

Software Engineering Exams – the influence of multiple-choice questions (MCQs)

Michael Ritzschke

Humboldt-Universität zu Berlin
Department of Computer Science
Software Engineering

13th Workshop "Software Engineering Education and Reverse Engineering"
Bansko, Bulgaria, 26 – 31 August 2013

A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

Contents

- Overview
- Three types of examination questions
- The influence of different assessments for MCQs
- Only MCQs?
- Conclusions

Overview of SE exams at HU

criterion		WS 12/13	WS 11/12
basis	time	120 min.	
	points	180	
number of	tasks	43	43
	subtasks	65	64
number of students	accepted	92	101
	registered	76	87
	participated	71	81

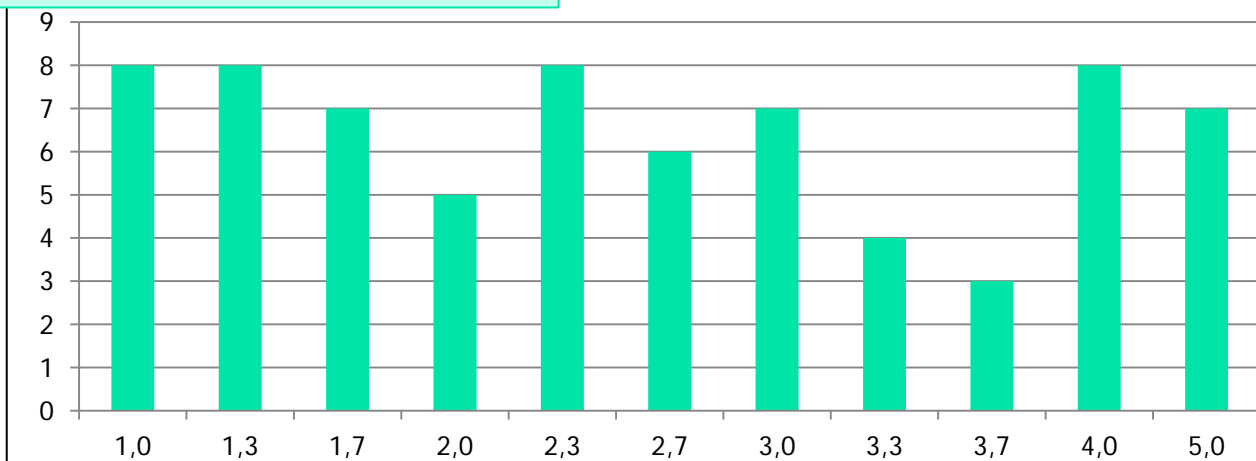
The main results in 2013

Scale

Mark	1,0	1,3	1,7	2,0	2,3	2,7	3,0	3,3	3,7	4,0
Points	153,0	144,0	135,0	126,0	117,0	108,0	99,0	90,0	81,0	72,0
%	85,0	80,0	75,0	70,0	65,0	60,0	55,0	50,0	45,0	40,0

