Questions and Answers about Questions and Answers:

Lessons from generating, scoring, and analyzing questions in a Reading Tutor for children

Jack Mostow

Project LISTEN (www.cs.cmu.edu/~listen)

AAAI Symposium on Question Generation keynote, Nov. 5, 2011, Arlington, VA

Questions about Questions

Target: What does it take to answer the questions?

Purpose: Why ask the questions?

Question type: In what form will the questions be output?

Answer type: In what form will responses be input?

Generation: How construct questions, answers, distracters?

Modality: What channels will convey questions and answers?

Assessment: How score answers? How generate feedback?

Evaluation: How to tell how well questions serve purpose?

What can questions target?

Reading

- Decoding: In "Word Swap," click on the "misread" word [Zhang ITS 08]

The lid sat in the loft.
was

What can questions target?

Reading

- Decoding: In "Word Swap," click on the "misread" word [Zhang ITS 08]
- Comprehension: *Click on the missing word.* [Mostow *TICL* 04]

"I am helping to lay up food for the winter," said the Ant, "and _____ you to do the same."

bother

recommend

chat

grasshopper

09]

Voc

Kno

11]
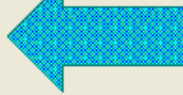
What can questions target?

Reading

- Decoding: In "Word Swap," click on the "misread" word [Zhang ITS 08]
- Comprehension: *Click on the missing word.* [Mostow *TICL* 04]
- Inter-sentential prediction: *Which will come next?* [Beck ITS 04]

Goodbye

Volume

Joe Valeri has read 3.0 minutes today.   Joe (level C) has read "Sentence Prediction Intervention" (level) 0 times, and 2478 words and 10 stories as

Back

Go

Which will come next?

"One doesn't get the offer of a beautiful princess and half a kingdom every day."

He might have a hundred horsemen, too, to back him up.

"That's the very thing for a man like me," thought the little tailor.

11/5/11 — Jack Mostow keynote — 5
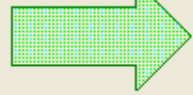
What can questions target?

Reading

- Decoding: In "Word Swap," click on the "misread" word [Zhang ITS 08]
- Comprehension: *Click on the missing word.* [Mostow *TICL* 04]
- Inter-sentential prediction: *Which will come next?* [Beck ITS 04]
- Monitor: *Did that make sense?*
- Self-question: *Why was the country mouse surprised?* [Chen AIED 09]
- Disengagement: hasty guessing [Beck AIED 05]

Vocabulary

- Recall: *Which means most like <word>?* [Aist 01]
- Recognize: *Which word means <definition>?* [*TICL* 04]
- Remind: Definition cloze [Gates QG 11]
- Disambiguate: *What does <word> mean here?*

Knowledge

- Fact: *Noiz arte du nahi duenak iritzia emateko aukera?* [Aldabe QG 11]
- Skill: *How many grams can a worker ant carry?* [Williams QG 11]
- Concept: …

How has Project LISTEN used questions?

1. Assess comprehension [Mostow et al., *TICL* 04]
2. Help comprehension [Beck et al., ITS 04]
3. Assess engagement [Beck, AIED 05]
4. Teach self-questioning [Mostow & Chen, AIED 09]
5. Model self-questioning [Chen et al., ITS 10]
6. Assess self-questioning [Chen et al., QG 11]
7. Help vocabulary learning [Gates et al., QG 11]

1. Assess comprehension [*TICL* 04]

Target: comprehend sentence

Purpose: assess comprehension while reading

Source: sentence in text

Question type: cloze

Answer type: multiple choice among 4 words from text

Generation: randomly pick sentence, word, and distracters

Modality: play recorded sentence and words; click on one

Assessment: original word? immediate correctness feedback

Evaluation: correlate against standard comprehension test

Student starts reading a story...

Goodbye

Jane Student has read 6 minutes today. Jane (level E) has read "The Ant And The Grasshopper" (level C) 0 times, and 93 words and 1 stories as of April 4, 2002 at 04:56:14 PM.

Back

Go

Project LISTEN Reading Tutor   Version: Jan 8 2002 15:24:55 F   Instructions: The Reading Tutor expects the student to read the whole sentence
Copyright 1995-1999
Carnegie Mellon University
U.S. Patent No. 5,920,838

Help

Say: In a field one s
Read together: In a
Play back last...

In a field one summer's day a Grasshopper was **hopping** about, chirping and singing to its heart's content.
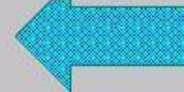
Now and then during the story…

Goodbye

Jane Student has read 8 minutes today.   Jane (level E) has read "The Ant And The Grasshopper" (level C) 0 times, and 93 words and 1 stories as of April 4, 2002 at 04:58:05 PM.

