Fine-grained Cognitive Assessment based on Free-form Input for Math Story Problems

Bastiaan Heeren   Johan Jeuring   Sergey Sosnovsky   Paul Drijvers
Peter Boon   Sietske Tacoma   Jesse Koops   Armin Weinberger
Brigitte Grugeon-Allys   Françoise Chenevotot-Quentin   Jorn van Wijk
Ferdinand van Walree

September 5, EC-TEL 2018, Leeds

Open Universiteit
www.ou.nl

Universiteit Utrecht

Cito

UNIVERSITÄT DES SAARLANDES

UPEC

LDAR   LABORATOIRE DE DIDACTIQUE ANDRÉ REVUZ   RECHERCHE EN DIDACTIQUE DES SCIENCES

Introduction

Using ICT for assessing mathematics achievement:

- ▶ Targeted at 12-15 years old pupils
- ▶ Setting up algebraic expressions and equations, and simplifying and solving them

Detailed assessment of free-form answers to math story problems:

- ▶ Analysis of intermediate steps
- ▶ Determining the high-level solution approach
- ▶ Identification of misconceptions

'Magical trick' task

numworx Setting up algebraic expressions >

LESSON   Setting up algebraic expressions

Task 05   **Magical trick?**

A student says to her peer:
"Choose a number, add 8, multiply the result by 3, subtract 4, add the initial number, divide by 4, add 2, and subtract the initial number.
You will end up with 7."

Is this true for any starting number? Explain your answer.

Your work

$(x+8)*3-4+x/4+2-x = 7$

$(3(x+8)-4+x)/4+2-x = 7$

$(3x+24-4+x) /4+2 = 7$

$(4x+20) /4+2-x = 7$

$x+5+2-x = 7$

$7=7$
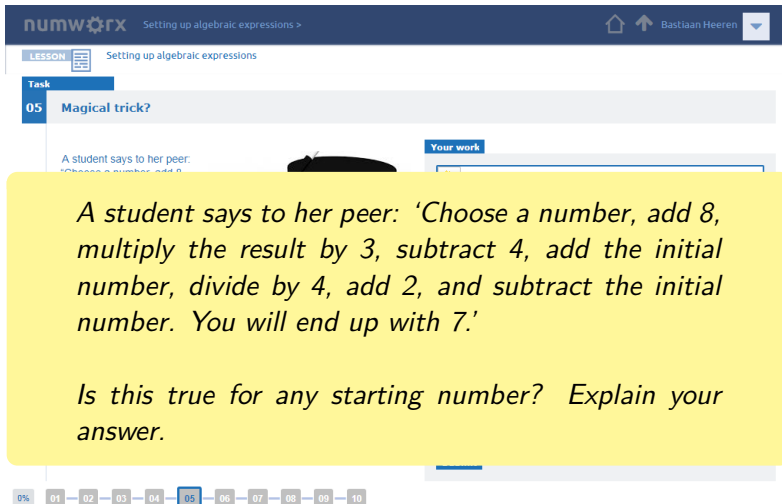
For any number x, we end up with 7.

Submit

0% 01 02 03 04 05 06 07 08 09 10

'Magical trick' task

LESSON 🖹 Setting up algebraic expressions

Task
05 **Magical trick?**

Your work

A student says to her peer: 'Choose a number, add 8, multiply the result by 3, subtract 4, add the initial number, divide by 4, add 2, and subtract the initial number. You will end up with 7.'

Is this true for any starting number? Explain your answer.