1,0 – best grade: excellent
 4,0 – just passed
 5,0 - failed

Results



- 64 students passed
- 7 students failed
- Average grade: 2,64

Our basis for the statistical evaluation: individual points for each task

students			tasks																		overall		
			1	2	3	4	5	13	14	15	16	44	45	50	51	52	60	61	62	63		64	
student nb.	nb.	points	1	1	2	1	3	a	b	c	d	a	b	b	c	33	39	40	41	42	43		
			1	1	2	1	3	1	1	2	1	3	3	3	3	3	3	2	2	3	2	180,0	
533591	1		1,0	1,0	2,0	2,0	2,0	1,0	1,0	2,0	1,0	2,0	3,0	3,0	0,0	4,0	3,0	2,0	2,0	2,0	2,0	1,0	151,0
539272	2		1,0	1,0	2,0	1,0	1,0	1,0	1,0	2,0	0,0	3,0	3,0	2,0	3,0	0,0	1,0	0,0	2,0	2,5	2,0		104,0
543474	3		1,0	1,0	2,0	2,0	4,0	1,0	1,0	2,0	0,0	3,0	3,0	3,0	3,0	1,0	3,0	1,0	2,0	2,5	2,0		107,0
537231	4		1,0	1,0	2,0	2,0	3,0	1,0	1,0	2,0	1,0	3,0	1,5	3,0	3,0	3,0	3,0	2,0	1,0	3,0	1,5		126,5
539658	5		0,0	1,0	2,0	2,0	2,0	1,0	1,0	2,0	0,0	3,0	3,0	0,0	0,0	0,0	3,0	0,0	0,0	3,0	2,0		98,0
53967	6		1,0	1,0	2,0	2,0	2,0	1,0	1,0	2,0	0,0	3,0	3,0	0,0	0,0	0,0	3,0	0,0	0,0	3,0	2,0		94,0
539690	7		1,0	1,0	2,0	2,0	2,0	1,0	1,0	2,0	0,0	3,0	3,0	0,0	0,0	0,0	3,0	0,0	0,0	3,0	2,0		122,5
535372	8		1,0	1,0	2,0	2,0	1,0	1,0	1,0	2,0	0,0	3,0	3,0	0,0	0,0	0,0	3,0	0,0	0,0	3,0	2,0		117,5
539748	9		0,0	0,0	0,0	0,0	0,0	1,0	1,0	2,0	1,0	3,0	3,0	3,0	3,0	1,0	1,5	1,0	1,0	3,0	1,0		131,0
540812	73		1,0	1,0	2,0	2,0	2,5	1,0	1,0	2,0	1,0	3,0	3,0	3,0	3,0	3,0	2,0	2,0	2,0	2,5	1,0		153,5
540789	74		1,0	1,0	2,0	2,0	2,0	1,0	1,0	2,0	1,0	3,0	1,5	3,0	3,0	3,0	3,0	2,0	2,0	3,0	2,0		157,5
540748	75		1,0	1,0	2,0	2,0	1,0	1,0	1,0	2,0	0,0	1,0	3,0	0,0	0,0	3,0	3,0	2,0	2,0	2,5	1,0		108,0
540755	76		0,0	1,0	2,0	2,0	1,5	1,0	1,0	2,0	0,0	3,0	3,0	0,0	0,0	2,0	1,5	0,0	0,0	0,0	0,0		89,5
537472	77		1,0	1,0	2,0	2,0	1,5	1,0	1,0	2,0	0,0	3,0	3,0	0,0	0,0	1,0	1,0	2,0	2,0	3,0	2,0		103,5
540727	78		0,0	1,0	2,0	2,0	1,0	1,0	1,0	2,0	0,0	3,0	3,0	0,0	0,0	0,0	2,0	2,0	1,0	1,5	0,0		90,0
533440	79		0,0	1,0	2,0	2,0	2,0	1,0	1,0	2,0	1,0	3,0	3,0	2,0	2,0	1,0	0,0	1,0	1,0	3,0	1,5		114,5
532115	80		1,0	1,0	2,0	2,0	2,5	1,0	1,0	2,0	1,0	3,0	3,0	3,0	2,0	3,0	3,0	2,0	1,0	2,0	1,0		150,0
529417	81		1,0	1,0	2,0	2,0	2,0	1,0	1,0	2,0	0,0	3,0	2,0	1,0	3,0	0,0	3,0	2,0	2,0	2,0	1,0		117,0
		min	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0		38,0
		max	1,0	1,0	2,0	2,0	4,0	1,0	1,0	2,0	1,0	3,0	3,0	3,0	3,0	4,0	3,0	2,0	2,0	3,0	2,0		162,0
		average %	66,7	96,3	96,3	96,9	56,6	93,8	93,8	95,1	46,9	85,0	78,6	42,4	38,7	58,4	74,9	58,0	71,3	74,1	54,3		63,1

2 points at task 4 for student 6

A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

Contents

- Overview
- Three types of examination questions
- The influence of different assessments for MCQs
- Only MCQs?
- Conclusions

Types of questions in 2013: 3 times more multiple choice than in 2012

year	criteria	knowledge	multiple choice	skills	overall
2013	quantity (subtasks)	35	15	15	65
	possible points	72	44	64	180
	amount %	40	24,4	35,6	
2012	quantity (subtasks)	45	5	14	64
	possible points	97	15	68	180
	amount %	53,9	8,3	37,6	

New in 2013: negative points for wrong answers

10. (3 points) On which basis cost estimation for software projects is possible?

