How challenging are Bebras tasks?

Carlo Bellettini   Violetta Lonati   Dario Malchiodi
Mattia Monga   Anna Morpurgo   Mauro Torelli

Dept. of Computer Science
Università degli Studi di Milano, Milan, Italy

0 $_{0011100011101}$

ALADDIN

Aladdin laboratorio di divulgazione
e didattica dell'informatica

ITiCSE — Vilnius, July 6, 2015

Bebras IRT analysis

M. Monga

Tasklets

Measurement model

Results

Conclusions

Bebras' *tasklets*

Tasklet
A small and moderately challenging task that enables an entertaining learning experience.

Tasklets should be [Dagienė & Futschek, 2008]:

- fun and attractive
- independent of specific curricular activities
- adequate for contestants' age
- solvable in three minutes

Tasklets are conceived and tuned during an international workshop with ≈ 80 participants from > 30 countries.

Psychologists made a test of laterality in the classroom consisting of three tasks and answers were stored in a computer. The tasks were:

1. Give a clap: they recorded whether left or right hand was above.

2. Look at the picture and immediately tell, which animal do you see: they recorded whether student saw a head of rabbit or duck.

3. Clasp hands: they recorded whether left or right thumb was above.
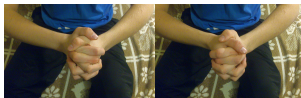
How many different codes should there be at least?
A) 1   B) 3   C) 8   D) 16.
Is this tasklet easy? medium? hard?

Measuring tasklet difficulty

Bebras IRT
analysis

M. Monga

Tasklets

Measurement
model

Results

Conclusions

How to *measure* tasklet difficulty? Not an easy task... even *a posteriori*.

- Number of failures? But the sample of solvers could be biased.
- Time spent on solution? Maybe it is just long to read.

We resorted to Item Response Theory model, a statistical approach routinely used in massive educational assessment like PISA.
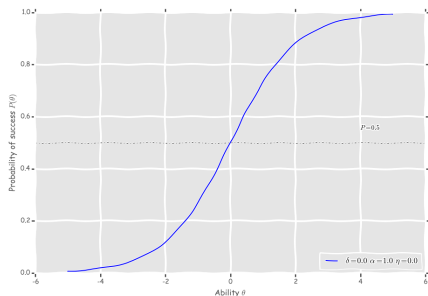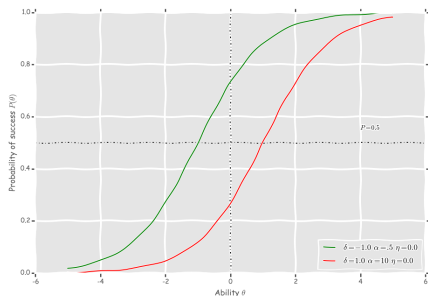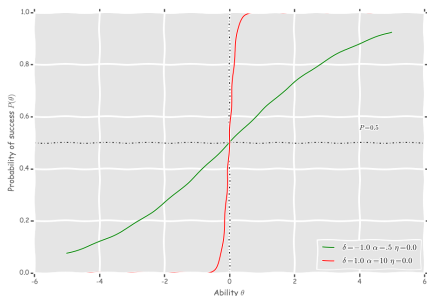
The IRT model

Bebras IRT
analysis

M. Monga

Tasklets

Measurement
model

Results

Conclusions

$$P(\theta) \quad = \quad \eta + \frac{(1-\eta)}{1 + e^{-\alpha \cdot (\theta - \delta)}}$$

$$\theta \quad = \quad \text{team } \textit{ability}$$

$$\delta \quad = \quad \text{tasklet } \textit{difficulty}$$

$$\alpha \quad = \quad \text{tasklet } \textit{discrimination}$$

$$\eta \quad = \quad \text{tasklet chance of being guessed}$$

The IRT model

Bebras IRT
analysis

M. Monga

Tasklets

Measurement
model

Results

Conclusions

$$P(\theta) \quad = \quad \eta + \frac{(1 - \eta)}{1 + e^{-\alpha \cdot (\theta - \delta)}}$$

$$\theta \quad = \quad \text{team } \textit{ability}$$

$$\delta \quad = \quad \text{tasklet } \textit{difficulty}$$

$$\alpha \quad = \quad \text{tasklet } \textit{discrimination}$$

$$\eta \quad = \quad \text{tasklet chance of being guessed}$$
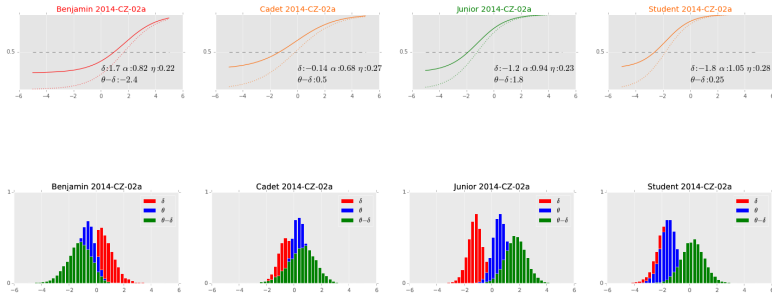
The IRT model

