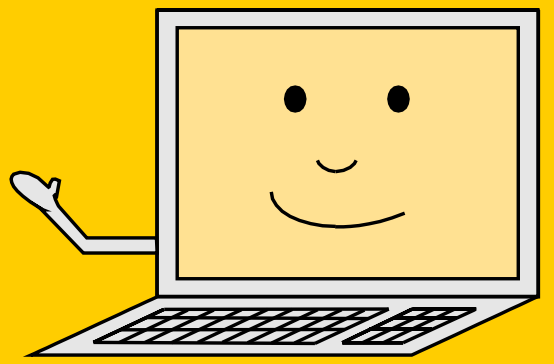
Computers as teaching assistants

Automated intelligent feedback in university statistics education



Sietske Tacoma
PhD candidate

United Nations (2013)

Principle 1: *Relevance, Impartiality, and Equal Access*

Official statistics provide an indispensable element in the information system of a democratic society

[…].

To this end, official statistics that meet the test of practical utility are to be compiled and made available

[…]

to honour citizens' entitlement to public information.

Fundamental Principles of Official Statistics*

United Nations

For more information: unstats.un.org

The General Assembly,

Recalling recent resolutions[1] of the General Assembly and the Economic and Social Council highlighting the fundamental importance of official statistics for the national and global development agenda,

Bearing in mind the critical role of high-quality official statistical information in analysis and informed policy decision-making in support of sustainable development, peace and security, as well as for mutual knowledge and trade among the States and peoples of an increasingly connected world, demanding openness and transparency,

Bearing in mind also that the essential trust of the public in the integrity of official statistical systems and confidence in statistics depend to a large extent on respect for the fundamental values and principles that are the basis of any society seeking to understand itself and respect the rights of its members, and in this context that professional independence and accountability of statistical agencies are crucial,

Stressing that, in order to be effective, the fundamental values and principles that govern statistical work have to be guaranteed by legal and institutional frameworks and be respected at all political levels and by all stakeholders in national statistical systems,

Endorses the Fundamental Principles of Official Statistics set out below, as adopted by the Statistical Commission in 1994[2] and reaffirmed in 2013, and endorsed by the Economic and Social Council in its resolution 2013/21 of 24 July 2013:

* General Assembly Resolution 68/261 adopted on 29 January 2014. The "titles" of the Principles are not part of the original text.

[1] These include General Assembly resolution 64/267 on World Statistics Day and Economic and Social Council resolutions 2005/13 on the 2010 World Population and Housing Census Programme, 2006/6 on strengthening statistical capacity and 2013/21 on the Fundamental Principles of Official Statistics.

[2] For the original preamble used on the occasion of the initial adoption of the Fundamental Principles in 1994, see chapter V of the report of the Statistical Commission on its special session (*Official Records of the Economic and Social Council, 1994, Supple-*

Principle 1: ***Relevance, Impartiality, and Equal Access***
Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.

Principle 2: ***Professional Standards, Scientific Principles, and Professional Ethics***
To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.

Principle 3: ***Accountability and Transparency***
To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.

Principle 4: ***Prevention of Misuse***
The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.

Principle 5: ***Sources of Official Statistics***
Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents.

Principle 6: ***Confidentiality***
Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.

Principle 7: ***Legislation***
The laws, regulations and measures under which the statistical systems operate are to be made public.

Principle 8: ***National Coordination***
Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system.

Principle 9: ***Use of International Standards***
The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.

A typical task from a university statistics course

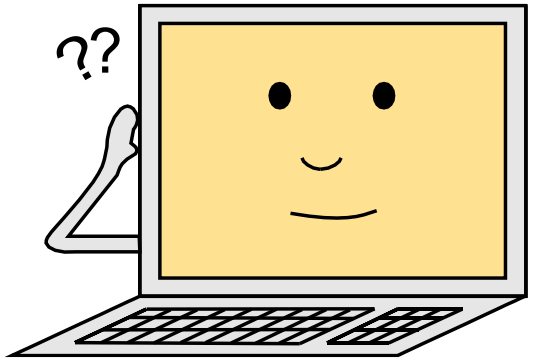
How would you react if the grade you received for an exam is much lower than you had expected? Research suggests that most students think they can handle such situations better than their peers, but some students think their coping is worse than that of their peers.

In this study, participants were asked to read a scenario of a negative event and indicate how this event would influence their well-being (-5: worsen much, +5: improve much). Next, they were asked to imagine this same event from the perspective of a peer. The difference between both judgements was noted.

Suppose that for the sample of $n = 25$ students the mean difference score was $M_D = 1.28$ points (own judgement minus judgement peer) with standard deviation $SD = 1.50$. Based on these data, can you conclude that there is a significant difference between the own judgements and judgements of peers? Use a test with $\alpha = .05$.

Question 1

How can computers be used to give useful automated feedback in multi-step hypothesis-testing tasks?



VanLehn, 2011

Multi-step hypothesis testing task

a The test variable μ is Choose

b Null hypothesis is Choose

Alternative hypothesis is Choose

Significance level is Choose

c The population parameters (temperatures in the 20th century) are:

Mean $\mu =$ [ ] °C Round off to 2 decimals.

Standard deviation $\sigma =$ [ ] °C

d Under the assumption that the null hypothesis is true, the data from the 21st century can be regarded as a sample. The sample parameters are:

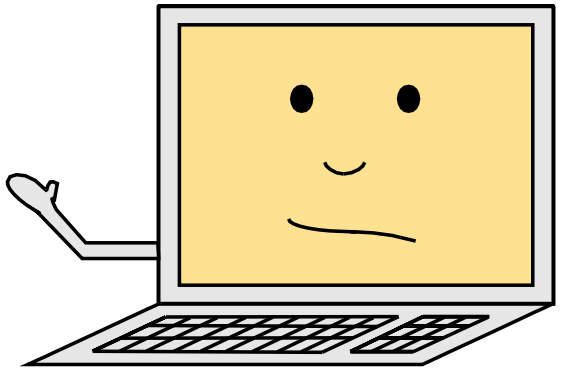
Mean $m =$ [ ] °C Round off to 2 decimals.

Sample size $n =$ [ ]

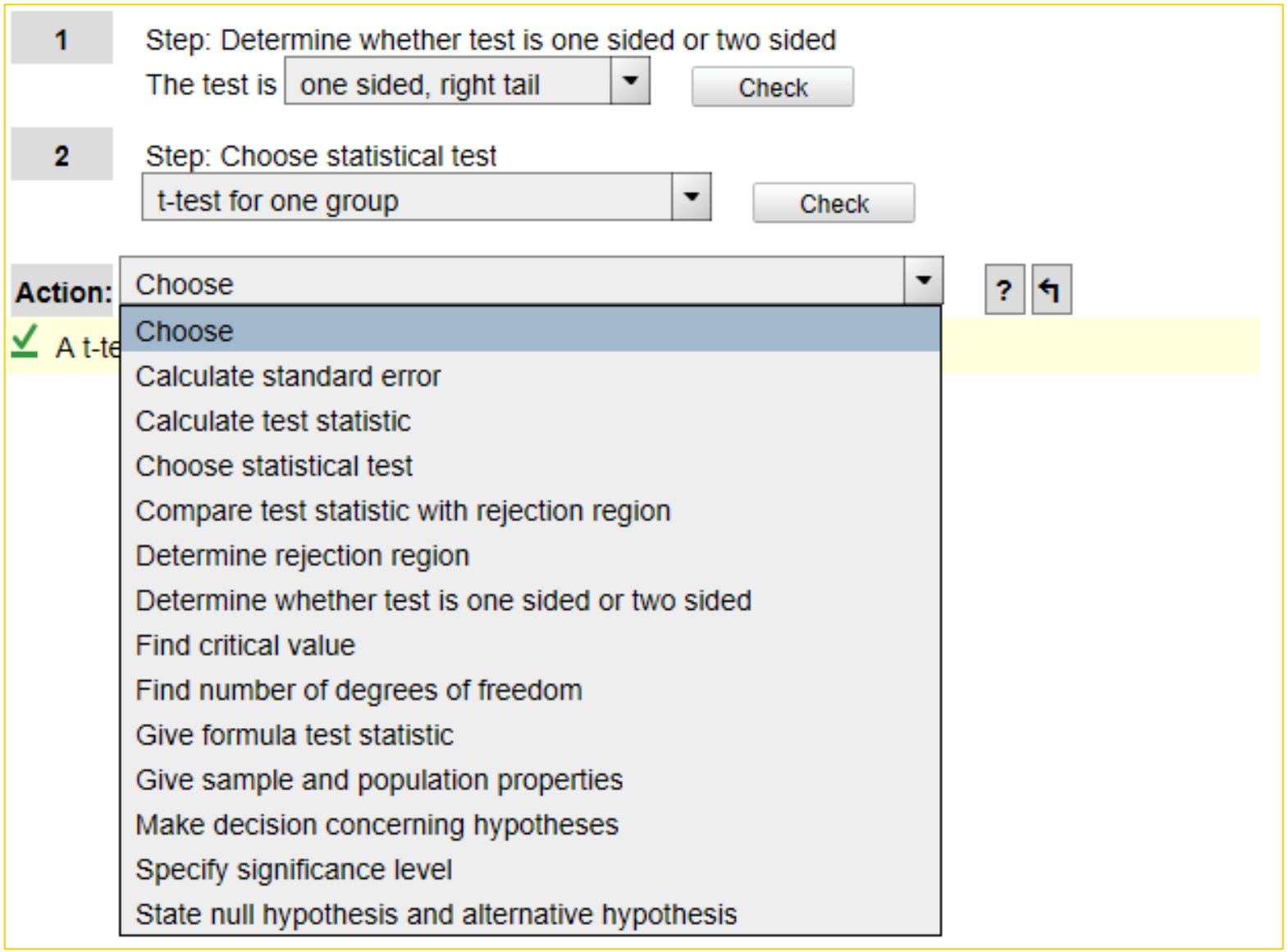
e Under the assumption that the null hypothesis is true, according to the $\sqrt{n}$ -rule the mean of 31 temperatures is normally distributed with:

Mean $\mu =$ [ ] °C Round off to 2 decimals.