- | | | |
|--|------------------------------|-----------------------------|
| a) Preliminary requirements specification | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| b) Requirements specification | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| c) Use Case Diagrams | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| d) previous projects | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| e) Division of a system to sub-systems | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| f) Information of the portion of a special phase as part of the whole system development | <input type="checkbox"/> yes | <input type="checkbox"/> no |

Assessment 2013:

- 0.5 points for each correct answer
- -0.25 points for a wrong answer

Reason: to prevent students from guessing answers

Results in 2013: less points for multiple choice questions

2013				
critierion	knowledge	multiple choice	Skills	overall
quantity (subtasks)	35	15	15	65
possible points	72	44	64	180
average points %	65,2	59,4	63,0	63,1
min	3,5	6	9	34
max	70	41,5	64	168,25

2012				
critierion	knowledge	multiple choice	Skills	overall
quantity (subtasks)	45	5	14	64
possible points	97	15	68	180
average points %	62,4	69,3	62,8	63,1
min	11	5,5	17	39
max	92,5	14	63	162

Reason:
negative points

MCOs with negative points:
assessment below the overall average
MCOs without negative points:
assessment above the overall average

A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

Contents

- Overview
- Three types of examination questions
- The influence of different assessments for MCQs
- Only MCQs?
- Conclusions

Three different assessments for MCQs

A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

- Negative points for wrong answers (2013)
- Zero points as a lower limit for all MCQs (i.e. never negative points in the total)
- Zero points for wrong answers (2012)

Example

for 3 possible assessments of MCQs

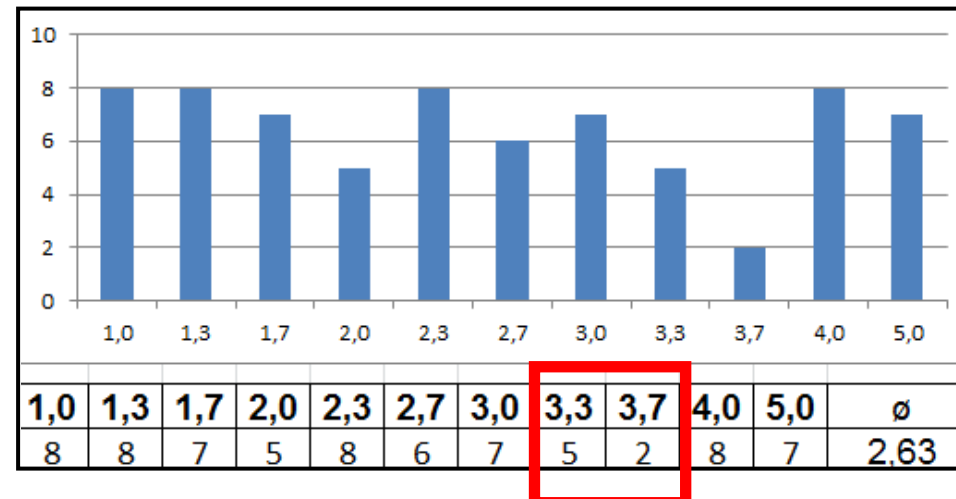
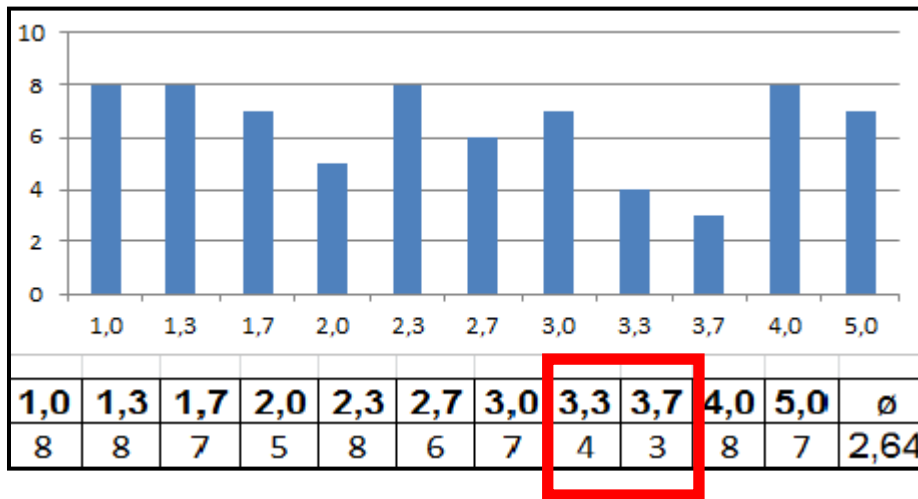
Assume: correct answers always „yes“

