# Spoken Dialog Systems for Tutoring

Amy Marsh

Ling 575

# Tutoring

- Idealized view – one-on-one work with an adult subject matter expert
- Can also include peer tutoring, group tutoring, computerized tutoring systems, asynchronous environments
- Research typically finds high effect sizes (up to 2.0)

# Why a Computerized Tutoring System?

- Human experts are extremely expensive
- Many of the reasons we think humans are superior turn out not to be true (Van Lehn 2011)
  - Detailed diagnostic assessments – humans use mastery information but don't diagnose a student's mental state
  - Choosing appropriate tasks – humans tend to follow a script
  - More student initiative – not really true
  - Broader domain knowledge – doesn't produce learning gains
  - Better able to motivate students – doesn't produce learning gains
  - Provide better scaffolding
  - Give better feedback

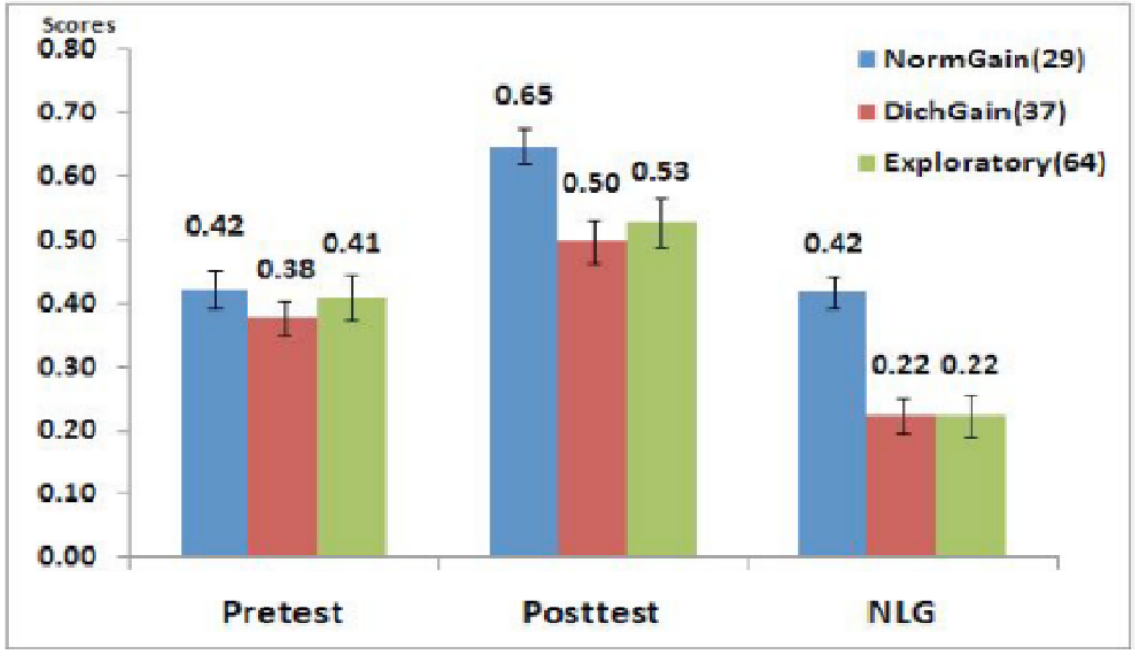
Kurt Van Lehn. (2011) The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems, *Educational Psychologist*, 46:4, 197-221.