Back

Go

Help

Say: "Why not come
Read together: "Wh
Play back last...

"Why not come and chat with me," said the Grasshopper, "instead of toiling and moiling in that way?"

11/5/11

10

insert a cloze question…

"I am helping to lay up food for the winter," said the Ant, "and _____ you to do the same."

bother

recommend

chat

grasshopper

- Reading Tutor reads question and choices aloud
- Target and distracters are words from same story

… just before a sentence.

Goodbye

Jane Student has read 8 minutes today.   Jane (level E) has read "The Ant And The Grasshopper" (level C) 0 times, and 102 words and 1 stories as of April 4, 2002 at 04:58:35 PM.

Project LISTEN Reading Tutor   Version: Jan 8 2002 15:24:55 F   Instructions: The Reading Tutor expects the student to read the whole sentence
Copyright 1995-1999
Carnegie Mellon University
U.S. Patent No. 5,920,838

Help

Say: "I am helping t...
Read together: "I ar...
Play back last...

"I am helping to lay up food for the winter," said the Ant, "and recommend you to do the same."

11/5/11

12

What do cloze questions test?

"Why bother about _____?"

food

winter

dying
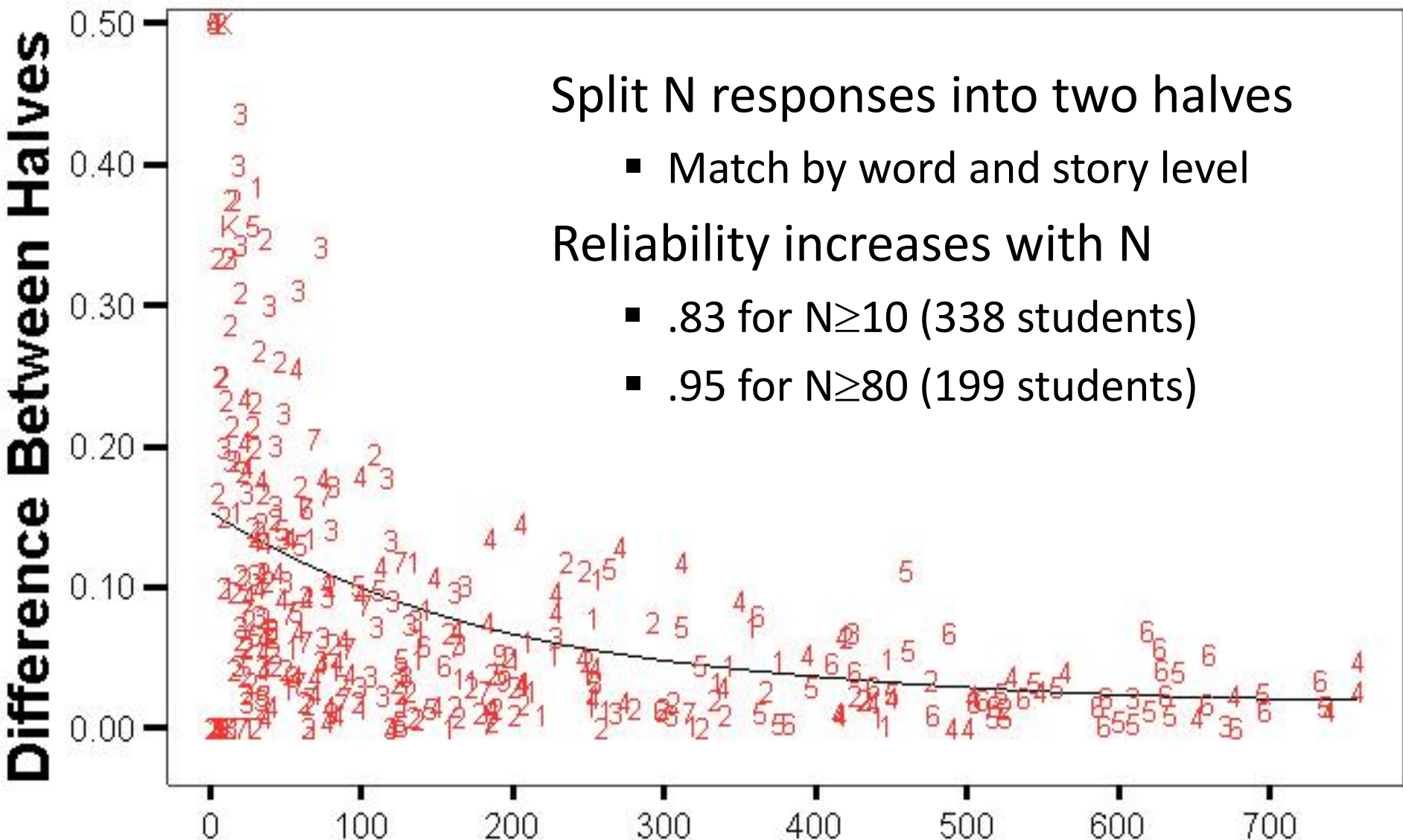
passed

Oct.01 – Mar.02 evaluation data

Reading Tutor asked 69,326 automated cloze questions
- 364 students in grades 1-9 at 7 schools
- 20-758 questions per student (median 136)
- 98% of questions answered – else *Goodbye*, *Back*, or timeout
- 24-88% correct (median 61%); 25% = chance

How much guessing? Hasty responses [J. Valeri]:
- 3,078 (4.5%) faster than 3 seconds, only 29% correct
- 3.9% per-student mean, but below 1% for most students
- Guessing rose from 1% in October to 11% by March

Reliability: Guttman split-half test

Split N responses into two halves
- Match by word and story level

Reliability increases with N
- .83 for N≥10 (338 students)
- .95 for N≥80 (199 students)

Difference Between Halves (y-axis) vs Number of Cloze Questions (x-axis)

What affects cloze difficulty?

Similarity of distracters to answer
- Part of speech [Hensler & Beck, ITS 06]
- Semantic class
- Consistency with local context
- Consistency with inter-sentential context

Vocabulary level of answer and distracters
- "Sight words" = 225 most frequent words of English [Dolch list]
- "Easy words" = 226…3,000
- "Hard words" = 3,001…25,000
- "Defined words" = marked as warranting explanation

Text level of story
- Grade K, 1, 2, 3, 4, 5, 6, 7