option number	Student answers		points 2013 (penalty)	Modifi- cation 1	Modifi- cation 2 (like 2012) (bonus)
	yes	no			
1	x		1	1	1
2		x	-0,5	-0,5	0
3		x	-0,5	-0,5	0
4		x	-0,5	-0,5	0
5			0	0	0
		total	-0,5	0	1

Modification 1: What happens if we use assessment zero points instead of negative points for multiple choice?

with reduction
(negative points possible)

with 0 points instead negative points



- 34 of 71 students would get more points (0,25...2,75)
- but only 1 Student would get a better grade

Modification 2: What happens if we use assessment without reduction for multiple choice?

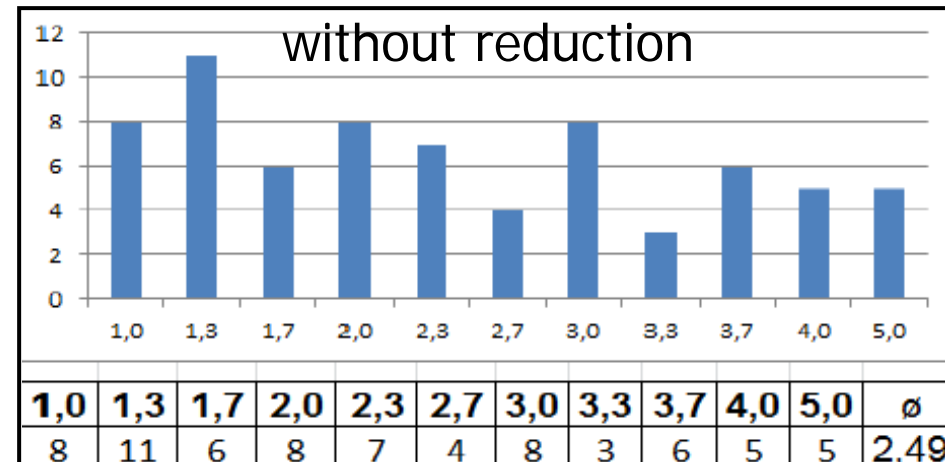
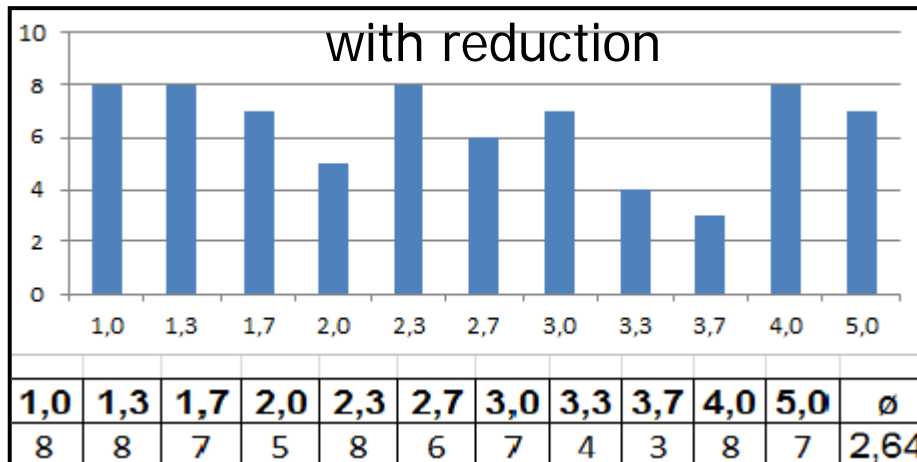


