

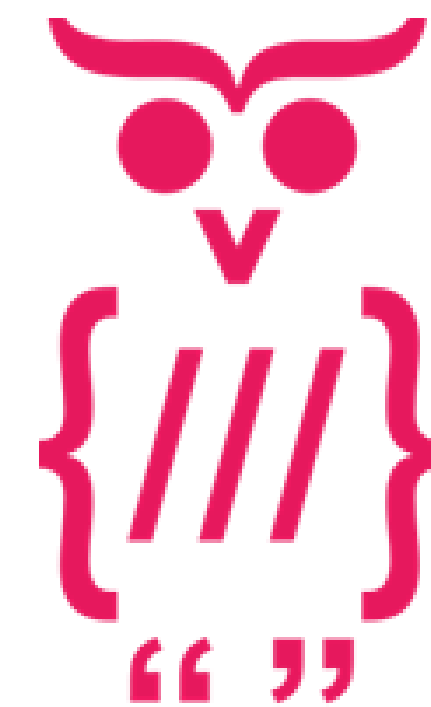


Analyzing task difficulty in a Bebras contest using Cuttle



Dutch Foundation for the Olympiad of Informatics

- Eljakim Schrijvers
(contest system, data)
- Willem van der Vegt
(research on task difficulty)
- Bouke van der Bijl
(presenting)



informatics
olympiade



Dutch Foundation for the Olympiad of Informatics

Three contests

- **Beverwedstrijd (Bebras contest)**
(primary and secondary education) 20000 contestants
- **National Olympiad in Informatics**
(secondary education) 400 contestants
- **CodeCup (game programming)**
(open, world wide) 100 contestants





International Challenge on Informatics
and Computational Thinking



- Started in 2004 in Lithuania
- In 2018 over 50 countries
- 2.7 million contestants
- Tasks from a common task pool, developed at the annual Bebras Workshop
- In the Netherlands:
 - 5 age groups (8-18 years)
 - 40 minutes for 15 questions
 - A second round at a university
- We offer the challenge as a contest





- We announce the expected difficulty level for each task
 - 5 easy
 - 5 medium
 - 5 hard
- The score for a task depend on the announced level:
 - Easy: +6 or -2
 - Medium: +9 or -3
 - Hard: +12 or -4

How Hard Will this Task Be? Developments in Analyzing and Predicting Question Difficulty in the Bebras Challenge

Three kinds of difficulty:

1. Content difficulty
2. Stimulus difficulty
3. Task difficulty

- Task designers tend to overestimate content diff.
- There are some tools to help predict difficulty level
- Easy to identify parameters would be helpful



Bebras Challenge

Task categories

- ALP: Algorithms and Programming
- DSR: Data, Data Structures, and Representations
- CPH: Computer Processes and Hardware
- COM: Communications and Networking
- ISS: Interactions, Systems, and Society.



Bebras Challenge

Answer categories

- Multiple choice text
- Multiple choice images
- Interactive
- Open answer integer
- Open answer text







- Here information by KIM on the properties of the system
- Here information by Kim on the use of the system