Cloze performance at 4 word levels x 8 text levels predicts Woodcock Reading Mastery Test comprehension (R = .84)

2a. Help comprehension [ITS 04]

Target: comprehend text by questioning

Purpose: scaffold comprehension while reading

Source: none

Question type: *Wh-*

Answer type: multiple choice

Generation: scripted generic questions and choices

Modality: play recorded prompt and words; click on one

Assessment: none

Evaluation: test efficacy on ensuing cloze performance

Generic *Wh-* questions: initial

Generic prompt (meta-question):

- *Click on a question you can answer, or click Back to reread the sentence*

Generic 1-word questions as choices:

- *Who? What? When? Where? Why? How? So?*

Evaluation: failed 2002 user test

- Meta-question confusing
- 1-word questions too vague to map to text

Generic *Wh-* questions: revised

Volume

Charlotte Tester has read 15.1 minutes today.
Charro (level B) has read "Sight Words" (level) 0
times, and 213 words and 1 stories as of July 15,

Back

Go

What has happened so far?

facts were given	a meeting
a problem is being solved	nothing yet
a problem has been solved	I don't know
a problem	
an introduction	
a mistake	

Jack Mostow keynote

2b. Help comprehension [ITS 04]

Target: comprehend text by predicting

Purpose: scaffold comprehension while reading

Source: none

Question type: *Which will come next?*

Answer type: multiple choice

Generation: answer = next sentence; distracters = 2 following

Modality: play recorded prompt and sentences; click on one

Assessment: original sentence? immediate correctness feedback

Evaluation: 41% right; test efficacy on ensuing cloze performance

Sentence prediction
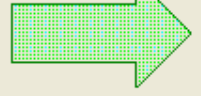
Which will come next?

"One doesn't get the offer of a beautiful princess and half a kingdom every day."

He might have a hundred horsemen, too, to back him up.

"That's the very thing for a man like me," thought the little tailor.

2003 evaluation data

Reading Tutor asked 23,372 randomly inserted questions

- 252 students in grades 1-4+? who used for an hour or more
- 6,720 generic *what, when, where*
- 1,865 *Which will come next?* sentence prediction
- 15,187 cloze
- On **average**, one question every 10 sentences

Evaluation results

Variable	Helps/Hurts	p
# 3W questions	☺	0.023
# sentence predictions	☺	0.072
# cloze questions	=	
# recent 3W questions	☹	0.074
# recent cloze questions	=	
Time since prior question (sec)	☺	0.036
Time since start of story (sec)	☹	0.14