2013 multiple choice

critierion	with reduction	without reduction	2012
possible points	44	44	15
average points %	59,4	69,8	69,3

We would get nearly the same result as in 2012

Main results for modification 2 would be much better



- all students would get more points (0,25...9,0), in average 4,5 points
- **28 of 71 students** would get a better grade (!)
- average: 2,49 instead of 2,64 (!)

Comparison between good, middle and bad student-groups

	Number of additional points in case of no reduction			average points %		number of better marks
	min	max	ø	with reduction	without reduction	
students						
10 best	0,25	4,50	1,88	83,6	87,4	0
10 worst	3,25	8,50	5,78	35,6	48,8	2
10 middle	2,25	7,50	5,36	55,5	67,7	7
all	0,25	9,00	4,57	59,4	69,8	28

- better students get only few points more
- worst students would get most additional points
- students in the middle would have most profit

A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

Contents

- Overview
- Three types of examination questions
- The influence of different assessments for MCQs
- Only MCQs?
- Conclusions

A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

Exams completely based on MCQs

Why?

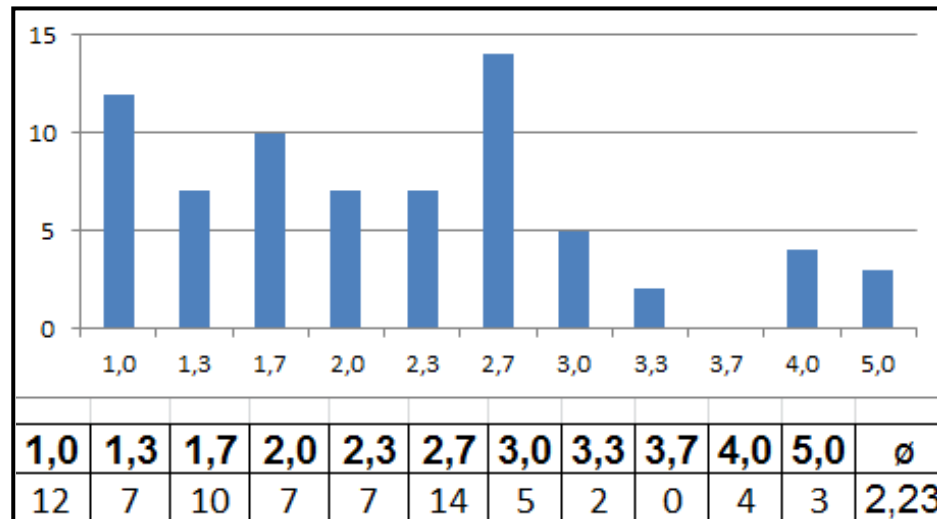
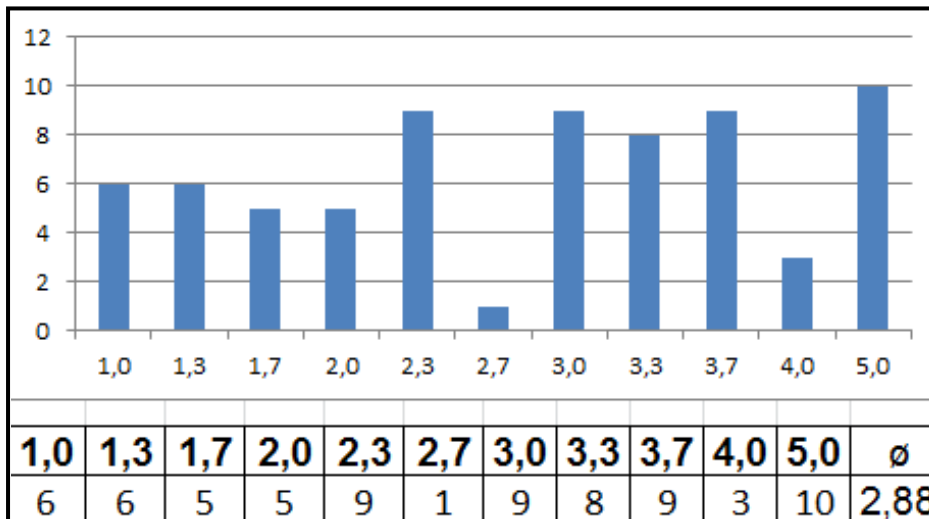
- Advantage: MCQs easier and faster to correct
- MCQs can be corrected by non-professionals or automatically
- Question: Will an exam completely based on MCQs properly reflect students performance?