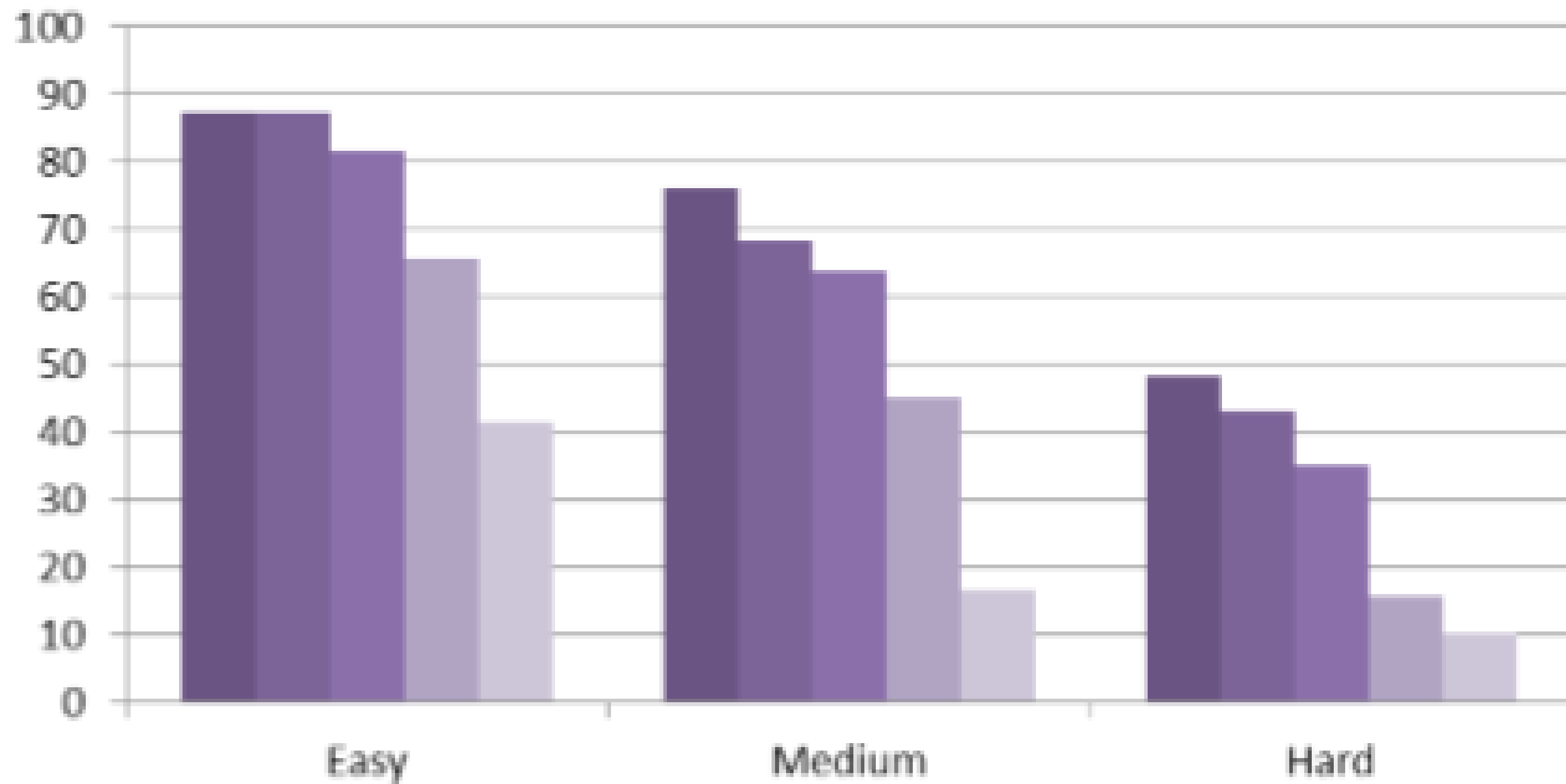
Questions

Analyzing task difficulty in a Bebras contest using Cuttle

1. It is possible to use the Cuttle system to collect data that can be useful for analyzing task difficulty in Bebras?
2. And can we formulate questions for future research, based on the findings using Cuttle?



Results 2017 age group VI



Difficulty estimations by task designer, workshop and in the contest

Task-ID	Original difficulty level	Workshop difficulty level	Contest difficulty level
2017-CA-12	III-easy	V-medium	VI-easy
2017-IS-01	V-medium	V-hard	VI-easy
2017-BE-05	IV-medium	IV-medium	VI-easy
2017-RU-03	II-medium	IV-easy	VI-easy
2017-IR-07	IV-easy	V-medium	VI-easy
2017-CA-07	V-hard	V-medium	VI-medium
2017-PL-02	V-hard	V-hard	VI-medium
2017-CH-01b	IV-easy	V-medium	VI-medium
2017-CZ-04c	V-medium	V-hard	VI-medium
2017-CH-07b	VI-hard	V-hard	VI-medium
2017-KR-07	IV-medium	VI-hard	VI-hard
2017-SK-12a	VI-medium	VI-hard	VI-hard
2017-UK-04	VI-hard	VI-hard	VI-hard
2017-KR-03	VI-medium	VI-hard	VI-hard
2017-SI-04	V-medium	V-medium	VI-hard

