harmful Actions	+	p < .0001	Medium (-0.61)
PS (Fewer) Unnecessary Actions	+	p < .0001	Medium (-0.54)
Individual Interviews	+	p < 0.01	Very Large (1.20)
Security Exercise	-	N.S.	Very Large (-1.30)
Network Design & Development	+	N.S.	Large (0.77)
Knowledge Test Total Score	+	p < 0.0001	Very Large (3.63)

IWAR 2 Results Summary DT & ITTC

Performance Measure	DT > ITTC?	Occur by Chance	Effect Size
Problem Solving (PS) Total	+	$p < .0001$	Very Large (1.13)
PS (Fewer) Harmful Actions	+	$p < 0.01$	Medium (-0.44)
PS (Fewer) Unnecessary Actions	+	$p < .0001$	Medium (-0.62)
Oral Review	+	$p < 0.05$	Large (0.90)
Security Exercise	-	N.S.	Negligible (-0.03)
Network Design & Development	+	$p < 0.01$	Very Large (1.41)
Knowledge Test Total Score	+	$p < 0.0001$	Very Large (2.63)

So What?

Cost of the Digital Tutor = $40-50M

IT Tutor Return on Investment

Option	Year											
	1	2	3	4	5	6	7	8	9	10	11	12
"A" School	A + 7 OJT	A + 7 OJT	A + 7 OJT	A + 7 OJT	A + 7 OJT	A + 7 OJT	A + 7 OJT	A + 7 OJT	A + 7 OJT	A + 7 OJT	A + 7 OJT	A + 7 OJT
"A" School & Digital Tutor	A + 7 OJT DT Dev	A + 7 OJT DT Dev	A + 7 OJT DT Dev	A + 7 OJT DT Dev	7 OJT DT Op	6 OJT DT Op	5 OJT DT Op	4 OJT DT Op	3 OJT DT Op	2 OJT DT Op	1 OJT DT Op	DT Op
Cost Differ-ence ($M)	($12.5)	($12.5)	($12.5)	($12.5)	$19.8	$72.2	$121.6	$171.0	$220.4	$269.8	$319.2	$368.7
12-Year ROI (3,076.2 – 1,563.6) = $1,512.6M												

Into the Future: ADL?

A thought:

The future is already here, but unrecognized and unevenly distributed.