What happens if we use only MCQs?

(Base: 15 questions 2013, same scale as shown at the beginning)

with reduction (penalty)

without reduction (bonus)



The result is strongly determined by the kind of assessment – in each case: MCQ exams acceptable

What happens if we use only MCQs?

with reduction?	only MCQs			∅	∅
	number of marks				
	better	worse	equal		
yes	15	39	17	2,88	2,64
no	35	14	22	2,23	2,49

→ **With reduction: MCQs lead to worse results**

Without reduction: MCQs lead to better results

What happens if we use only MCQs?

with reduction?	only MCQs			∅	overall
	better	worse	equal		
yes	15	39	17	2,88	2,64
no	35	14	22	2,23	2,49

With reduction: MCQs lead to worse results

→ Without reduction: MCQs lead to better results

A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

Contents

- Overview
- Three types of examination questions
- The influence of different assessments for MCQs
- Only MCQs?
- Conclusions

A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

Conclusions

- There are many references about using MCQs
- For instance in /2/ we find a good historical overview and some reflections for writing good multiple-choice tests, /3/ presents a checklist for writing effective MCQs ... and so on
- Writing good multiple-choice tests is difficult

/2/ Simon: Wrong is a relative concept: part marks for multiple-choice questions. Proceedings of the 13th Australasian Computing Education Conference (ACE 2011), Perth, Australia, January 2011. CRPIT, Vol. 114, S. 47-53

/3/ Woodfort, K., Bankroft, P.: Multiple Choice Questions Not Considered harmful. Proceedings of the 7th Australasian Computing Education Conference (ACE 2005), Newcastle, Australia. CRPIT, Vol. 42, S. 109-115

A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

Conclusions

- An additional problem we offer in this presentation:
The bonus-penalty kind of assessment
- The **bonus assessment** leads to significant better results: if we would have used it in 2013 instead of the **penalty assessment** 40% of the students would get a better grade (!)
- For better students the type of assessment is insignificant, **most profit** from the bonus assessment would have **students with mean accomplishments**

A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

Conclusions

- The bonus assessment allows guessing answers, but the penalty assessment is disputed in Germany (action at law are possible) /4/
- Students at Germany called a court because of the penalty system. The court gave them right and prohibited negative points. Reason: Positive points (positive knowledge) will be neglected by wrong answers at other questions.

/4/ <http://www.pflichtlektuere.com/16/05/2012/multiple-choice-verwirrung-um-minuspunkte/>

A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

Conclusions

Thank you for your attention!

References

- /1/ Ritzschke, M.: Software Engineering exams of Bachelor students - some conclusions. 12th DAAD-workshop "[Joint Course in Software Engineering](#)". Opatija, Croatia, September, 3rd - 8th 2012. [Paper \(pdf\)](#)
- /2/ Simon: Wrong is a relative concept: part marks for multiple-choice questions. Proceedings of the 13th Australasian Computing Education Conference (ACE 2011), Perth, Australia, January 2011. CRPIT, Vol. 114, S. 47-53
- /3/ Woodfort, K., Bankroft, P.: Multiple Choice Questions Not Considered harmful. Proceedings of the 7th Australasian Computing Education Conference (ACE 2005), Newcastle, Australia. CRPIT, Vol. 42, S. 109-115
- /4/ <http://www.pflichtlektuere.com/16/05/2012/multiple-choice-verwirrung-um-minuspunkte/>