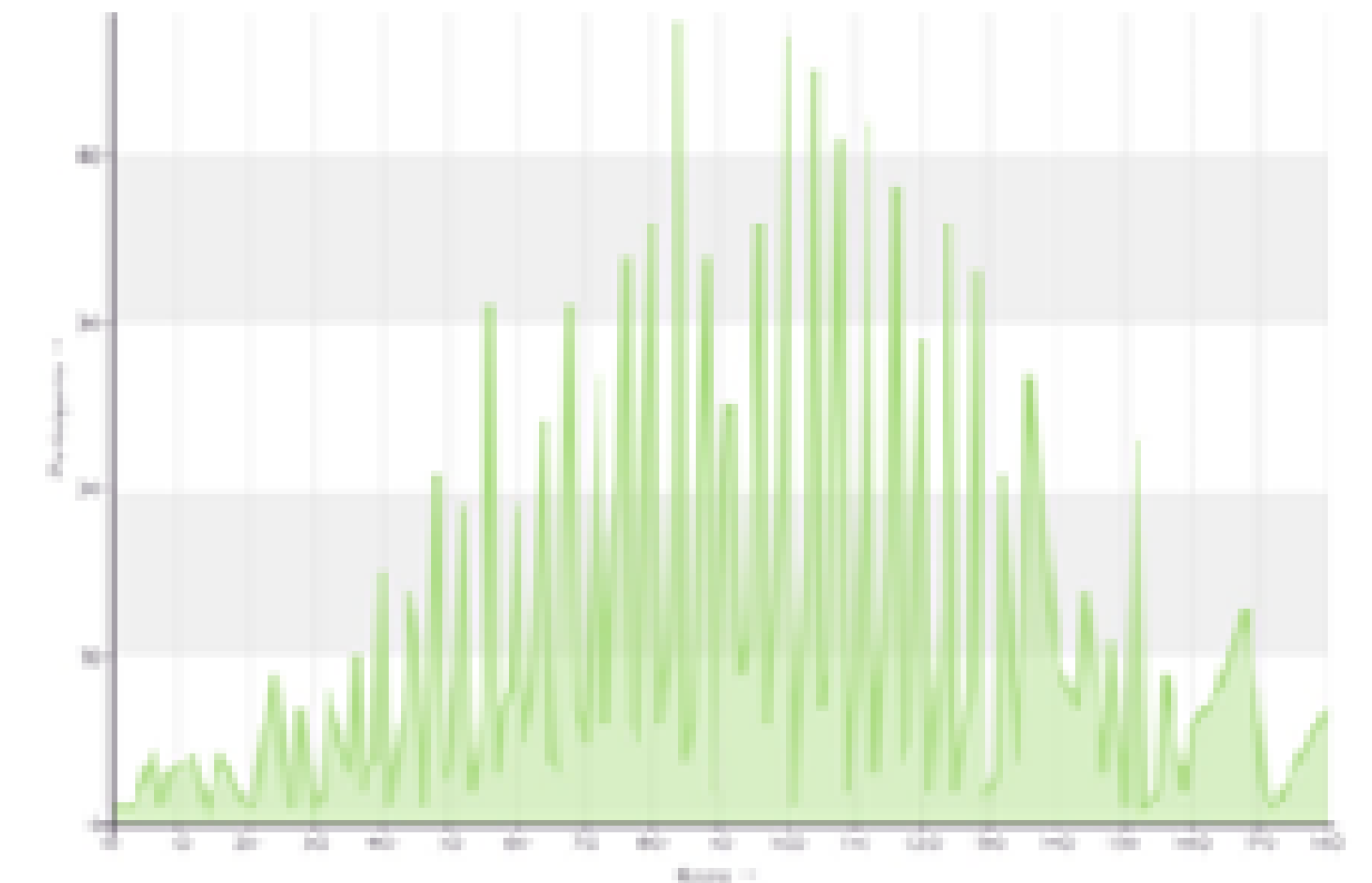