Bebras IRT
analysis

M. Monga

Tasklets

Measurement
model

Results

Conclusions

$$P(\theta) \quad = \quad \eta + \frac{(1-\eta)}{1 + e^{-\alpha \cdot (\theta - \delta)}}$$

$$\theta \quad = \quad \text{team } ability$$

$$\delta \quad = \quad \text{tasklet } difficulty$$

$$\alpha \quad = \quad \text{tasklet } discrimination$$

$$\eta \quad = \quad \text{tasklet chance of being guessed}$$

Stochastic fitting of the model

684 teams (2784 pupils) in 4 categories (Benjamin [6th-7th],
Cadet [8th-9th], Junior [10th-11th], Student [12th-13th]),
11'483 answers.
IRT model fitted with a Markov Chain Monte Carlo approach
(implementend in Stan, http://mc-stan.org/, a probabilistic
programming language for Bayesian statistical inference)

Data available at: https://bitbucket.org/mmonga/bebrastan

IRT analysis

Bebras IRT analysis

M. Monga

Tasklets

Measurement model

Results

Conclusions

Interestingly enough, Juniors showed a greater ability than Students (high profile Students went to Olympiads?)
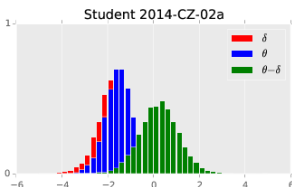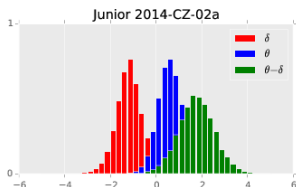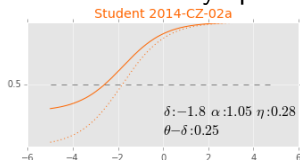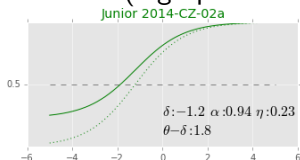
The tasklet is "easy" but since Students have low performances on average, it was classified as *medium*.

Easier than expected

Some tasklets resulted easier than expected.
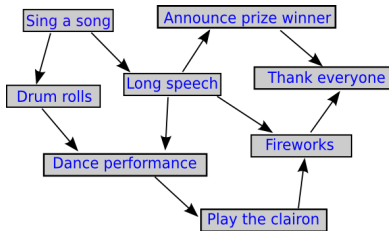
In the beaver community there are many steps involved when organizing a ceremony. In order to organize a proper ceremony these steps must be taken in the correct order.
The arrows in the picture indicate which step(s) must be taken before another step can be taken. Make a proper ceremony.

Rated hard for Cadet, medium for Junior. ⤳ Easy: the authors rated the tasklet focusing on the general problem of topological sort instead of the *small* instance.

Harder than expected

We rescaled difficulty because teams may work in parallel (original tasklet was conceived for an individual player): but the real issue is in how to approach the problem or younger pupils might not notice the opportunity.

Check which order (out of four) is compatible with the observations of taller beavers in front and behind each one.
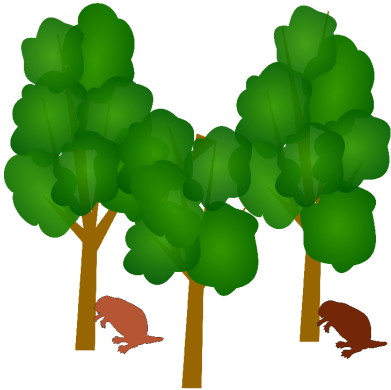
Much harder for youngest

Three trees, two beavers: a beaver cuts a tree in 4 minutes, but they never work in parallel on the same tree. What is the minimum time for cutting all the trees?

Open question: the players had to write what should happen in each minute.
Very hard for Benjamins and Cadets (with a high *discrimination*), very easy for Juniors.

Conclusions

Predicting tasklet difficulty is hard!

- Quantitative analyses of the answers may give insight on common mistakes in prediction (e.g., consider the general complex problem instead of the small instance)

- Know your population! (e.g., Students show often a lesser average ability than Juniors. . .)

- Future work: use quantitative data to improve ranking or to get tuned packages of tasklets for specific classroom purposes.