# Can a computerized system provide scaffolding and feedback?

- Cordillera (Chi et al, 2010) - spoken dialog system for introductory physics
- Tutoring Decisions:
  - Elicit/Tell – should you tell the student the next step, or elicit it from the student?
  - Skip/Justify – should you justify the step just taken, or not?
- Can you use reinforcement learning to determine correct strategy?
  - Tutoring dialogs are very long – lots of states
- Reward: learning gain from pretest to posttest
- Separate strategies for different topics (i.e. kinetic energy, potential energy)

# Cordillera (Chi et al, 2010)

- Random-Cordillera (Exploratory) – decision made randomly
- DichGain-Cordillera –17 features
- NormGain-Cordillera –50 features, more training data

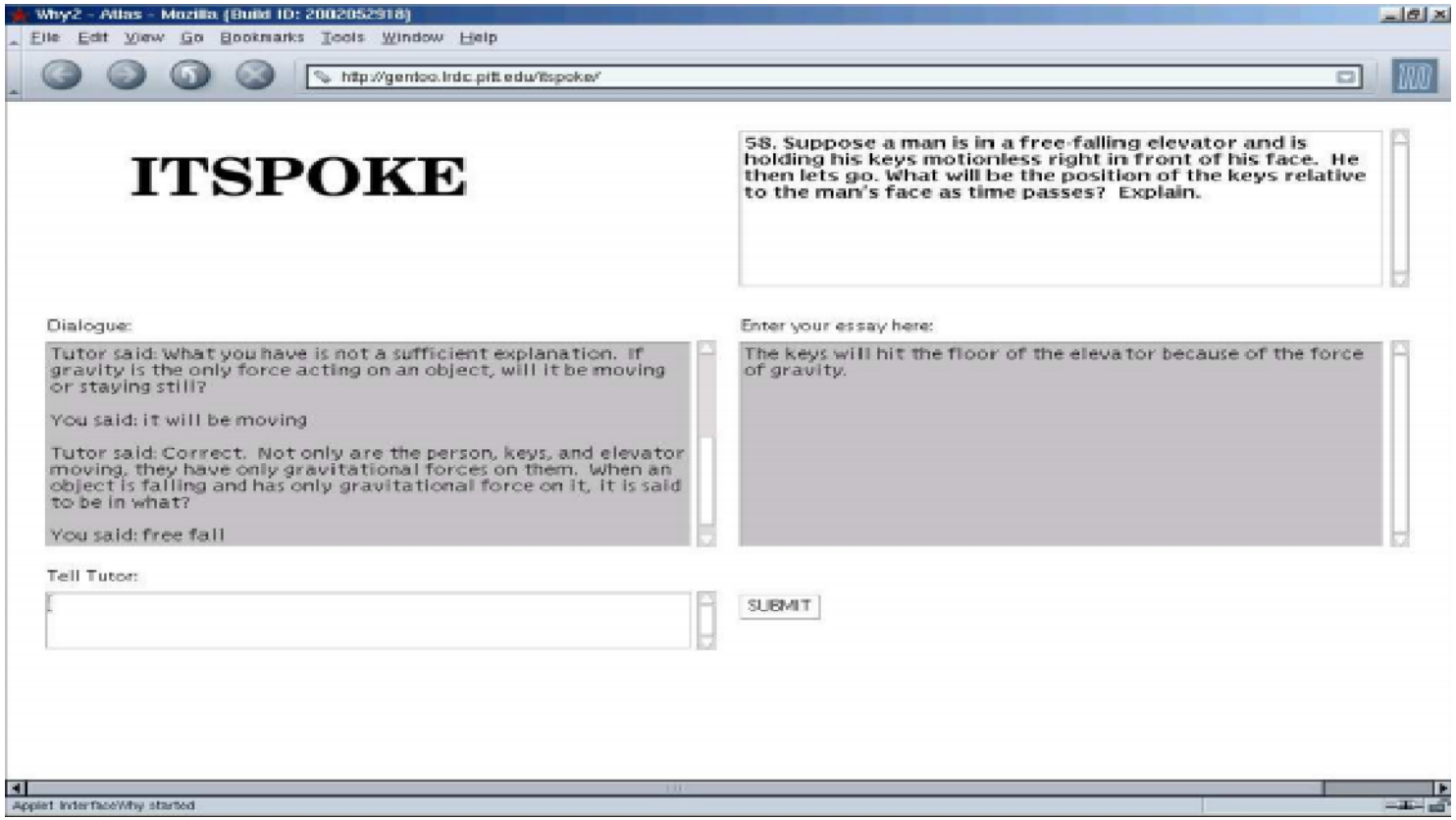


# Can a computerized system provide scaffolding and feedback? - Yes

- Most useful feature: step difficulty
- Features related to student's engagement in dialog also useful
- Features related to student's prior performance and background not useful