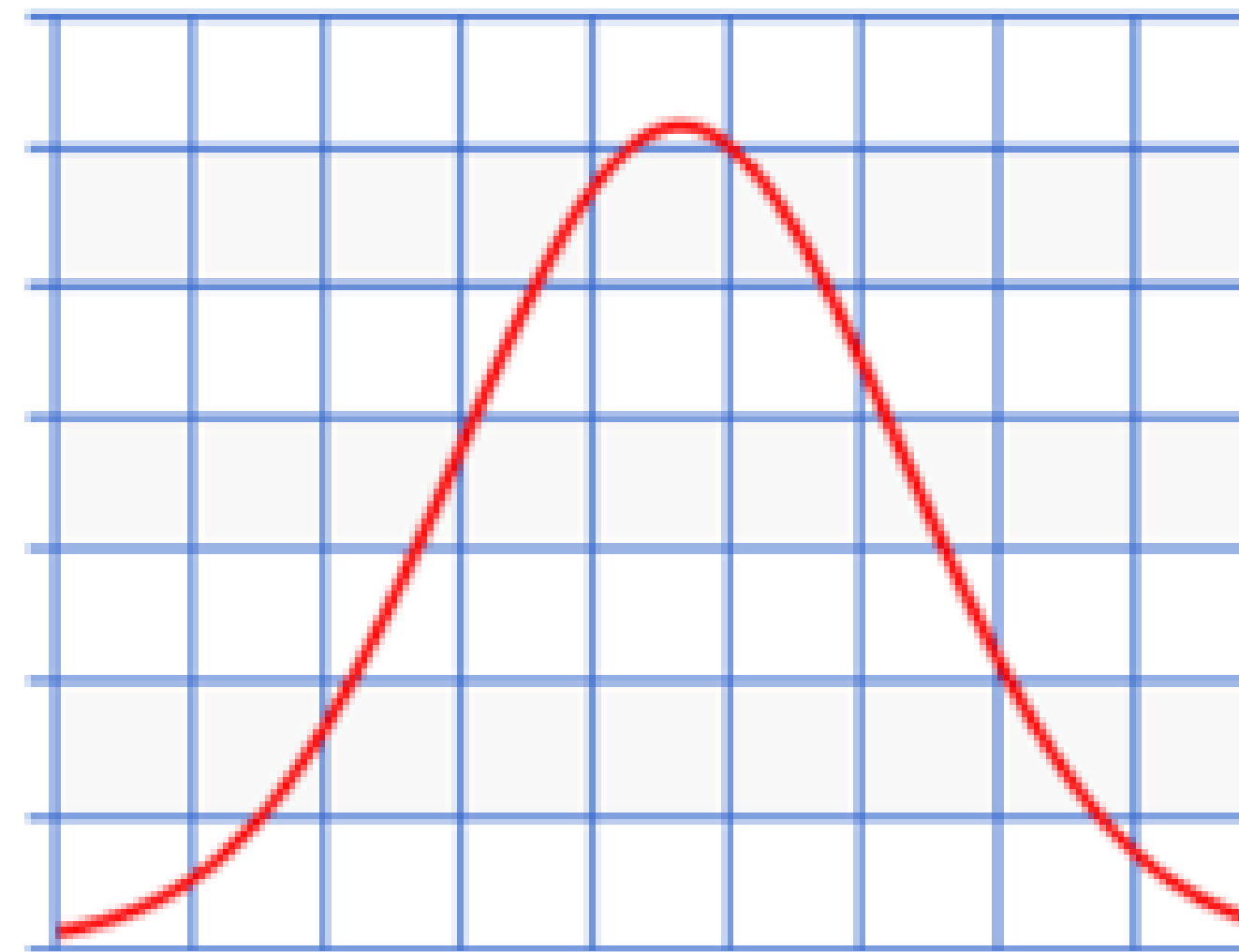