Standard deviation $\sigma =$ [ ] °C

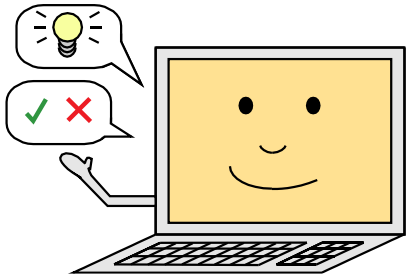


Multi-step task with step selection



Feedback design – example 1

Experimental condition



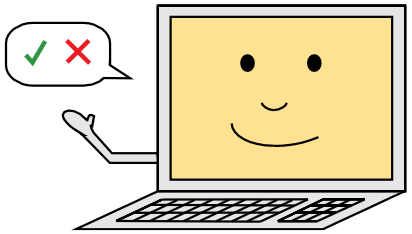
2 Step: Find critical value

z_{crit} = 1.33 Check

Action: Choose ? ↰

✗ What exactly is the critical value and what is the test statistic?

Control condition

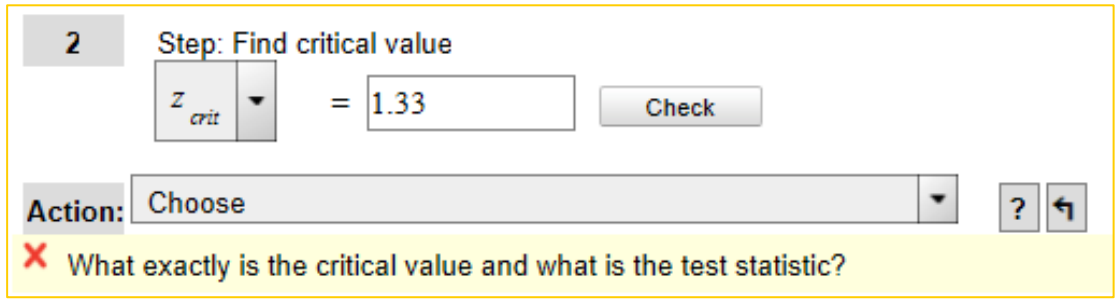


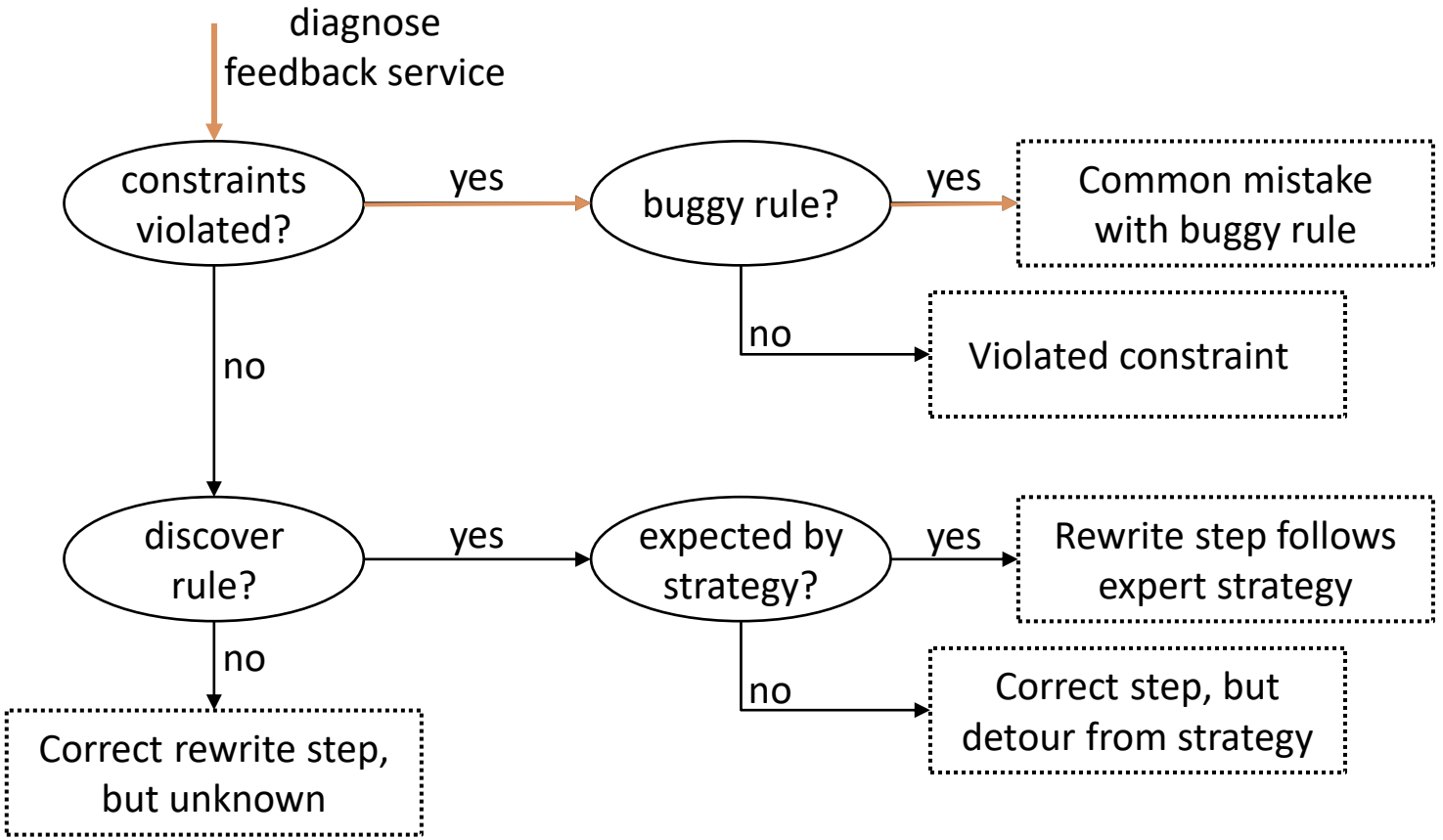
2 Step: Find critical value

z_{crit} ✓ = 1.33 ✗

Action: Choose ↰

Domain reasoner – rules and constraints