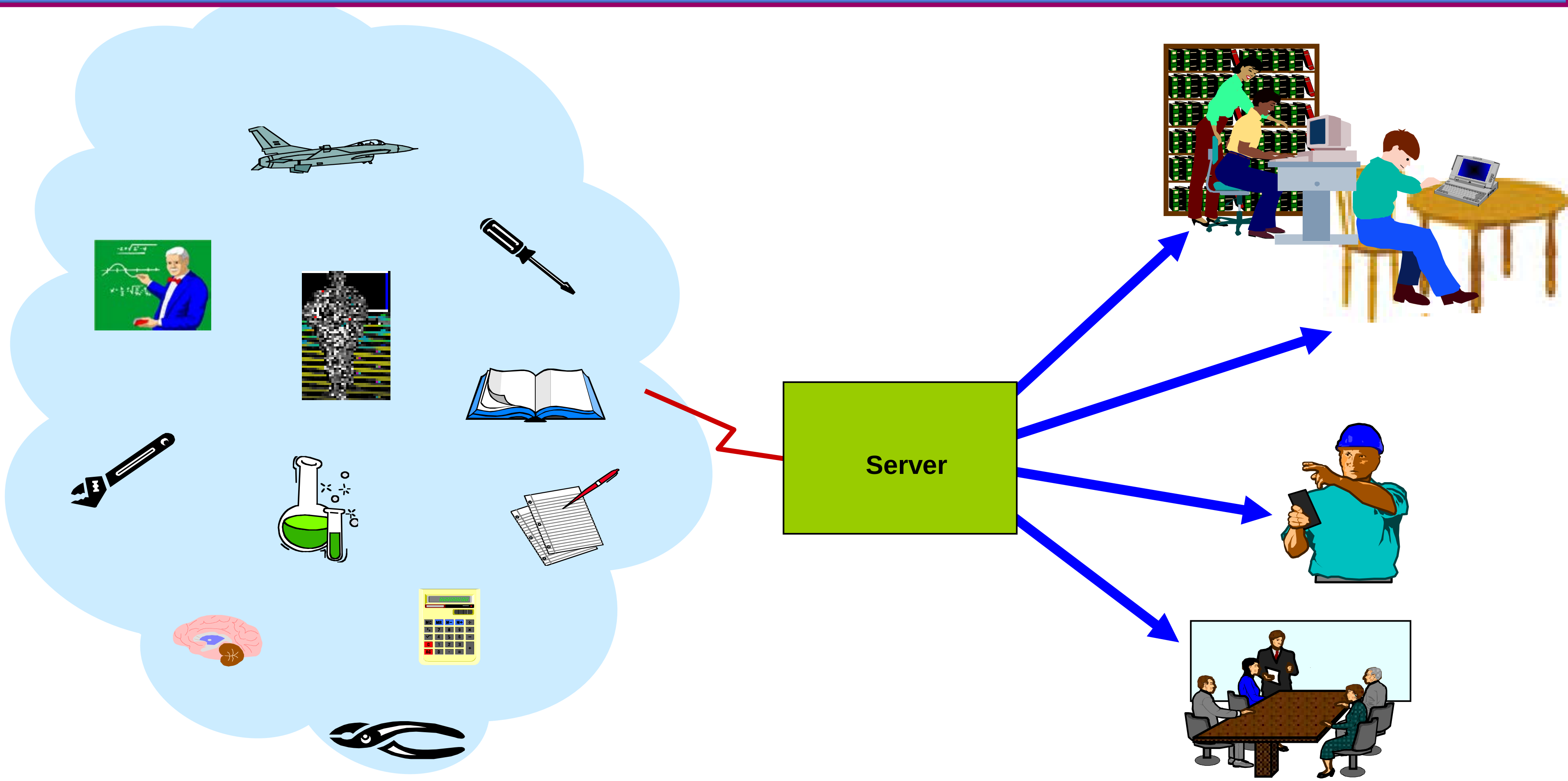
The Past as Prologue: The ADL Vision

Intelligent Tutoring Systems
Moore' s Law
Global Information Grid
Natural Language Interaction
Electronic Performance Aids
Distributed Learning Capabilities
Object Oriented Applications
Hand-Held Computers
Simulations and Games
Etc.

→

Personal Learning Associates

Hence, the ADL Vision



Shareable instructional objects from across the World Wide Web

Assembled in real-time, on-demand

To provide learning and assistance anytime, anywhere via guided dialogues

Eventually …

- **Anywhere, Anytime Learning Integrated with Performance/Decision Aiding**
  (Integrating the supply and demand side of learning)

- **Fewer Lessons, More Learning**
  (Learning as conversation)

- **Fewer Tests -- More Assessment (‘stealth assessment’)**
  (Continuous, Unobtrusive)

- **Personal Learning Associates (‘distributed’)**
  (In classrooms and out – anytime, anywhere)

Instruction (and Performance/Decision Aiding) as Individualized Tutorial Conversation

In Conclusion …

Implications

- **The fiscal folly of inadequate residential training.**
- **Training and education are investments not expenses.**
- **Increase amount (and depth) learned over reducing time to learn.**
- **We can greatly accelerate the acquisition of expertise – and should – the Digital Tutor is not the only example (e.g., Sherlock, IMAT).**
- **We can similarly accelerate acquisition of basic skills – (e.g., reading studies, math).**
- **Spend the money -- ROI is relatively insensitive to development costs <u>at scale.</u>**

A Thought …

The difficult, intransigent issues may be organizational, administrative, and structural. Not technological.

Finally …

- **Digital Tutoring can accelerate the development of targeted problem solving expertise**
- **Digital Tutors are expensive to build**
- **We should build them anyway, because of large:**
  - **Monetary Return on Investment**
  - **Operational Return on Investment**

Questions? Comments? Objections? Complaints? Thank you!