# Why a Spoken Dialog System for Tutoring?

- Student learning improves when they explain their thinking
- Responding appropriately to student emotion improves persistence
- Responding appropriately to student uncertainty improves learning




ITSPOKE (Litman & Silliman, 2004)

- Student types answer to qualitative physics problem
- System engages in dialog with student to correct and extend the essay
- Spoken dialog interface to Why2-Atlas, a text-based tutoring system

# ITSPOKE

- Finite State Dialog Manager: Question-Answer-Response
- Correct answer – go to next question
- Incorrect answer to an easy question – system gives correct answer and explanation
- Incorrect answer to a hard question – enters remediation subdialog

**ITSPOKE$_4$**: Good. Does the packet have an acceleration? If yes please specify its direction. *[SAQ, Advance, D1]*

**STUDENT$_4$**: horizontally forward yes *[non-uncertain, incorrect]*

**ITSPOKE$_5$**: Well... To find the packet's acceleration we will use Newton's 2nd law. What is the definition of this law? *[LAQ, Push, D2]*

**STUDENT$_5$**: an object in motion tends to stay in motion until it's act by an outside force *[uncertain, incorrect]*

**ITSPOKE$_6$**: Well... Newton's 2nd law says that the net force acting on an object equals the mass times the acceleration of the object. We can define Newton's 2nd law as F=m*a. Hence what can we determine about the acceleration of the packet given that we know the net force acting on it is not zero and points straight down? *[DAQ, Advance, D2]*

**STUDENT$_6$**: I don't know *[uncertain, incorrect]*

**ITSPOKE$_7$**: Well... *< further discussion of Newton's 2nd law... >* So once again on the original question. Does the packet have an acceleration? If yes please specify its direction. *[SAQ, PopUp, D1]*

Figure 1: ITSPOKE Annotated Dialogue Excerpt

# Responding to Student Uncertainty (Pon-Barry et al, 2006)

- Pretest – Work through problem – Posttest – Work through additional problem
- Normal Control Condition: Original ITSPOKE
- Experimental Condition: Treat uncertain correct answers as incorrect
- Random Control Condition: Randomly treat some correct answers as incorrect
- Wizard-of-Oz to categorize responses as correct/incorrect and certain/uncertain

# Experimental Results

- Different conditions had no impact on posttest scores
- Students who were correct and uncertain were more likely to remain correct in experimental group
- Students were less likely to remain uncertain of correct answers, but not statistically significant
- Further work with longer dialogs, better feedback for uncertain correct answers

# Automatically Detecting Uncertainty (Forbes-Riley et al, 2007)

- Labeled corpus – certain, uncertain, correct, incorrect
- Features:
  - Previous Question: Short Answer, Long Answer, Deep Answer, Repeat
  - Discourse Structure Depth: main dialog vs subdialog
  - Discourse Structure Transition: transitioning in and out of subdialog, continuing at current level

# Significant Features

- Long Answer Question – more uncertain answers
- Deep Answer Question – more uncertain and incorrect answers
- Short Answer Question – fewer uncertain and incorrect answers
- Main dialog – more correct, certain answers
- Subdialogs – more incorrect, uncertain answers
- Returning from subdialog to main dialog – more incorrect, uncertain answers

# Issues in Spoken Dialog Tutoring Systems

- Evaluation
- Using features of student speech
- Multimodality
- Mismatch between speech and actions