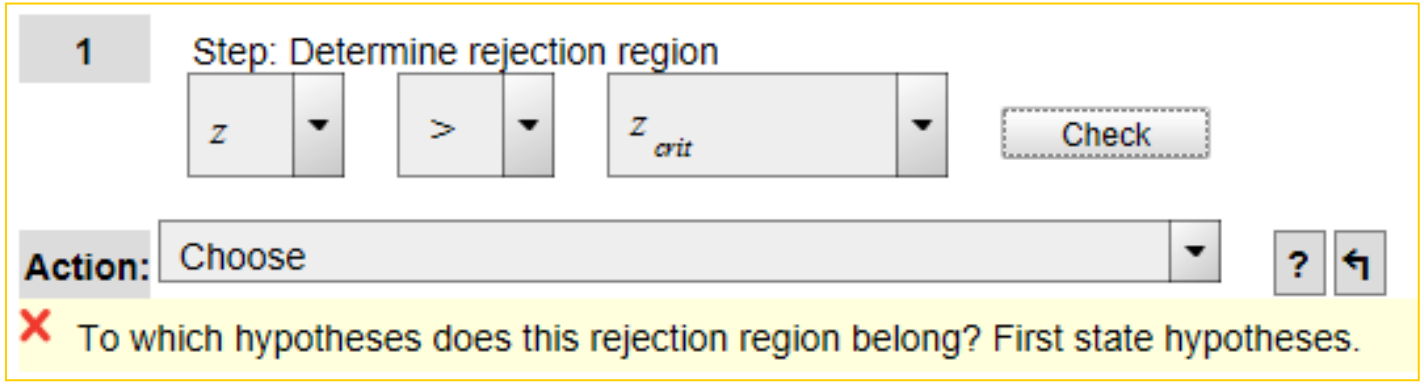




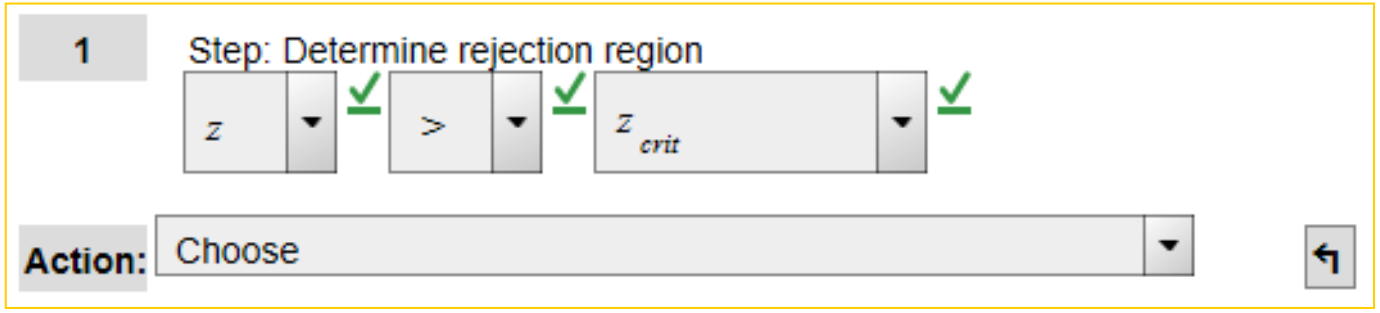
Heeren & Jeuring, 2014

Feedback design – example 2

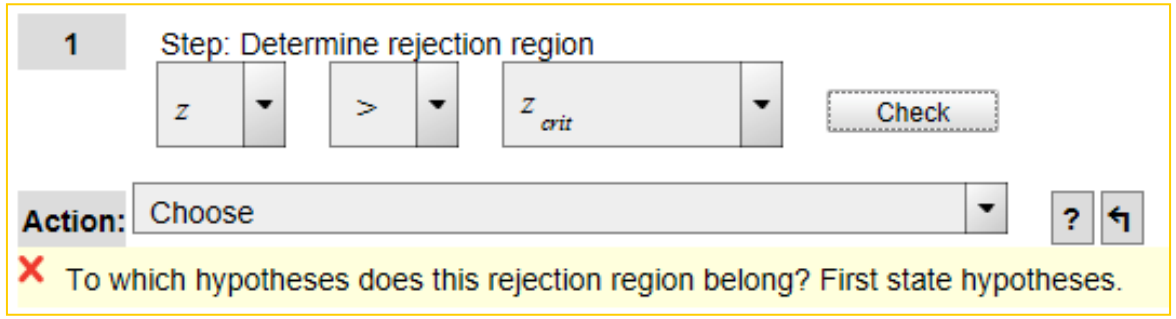
Experimental condition

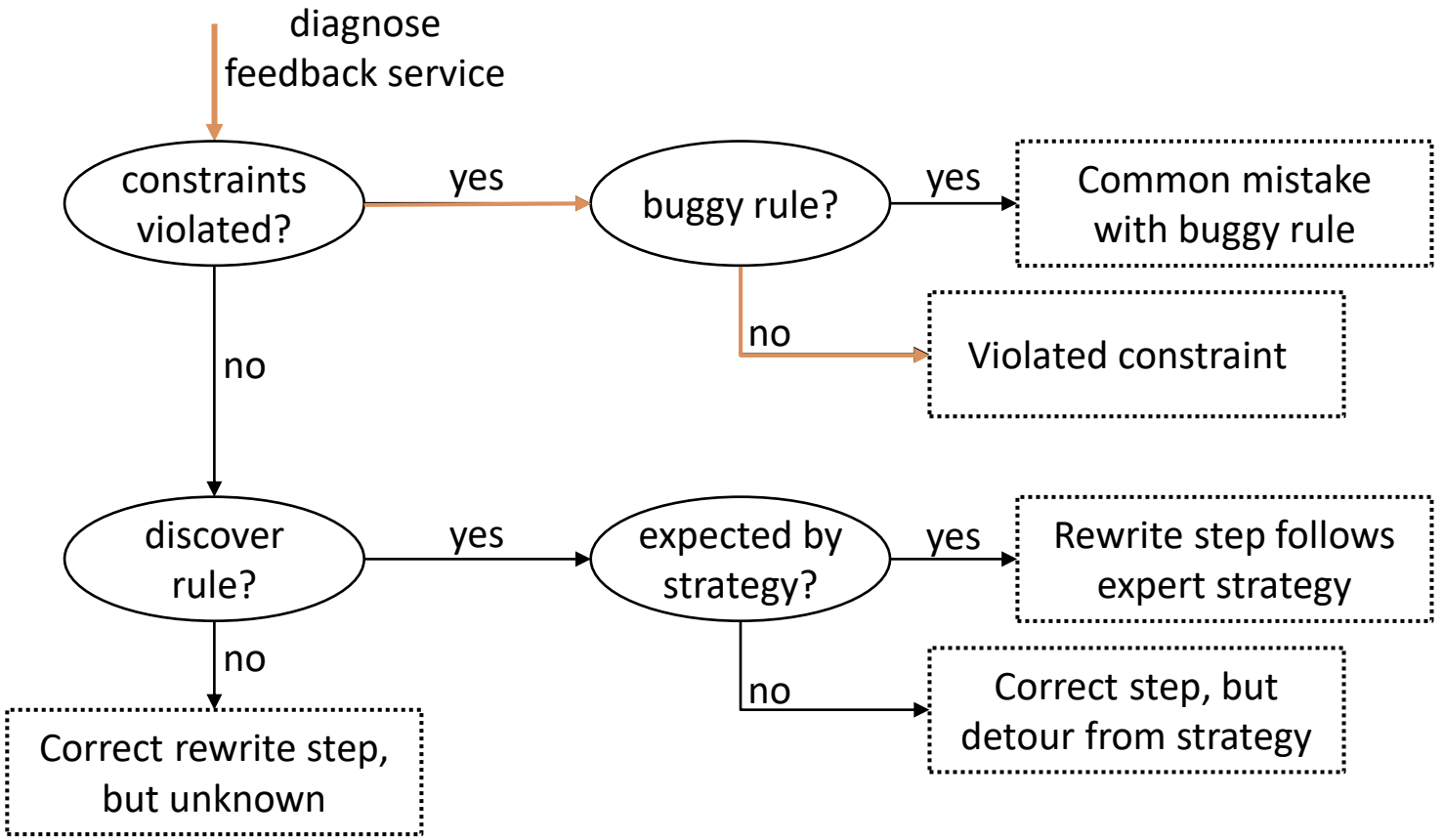


Control condition



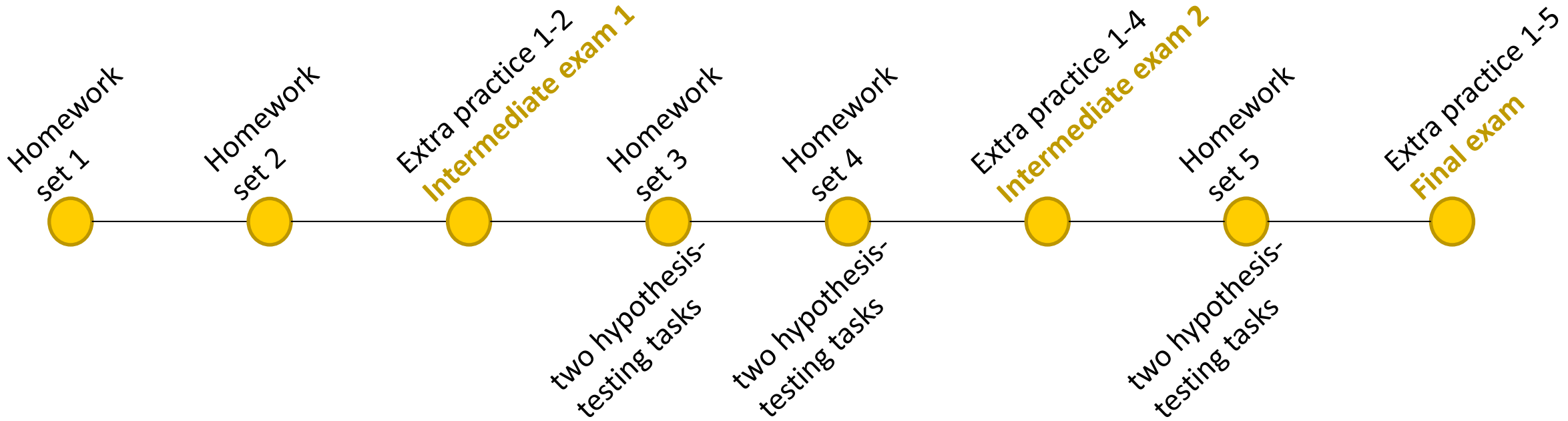
Domain reasoner for second example



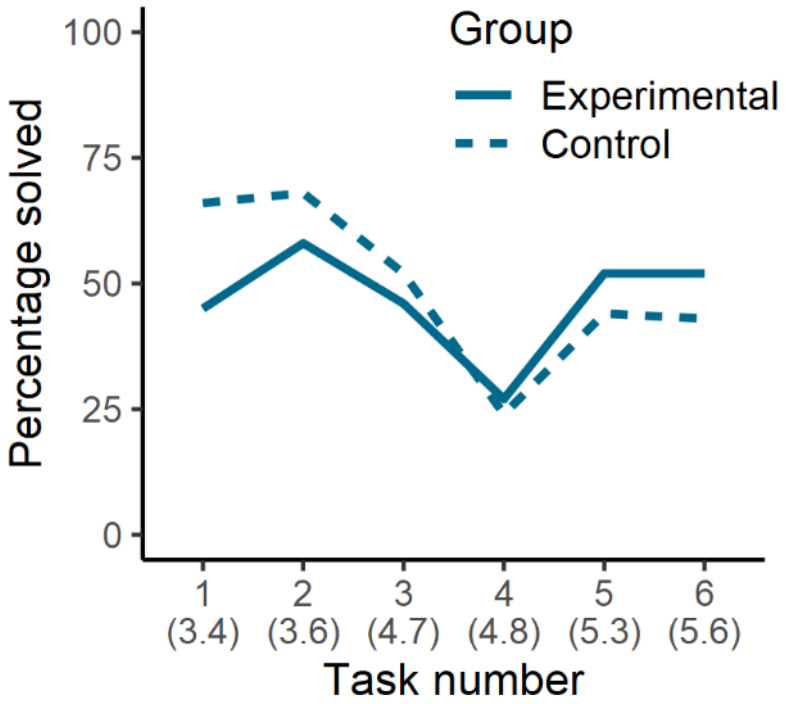


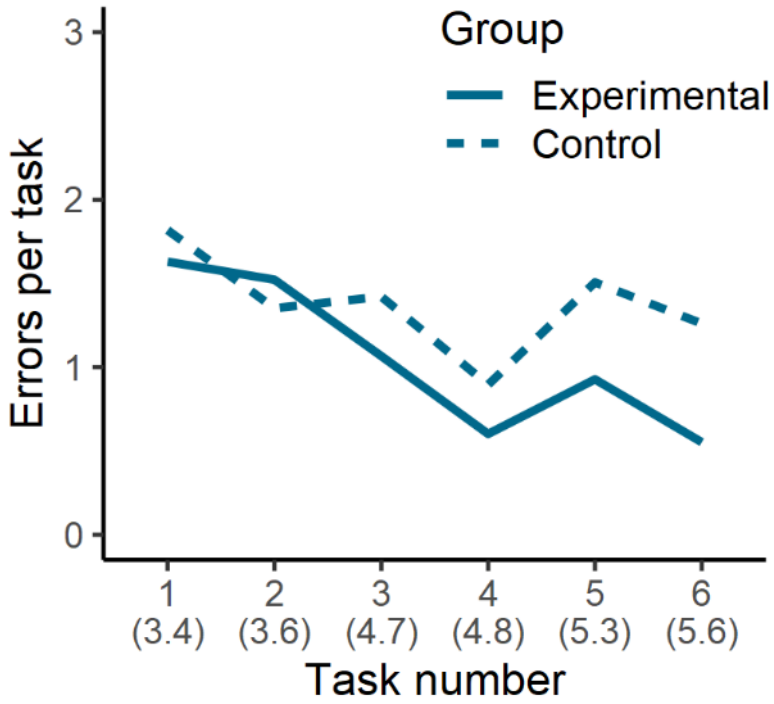
The experiment

- University statistics course for social sciences
- Homework in online learning environment