0% 01 — 02 — 03 — 04 — 05 — 06 — 07 — 08 — 09 — 10

Assessment approaches

- Conventional assessment tests are usually unidimensional: instead, we focus on obtaining a detailed picture, with the pupil's strengths and weaknesses
- Easily verifiable input (e.g. multiple choice questions) may restrict the complexity of the tasks
- Structuring the interaction provides scaffolding to a pupil, which is less desirable in an assessment scenario
- Asking follow-up questions about intermediate steps may extend the assessment session

Assessment approaches

- ▶ Conventional assessment tests are usually unidimensional: instead, we focus on obtaining a detailed picture, with the pupil's strengths and weaknesses
- ▶ Easily verifiable input (e.g. multiple choice questions) may restrict the complexity of the tasks
- ▶ Structuring the interaction provides scaffolding to a pupil, which is less desirable in an assessment scenario
- ▶ Asking follow-up questions about intermediate steps may extend the assessment session
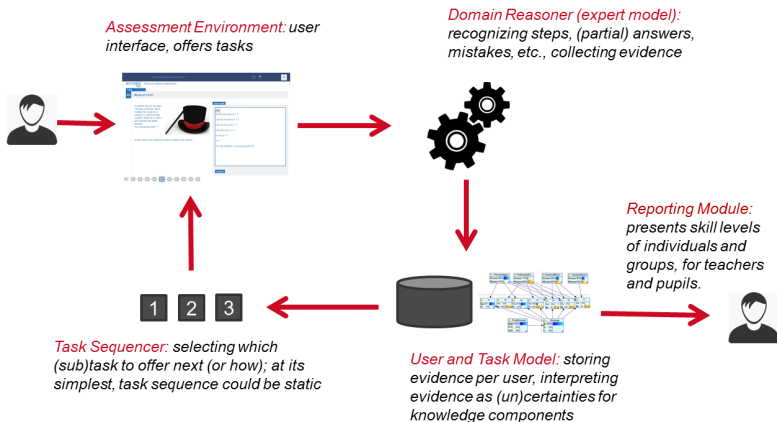
Our research goal: analysing free-form input to math story problems in an assessment setting

High-level architecture

Assessment Environment: user interface, offers tasks

Domain Reasoner (expert model): recognizing steps, (partial) answers, mistakes, etc., collecting evidence

Reporting Module: presents skill levels of individuals and groups, for teachers and pupils.

Task Sequencer: selecting which (sub)task to offer next (or how); at its simplest, task sequence could be static

User and Task Model: storing evidence per user, interpreting evidence as (un)certainties for knowledge components

Examples of solution approaches

```
((5+8)*3-4+5)/4+2-5 = 7
```
▶ global, arithmetic

```
5+8 = 13; 13*3 = 39;
39-4 = 35; 35+5 = 40;
40/4 = 10; 10+2 = 12;
12-5 = 7
```
▶ step-by-step, arithmetic

Examples of solution approaches

```
((5+8)*3-4+5)/4+2-5 = 7
```
▶ global, arithmetic

```
5+8 = 13; 13*3 = 39;
39-4 = 35; 35+5 = 40;
40/4 = 10; 10+2 = 12;
12-5 = 7
```
▶ step-by-step, arithmetic

```
5+8*3-4+5/4+2-5 = 7
```
▶ global, arithmetic, priority mistakes

```
5+8 = 13*3 = 39-4 = 35+5
= 40/4 = 10+2 = 12-5 = 7
```
▶ step-by-step, arithmetic, erroneous

Examples of solution approaches

((5+8)*3-4+5)/4+2-5 = 7
▶ global, arithmetic

5+8 = 13; 13*3 = 39;
39-4 = 35; 35+5 = 40;
40/4 = 10; 10+2 = 12;
12-5 = 7
▶ step-by-step, arithmetic

5+8*3-4+5/4+2-5 = 7
▶ global, arithmetic, priority mistakes

5+8 = 13*3 = 39-4 = 35+5
= 40/4 = 10+2 = 12-5 = 7
▶ step-by-step, arithmetic, erroneous

(x+8)*3 = 3x+24
3x+24-4 = 3x+20
3x+20+x = 4x+20
(4x+20)/4 = x+5
x+5+2 = x+7
x+7-x = 7
▶ step-by-step, algebraic

x+8*3-4+x/4+2-x
= 2x+20/4+2-x
= 2x+5+2-x
= x+7
▶ simplification mistakes

Domain reasoner

Analysis of input proceeds in 3 phases:

1. Extract mathematical expressions
   – Some natural language-specific pre-processing
   – Compare x+3 with 3x5=15

2. Parse extracted expressions

3. Recognise solution approach
   – View recognition as a parsing problem
   – Recogniser must be flexible enough to deal with imperfections and mistakes

▶ For some answers, the analysis fails to produce evidence

Task models and user models

- We use Bayesian Networks for inference under uncertainty
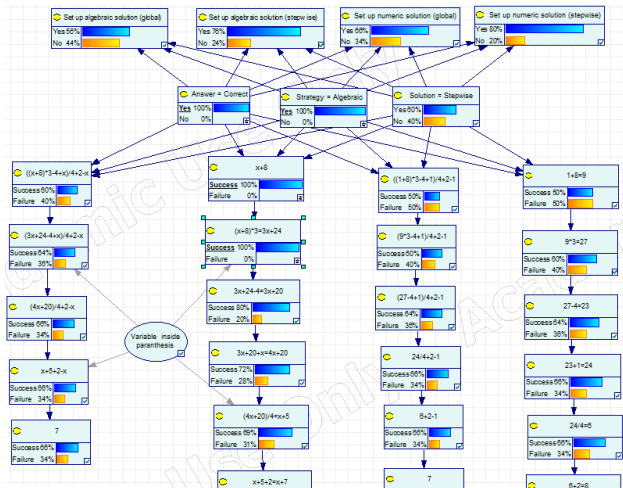
Preliminary evaluation

We tested the domain reasoner on a 'magical trick' dataset:

- 2956 free-form answers, collected between 2011 and 2015
- For nearly 80%, the solution approach was recognised

We organised two small-scale pilots:

- Pilots in Germany (N=19) and the Netherlands (N=22)
- Questionnaire, then 10 tasks, followed by a short survey
- Questionnaire focused on prior knowledge and usability
- Pupils think they did well in the test
- Again, the solution approach was recognised for nearly 80%

Results for small-scale pilots

	Task	German pilot						Dutch pilot				
		N	Al	Ar	Gr	Em	Un	N	Al	Ar	Em	Un
1.	Making a square	19	.53			.05	.42	22	.55			.45
2.	Matryoshka	17		.76		.06	.18	12		.17	.33	.50
3.	Car rental	18	.39		.39	.06	.17	22	.82			.18
4.	Pattern	18	.11	.67			.22	19	.11	.53		.37
5.	Magical trick	18	.06	.11		.56	.28	20			.70	.30
6a.	Rectangle area	18	.94			.06		22	.95			.05
6b.	Rectangle area	18	.67			.06	.28	22	.77			.23
6c.	Rectangle area	18				.67	.33	22	.36		.55	.09
7b.	Theatre rate	18	.33		.11	.50	.06	21	.76		.05	.19
9a.	Area of triangle	15	.33			.47	.20	22	.77		.05	.18
9b.	Area of triangle	15	.20			.60	.20	22	.73		.05	.23
9c.	Area of triangle	15	.13			.73	.13	22	.77		.23	
10.	V-pattern	15	.53			.27	.20	22	.73		.05	.23
	Overall	222	.33	.12	.04	.30	.21	270	.59	.04	.14	.22

See the paper for more details

Workshop on assessment in mathematics

Advise-Me

On Friday October 19, 2018, we are organising a multiplier event in Heerlen, the Netherlands. On this day, we will give an overview of the results of the project, and have asked two international experts to give their view on assessment in mathematics. You are very welcome to attend (registration is required).

- ▶ Presentations from the Advise-Me strategic partnership
- ▶ Invited speakers: Chris Sangwin (University of Edinburgh) and Vanda Luengo (LIP6)

Conclusions

- We have developed a framework for fine-grained cognitive assessment of free-form solutions to math story problems
- Our domain reasoner extracts the mathematics, analyses intermediate steps, and tries to recognise the solution approach
- We use Bayesian task models and a user model for the inference, storage, and update of user knowledge
- The solution approach was recognised in nearly 80% of the answers for two small-scale pilot studies
- In the future, we will organise more extensive evaluations