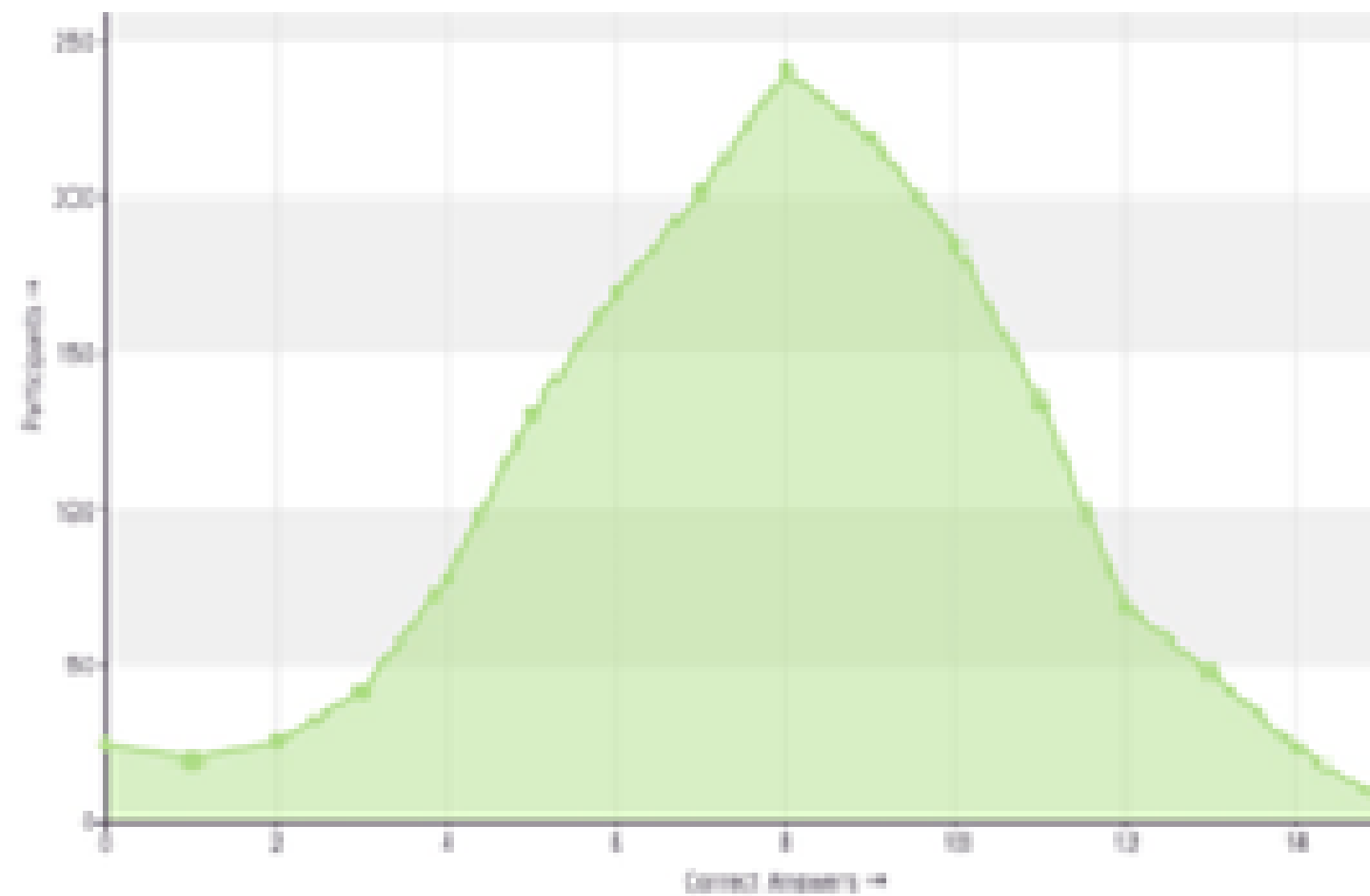