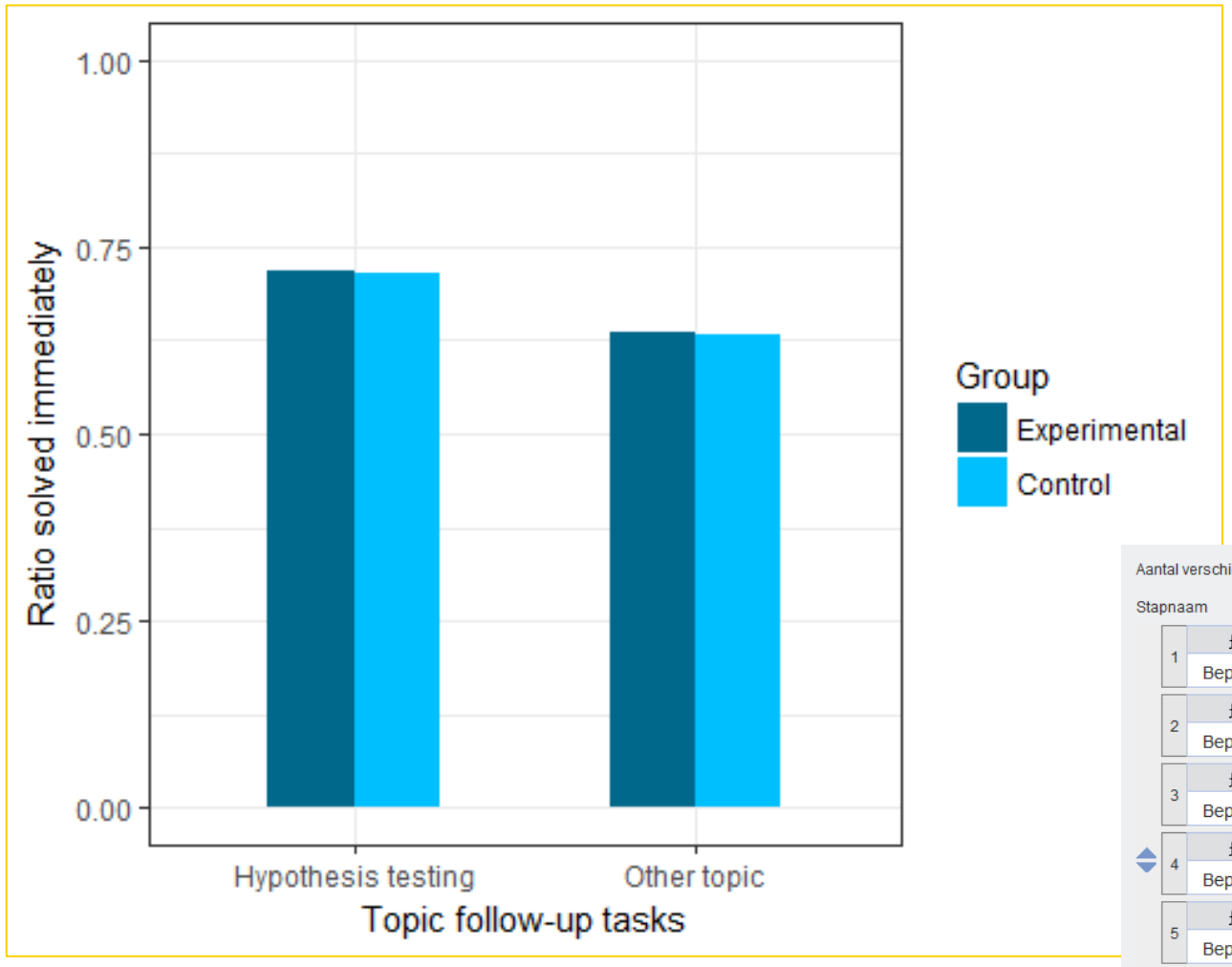
Feedback matters …

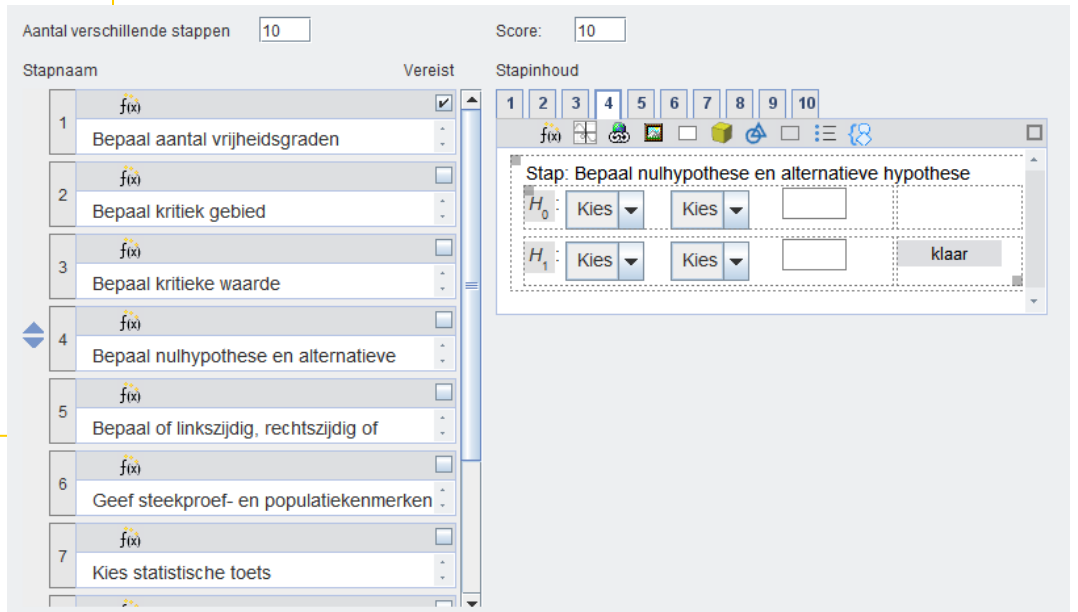




Experimental: 163 students
Control: 151 students

… and steps matter as well





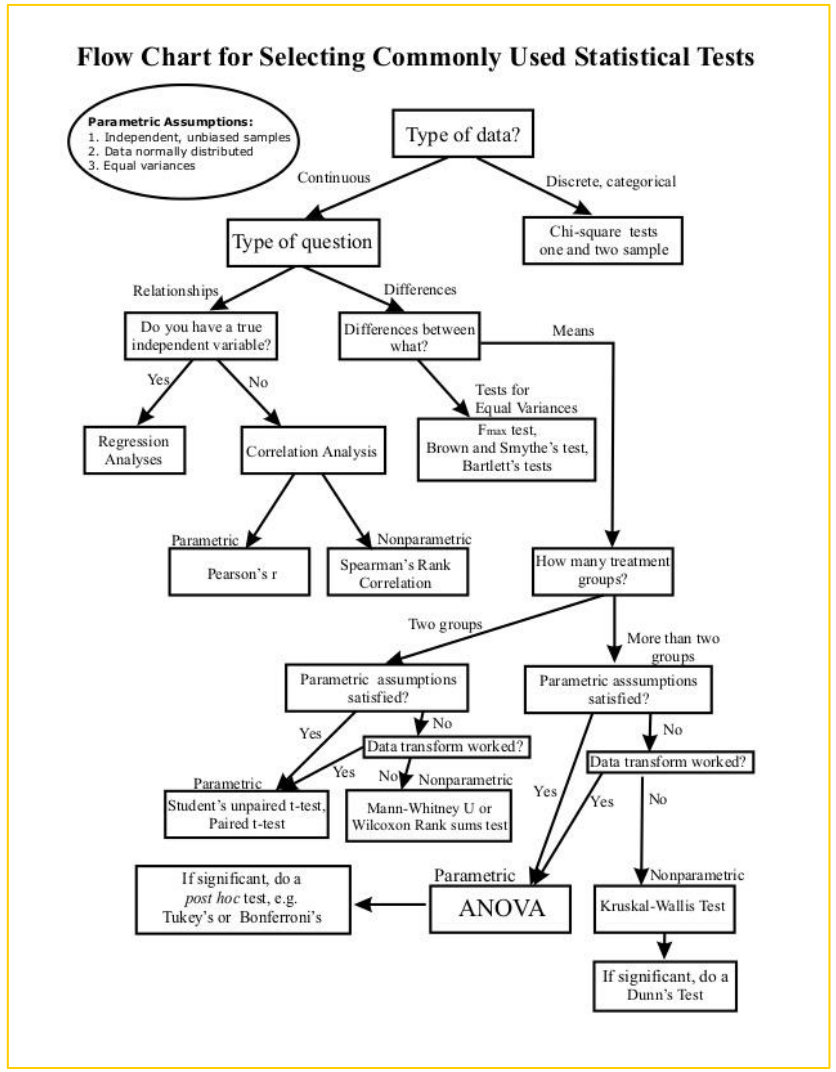
Question 1

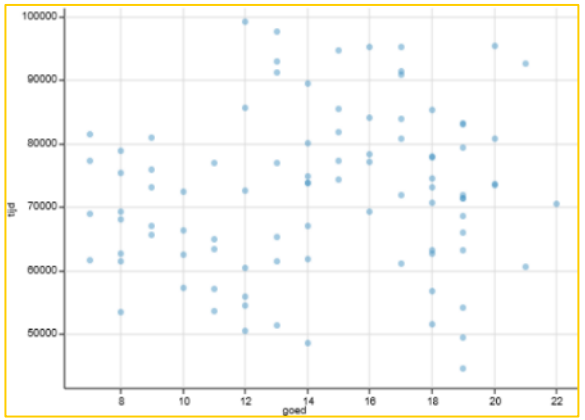
How can computers be used to give useful automated feedback in multi-step hypothesis-testing tasks?

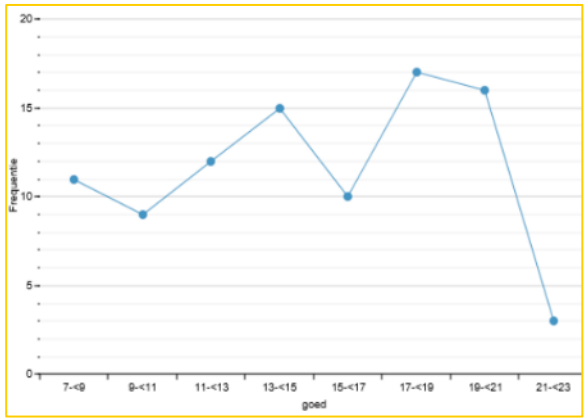
Use a domain reasoner to provide feedback on the logical reasoning and coherence of solutions

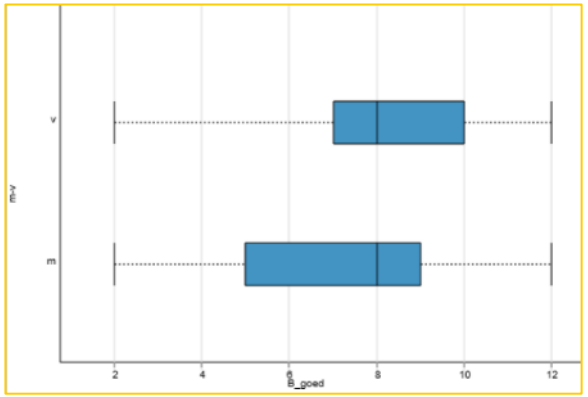
Demo video: https://youtu.be/toXFJhFJI5w

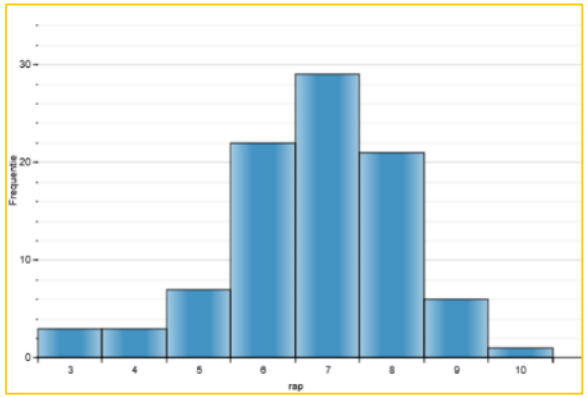
Statistical proficiency









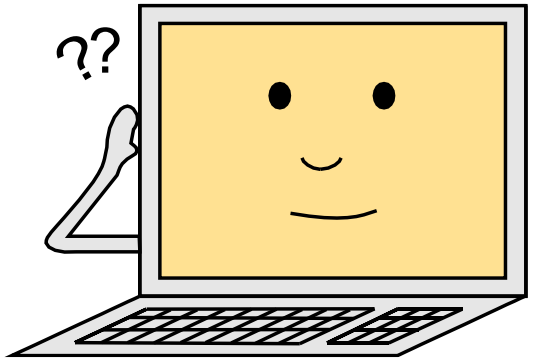


Source: https://www.slideshare.net/1171097100/choosing-appropriate-statistics-test-flow-chart

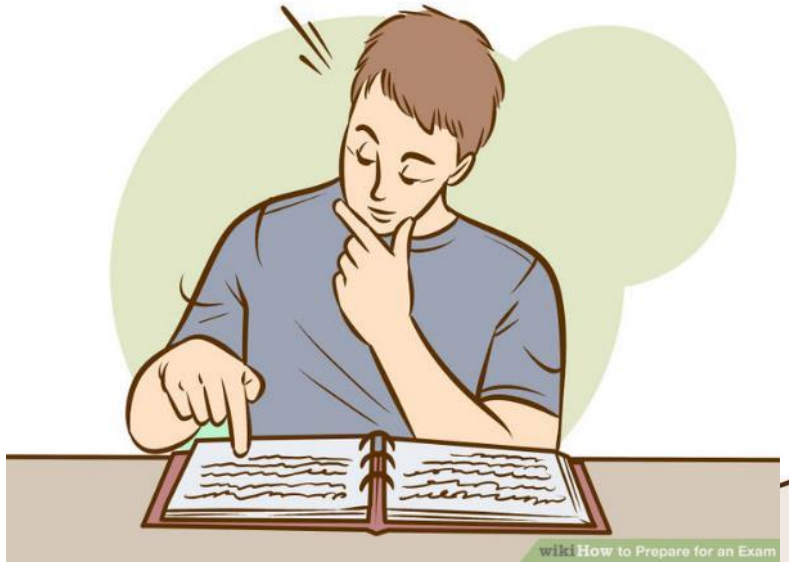
Castro Sotos et al., 2007

Question 2

How can computers be used to provide intelligent feedback about statistical proficiency?



Exam preparation...



LESSON Module 1 - Descriptive statistics

Exercise 7

Chapter 1

The frequency table below is part of the SPSS output that is generated from data collected in a study about coordination by boys and girls. The table displays the time it took the children to throw a ball into a net (variable 'duration' in minutes).

duration

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	1	5,0	5,0	5,0
	2	2	10,0	10,0	15,0
	3	3	15,0	15,0	30,0
	4	4	20,0	20,0	50,0
	5	3	15,0	15,0	65,0
	6	1	5,0	5,0	70,0
	7	5	25,0	25,0	95,0
	8	1	5,0	5,0	100,0
	Total	20	100,0	100,0	

Formulas

a What is the dependent variable?

time until the ball is thrown into the net ✓

Is this variable continuous or discrete?

continuous ✓

Hint