Effect of recent questions

Time since previous question (sec)
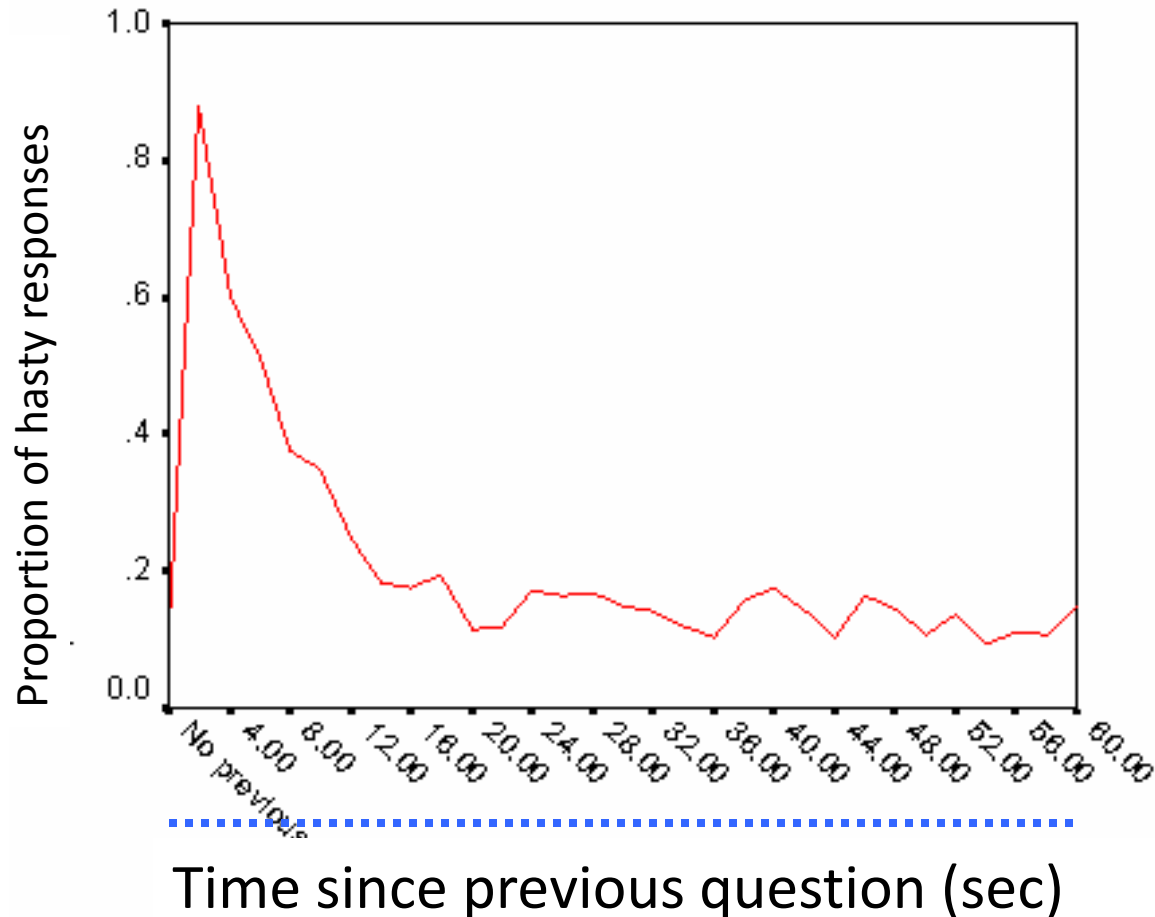
- Response time < 3 sec indicates disengagement
- Use to track (dis-)engagement [Beck AIED 05]

Effect of question type

QG costs (lowest → highest)

	None	Trivial	Simple	Complex	Manual
Generation					
Question		Generic	Cloze	NLP	Scripted
Answer	None	Given	= text	NLP	Scripted
Distracter	None	Generic	From text	NLP	Scripted
Modality					
Output		Menu	Text	TTS, visual	Prerecorded
Input	None	Click	Keyboard	ASR, gesture	Transcribed
Assessment					
Scoring	None	M/C	= answer	NLP	Manual
Feedback	None	Generic	answer	NLP	Scripted

Jack Mostow keynote

QG costs (lowest → highest)

	None	Trivial	Simple	Complex	Manual
Generation					
Question		Generic	Cloze	NLP	Scripted
Answer	None	Given	= text	NLP	Scripted
Distracter	None	Generic	From text	NLP	Scripted
Modality					
Output		Menu	Text	TTS, visual	Prerecorded
Input	None	Click	Keyboard	ASR, gesture	Transcribed
Assessment					
Scoring	None	M/C	= answer	NLP	Manual
Feedback	None	Generic	answer	NLP	Scripted