Categories, CS topics and answer types in the contest

Task-ID	Category	Computer Science Topic	Answer type
2017-CA-12	DSR	Dynamic programming	Multiple Choice Text
2017-IS-01	ALP	Sequence, binary system	Multiple Choice Text
2017-BE-05	(ALP/DSR)	A path in a graph	Multiple Choice Images
2017-RU-03	DSR	Gray code	Interactive
2017-IR-07	COM/ISS/ALP	Search in social network graph	Multiple Choice Text
2017-CA-07	ALP	Assignment problem	Interactive
2017-PL-02	(ALP/DSR)	Levenshtein distance	Open Ended Integer
2017-CH-01b	(ALP)	Programming in a maze	Interactive
2017-CZ-04c	ALP	A path in a graph	Interactive
2017-CH-07b	(ALP/DSR)	Maximum flow problem	Open Ended Integer
2017-KR-07	(DSR)	Image compression	Multiple Choice Text
2017-SK-12a	(ALP)	Turing machine	Multiple Choice Images
2017-UK-04	ALP	Assignment problem	Multiple Choice Text
2017-KR-03	(ALP)	Optimization, scheduling	Open Ended Text
2017-SI-04	(ALP/DSR)	Binary counting	Open Ended Integer



Number of correct answers and score distribution, compared with a normal distribution



General analytics for the contest

Group Analytics	
analytics score lowest	0
analytics score highest	180
analytics score average	92.9
analytics score stddev	32.4
analytics relative stddev	0.3
analytics max possible	180
analytic avg p score	51.6
analytics participants	1621

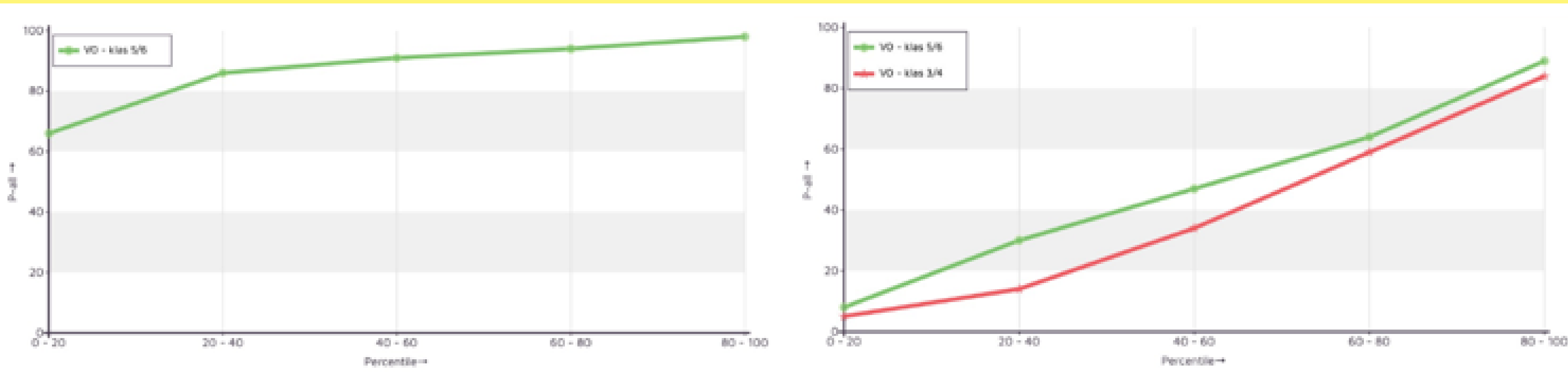


Used measures

- P_{all} is the percentage of correct answers across all participants
- R_{it} gives the correlation between the score of a task and the overall score as a percentage
- %NA is the percentage of contestants that did not answer the question



Specific task details



Results per category

Categories	n	P _{all}	R _{it}	%NA
ALP/DSR	4	45.5	40.8	13.01
Only ALP	8	50.1	43.9	15.41
Only DSR	3	67.2	44.5	9.75



Relation category and difficulty

Tasks on algorithms (ALP) are harder than tasks on information and structures (DSR)

- ALP
 - Dynamic, have impact on working memory; you have to be able to keep track of changes
- DSR
 - Static, you need to grasp the relations and answer the question

The combination is even harder



Results per answer type

Answer type	n	Pall	Rit	%NA
MC Text	5	59.8	43.0	10.25
MC Images	2	62.4	45.0	10.37
Open Ended Integer	3	31.6	39.7	13.87
Open Ended Text	1	15.7	45.5	32.10
Interactive	4	62.7	44.6	14.72

These results are as expected



P-values for other age groups

	III	IV	V	VI
2017-IS-01			69.0	86.4
2017-RU-03	22.8	35.7	53.8	65.7
2017-CA-07			61.8	75.9
2017-PL-02		34.9	49.3	68.1
2017-CH-01b		51.7	62.0	63.8
2017-CZ-04c	3.5	6.9	20.5	45.2
2017-CH-07b		4.1	9.2	16.5
2017-KR-07			39.9	48.4
2017-SK-12a			32.6	43.1



Conclusions

It is possible to use Cuttle to get useful information about question difficulty

First observations:

- The more open the answer type, the harder the task
- Dynamic tasks are harder than static ones

Future research:

- Why are some questions much harder for younger kids and have other questions about the same difficulty for all age groups?