b Which percentage of children needed longer than 6.5 minute to throw the ball into the net?

Hint

[ ] %

c What is the percentile rank of 5.499 minutes?

Hint

[ ] e percentile

d Which figure would be the best graphical representation for the variable 'duration'? Multiple answers are possible

Hint

Histogram | Polygon | Bar chart

and click Check

01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 Partial scores

Variables

Percentile

Representations

Inspectable student model

Category	Score
⊟ **Variables and representations**	81%
Properties variables	81%
Appropriateness representations	75%
Frequency table	100%
Cumulative frequency and percentile	75%
⊟ **Measures of center and spread**	44%
Finding measures of center	17%
Appropriateness measures of center	50%
Range	67%
Standard deviation	63%

Category	Score
⊞ **Variables and representations**	81%
⊞ **Measures of center and spread**	44%

Bull & Kay, 2010

LESSON Module 1 - Descriptive statistics

Exercise 7

The frequency table below is part of the SPSS output that is generated from data collected in a study about coordination by boys and girls. The table displays the time it took the children to throw a ball into a net (variable 'duration' in minutes).

duration

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	1	5,0	5,0	5,0
	2	2	10,0	10,0	15,0
	3	3	15,0	15,0	30,0
	4	4	20,0	20,0	50,0
	5	3	15,0	15,0	65,0
	6	1	5,0	5,0	70,0
	7	5	25,0	25,0	95,0
	8	1	5,0	5,0	100,0
	Total	20	100,0	100,0	

Formulas

a What is the dependent variable?

time until the ball is thrown into the net ✓

Is this variable continuous or discrete?

continuous ✓

Hint

b Which percentage of children needed longer than 6.5 minute to throw the ball into the net?

Hint

[ ] %

c What is the percentile rank of 5.499 minutes?

Hint

[ ] e percentile

d Which figure would be the best graphical representation for the variable 'duration'? Multiple answers are possible

Hint

Histogram | Polygon | Bar chart

and click Check

01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 Partial scores

Chapter 1

Dependent/ Variables

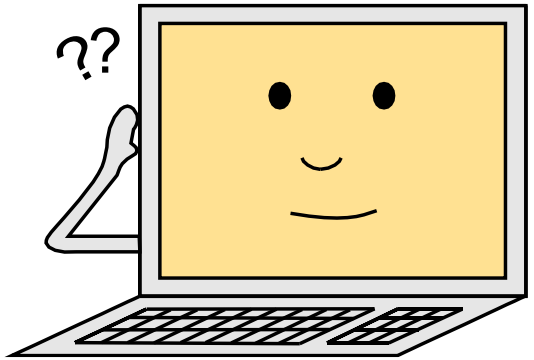
Continuous/ discrete

Percentile

Representations

Question 3

Do the student models that we have designed make sense?



Student data

Student	Task	Attempt
Student a	07a1	✗
Student a	07a1	✓
Student a	07a2	✓
Student a	07b	✗
Student a	07b	✓
Student a	07c	✓
Student a	07d	✓
Student b	07a1	✓
Student b	07a2	✗
Student b	07a2	✓
Student b	07b	✓
…	…	…

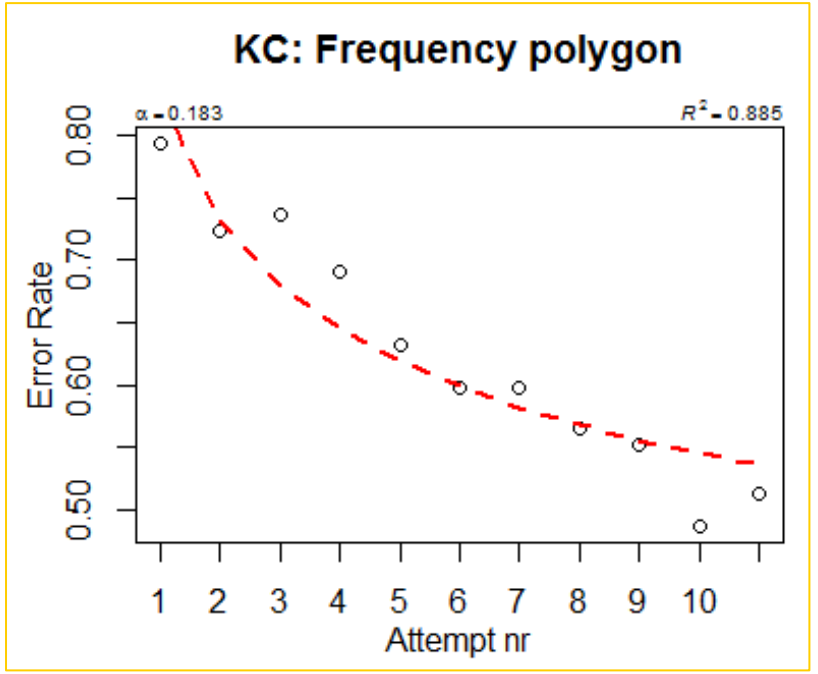
"Knowledge Component"

Student	KC	Attempt
Student a	Percentile	✗
Student a	Percentile	✓
Student a	Percentile	✓
Student b	Percentile	✓
Student a	Variables	✗
Student a	Variables	✓
Student a	Variables	✓
Student b	Variables	✓
Student b	Variables	✗
Student b	Variables	✓
Student a	Representations	✓
…	…	…

Learning curve analysis

- $Error\ rate\ n'th\ attempt = \frac{number\ of\ incorrect\ n'th\ attempts\ on\ KC}{total\ number\ of\ n'th\ attempts\ on\ KC}$

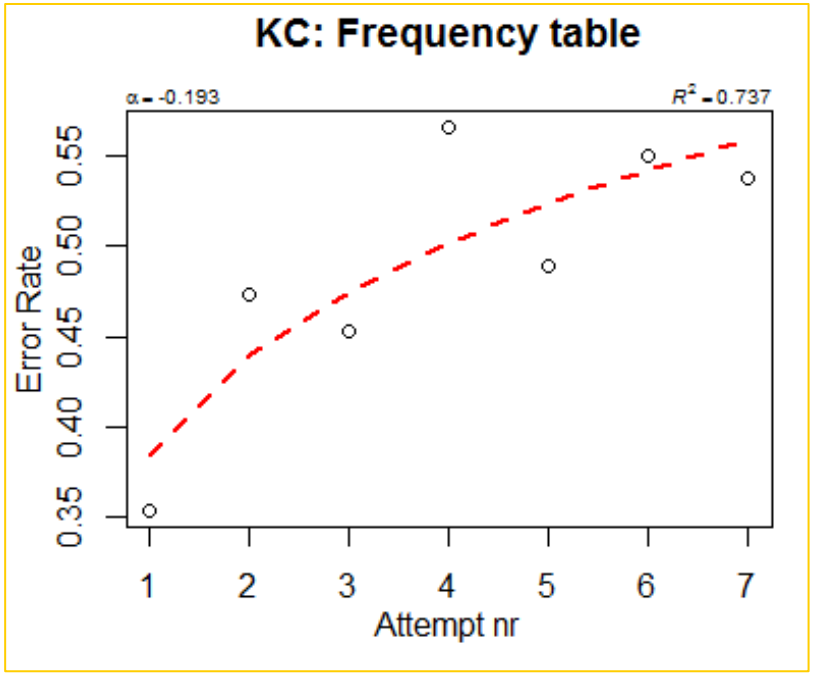
- Fitting function: $Error\ Rate = B \cdot AttemptNo^{-\alpha}$

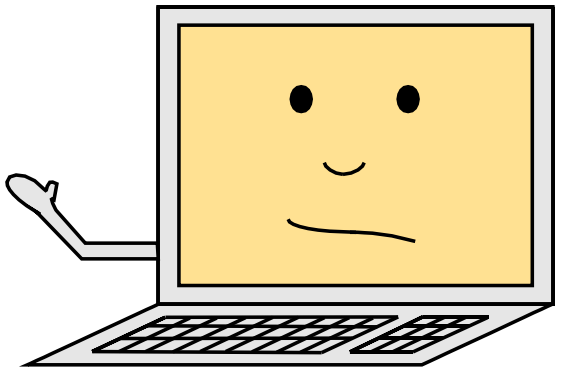


Martin, Mitrovic, Koedinger, & Mathan, 2011

Our learning curve analysis

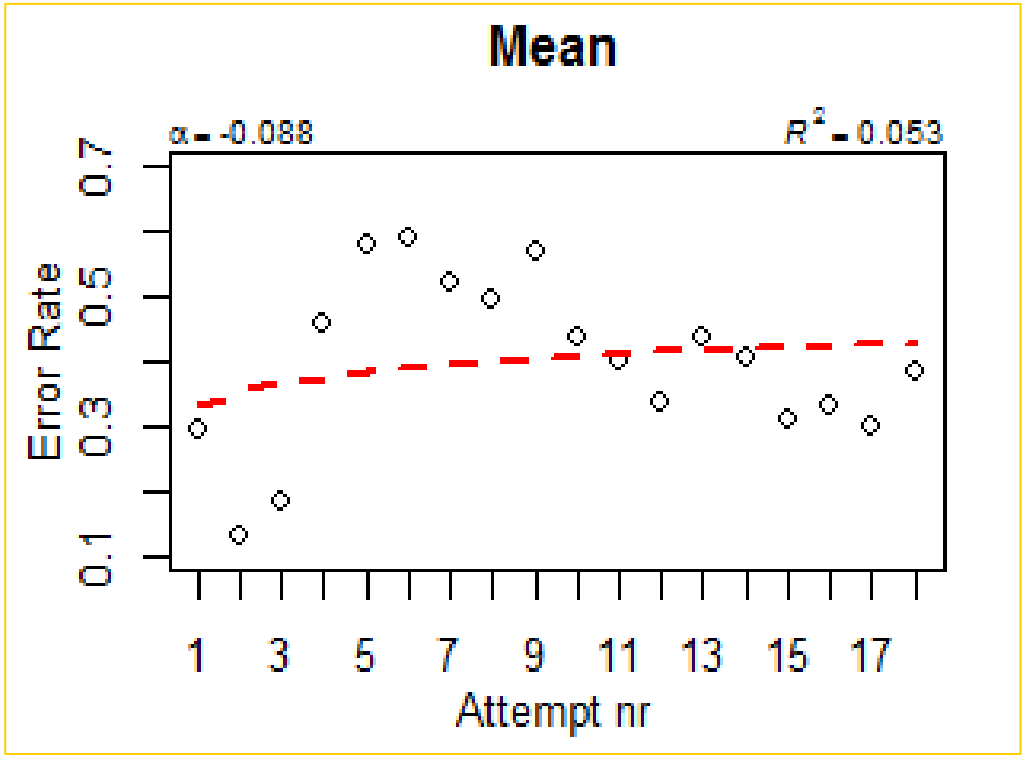
- Five homework sets
- 160 students
- 71 KCs
  - 34 KCs: decreasing learning curve
  - 15 KCs: not enough information
  - 22 KCs: increasing learning curve





Tacoma, Sosnovsky, Boon, Jeuring, & Drijvers, 2018

Example 1


Mean
α = -0.088
R^2 = 0.053
Error Rate
0.1
0.3
0.5
0.7
1
3
5
7
9
11
13
15
17
Attempt nr

Some tasks about the mean

Task 9

Given are the following scores:
8 7 8 8 4 9 10 7 8 8 9 8

c. What is the mean of this distribution?

Task 11

Which measures of center are appropriate for the following situations?

a. The number of bedrooms that houses in a specific street have.

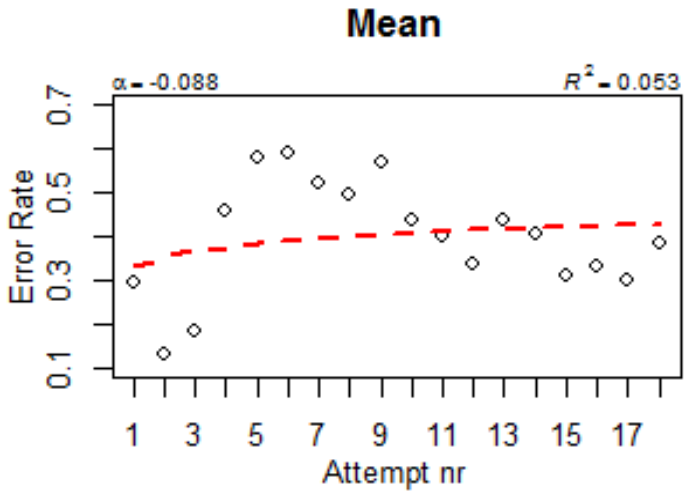
Mean | Median | Mode

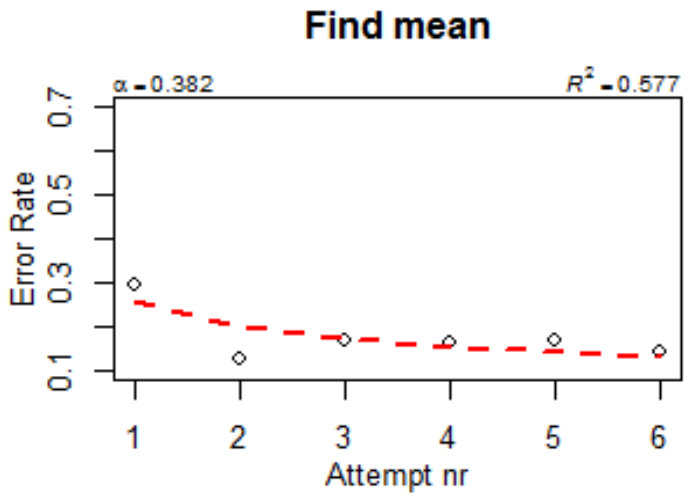
Task 14a

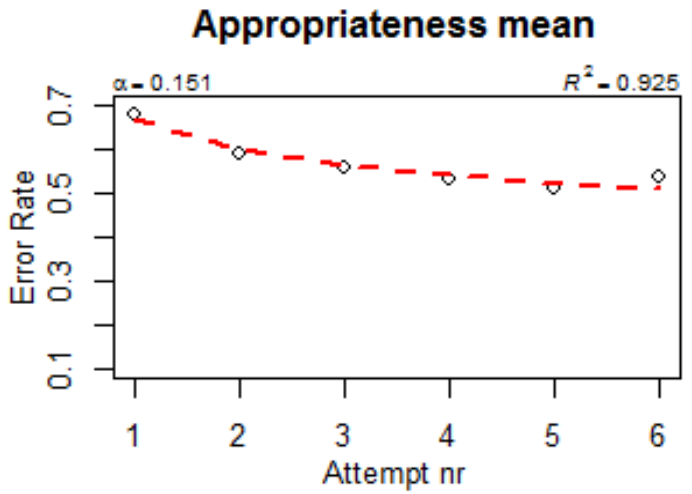
Calculate the standard deviation of the following sample of $n = 5$ scores: 0 6 7 8 14

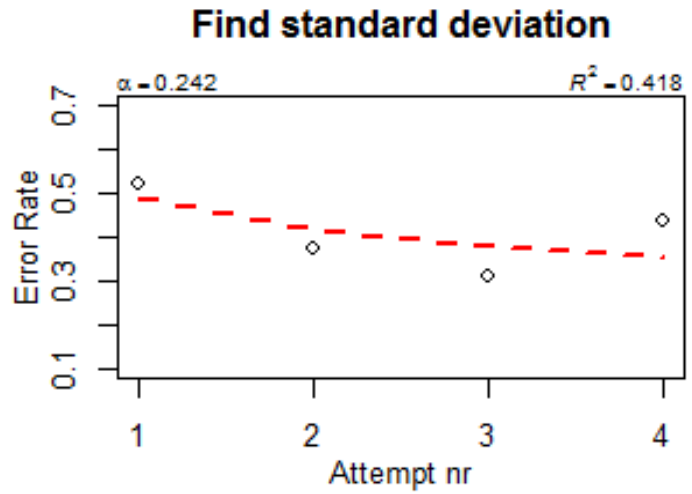
Standard deviation:

Multiple perspectives and concepts

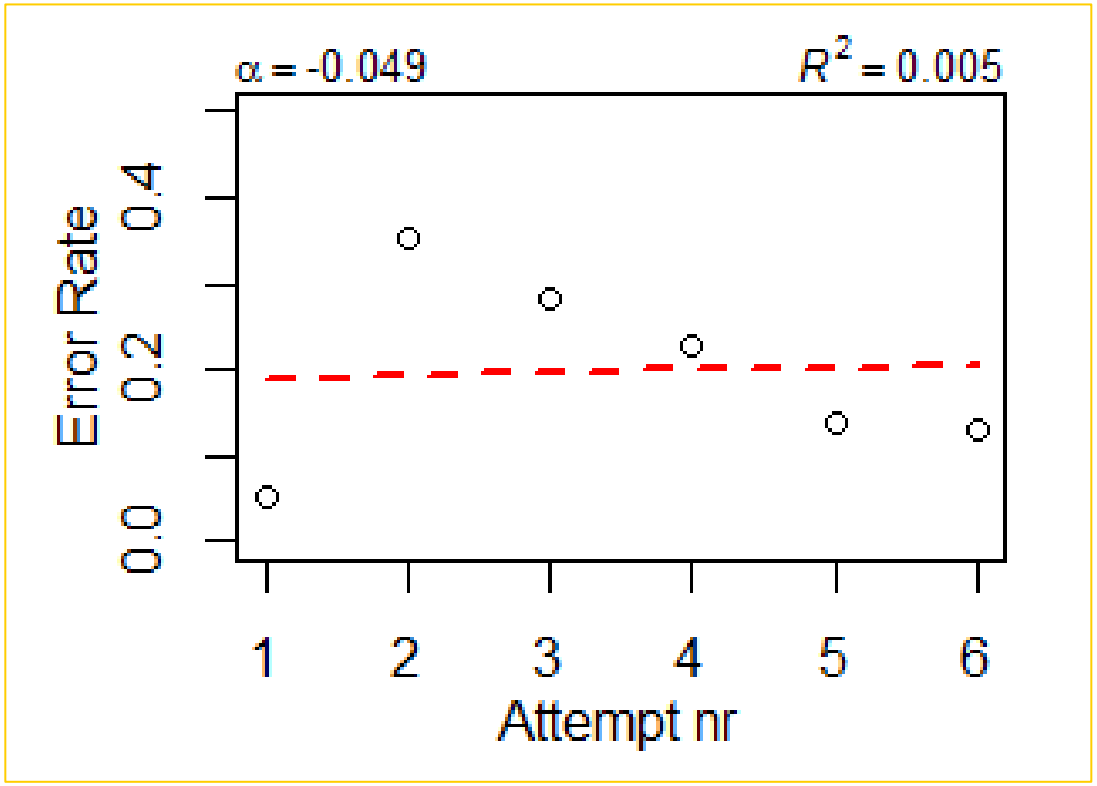








Example 2



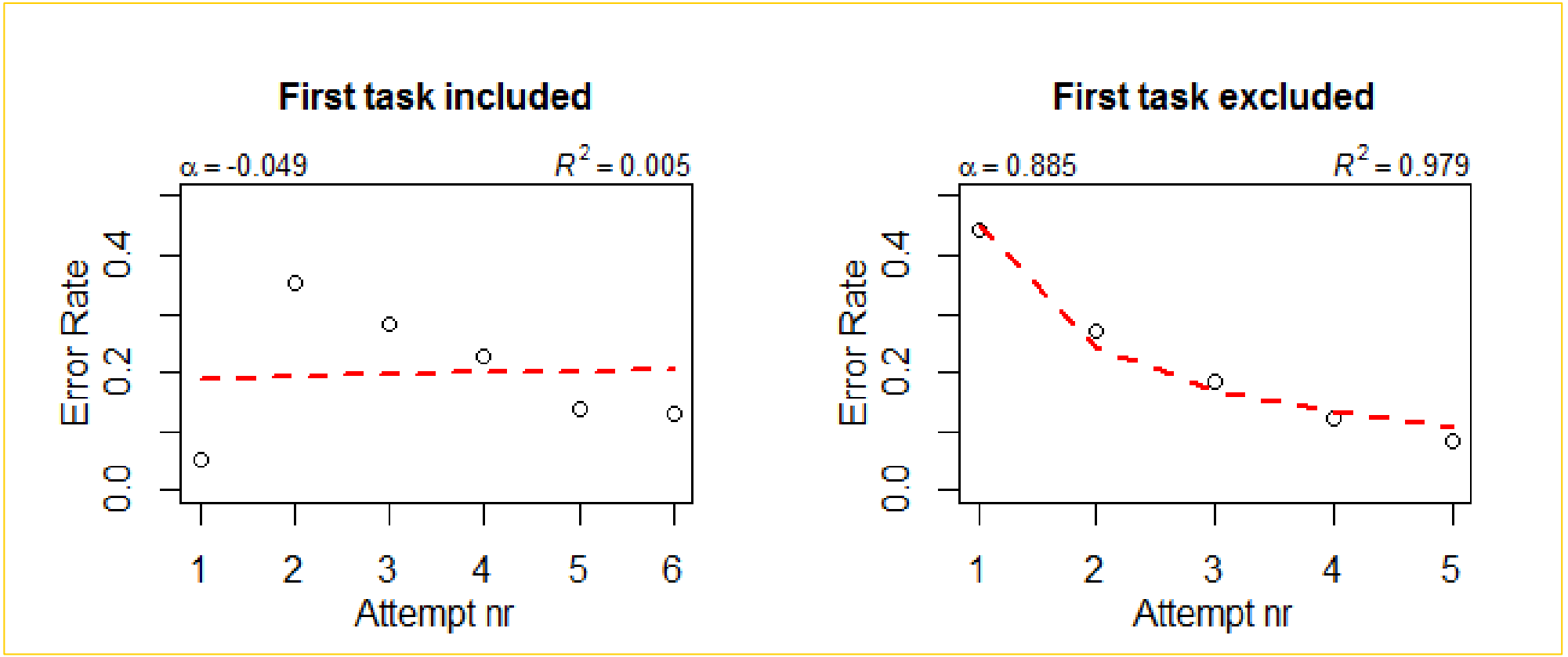
Knowledge component: Significance level

Some tasks about the significance level

Shown in the Digital Mathematics Environment (DME)

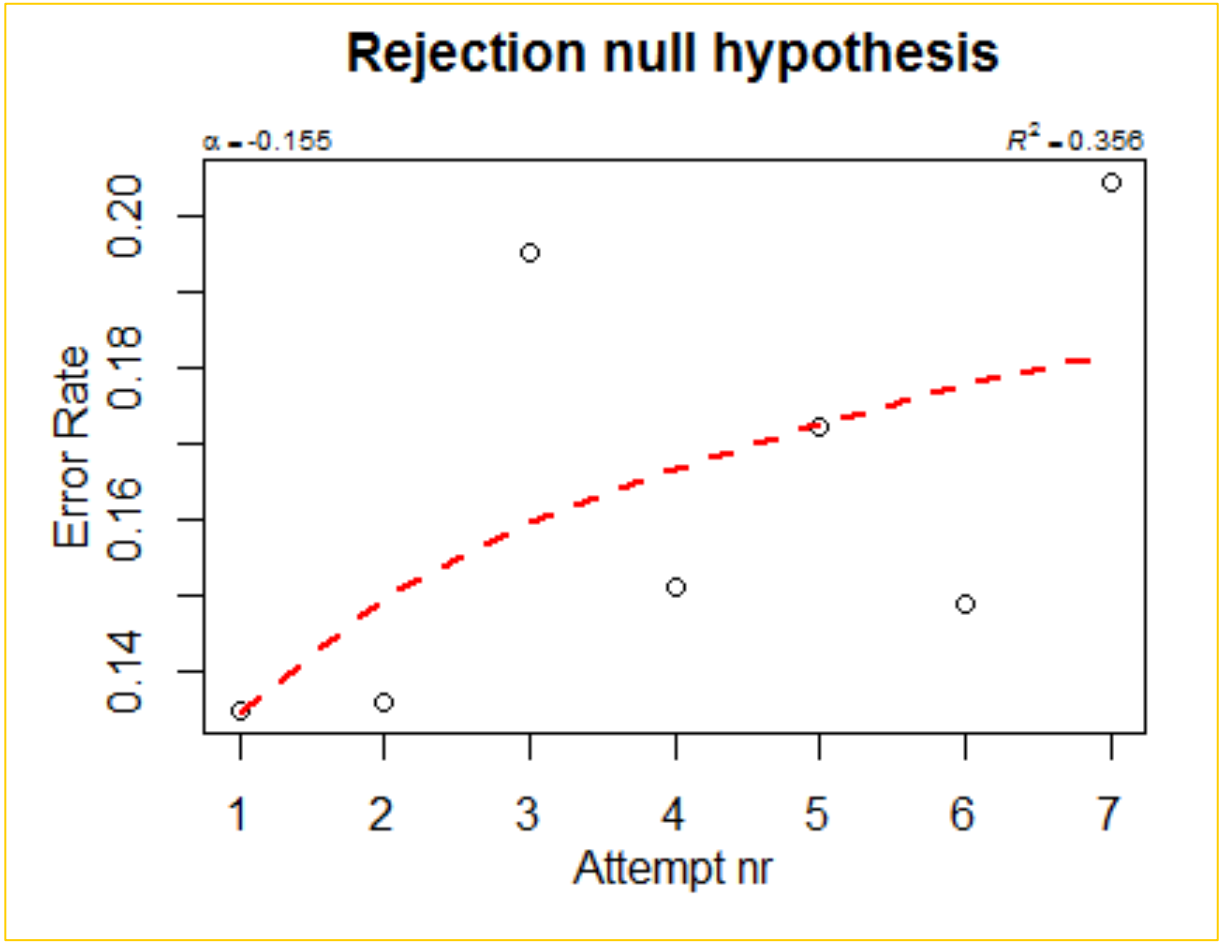
www.dwo.nl/leerling

Easy first tasks



Knowledge component: Significance level

Example 3


Rejection null hypothesis
α = -0.155
R^2 = 0.356
Error Rate
0.14
0.16
0.18
0.20
1
2
3
4
5
6
7
Attempt nr

Example task

Task 1

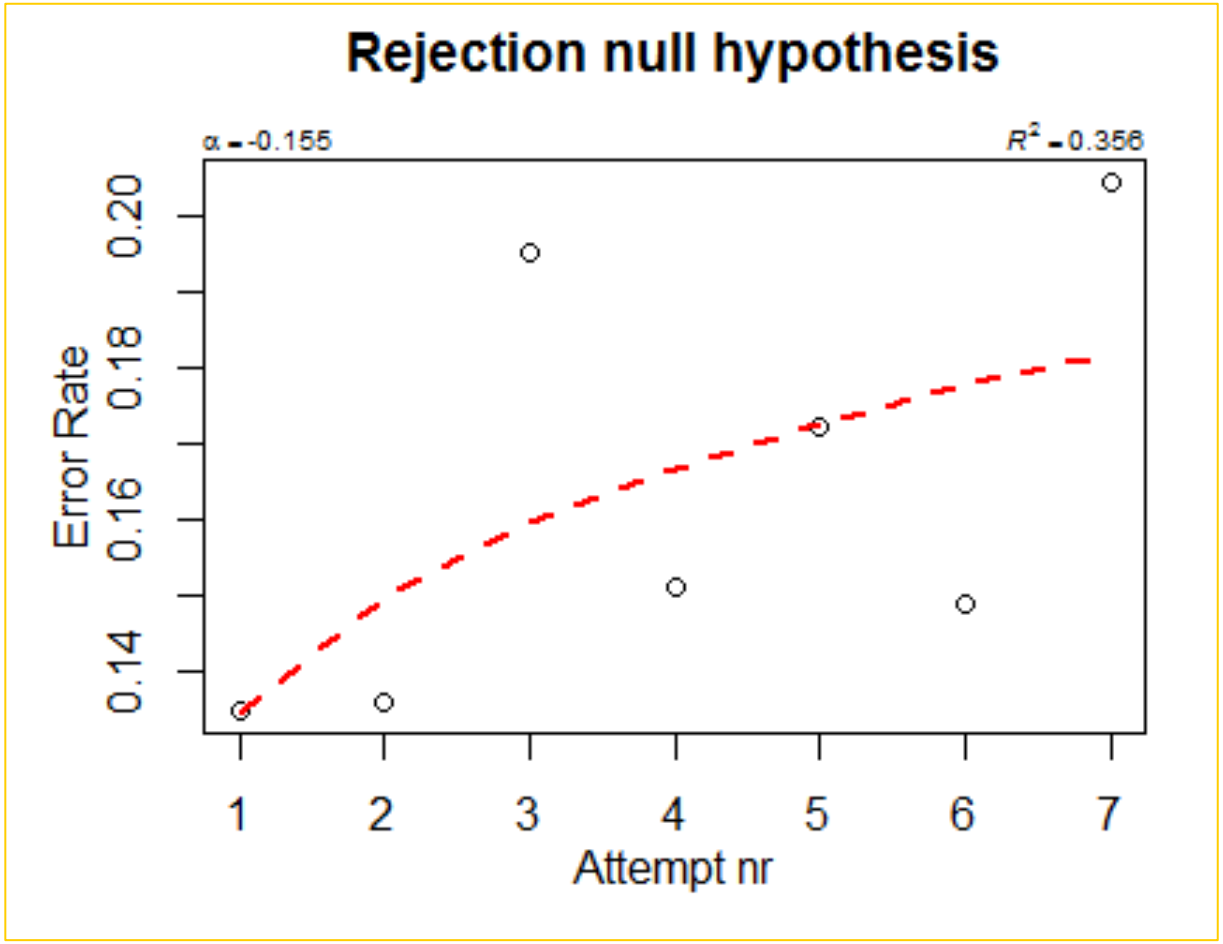
Context about learning ability of under achievers and corresponding data.

d. Does the mean learning ability score of the 25 under achievers differ significantly from 5?

Choose since the value lies Choose the rejection region.

Choose
yes
no

All tasks too easy



Question 3

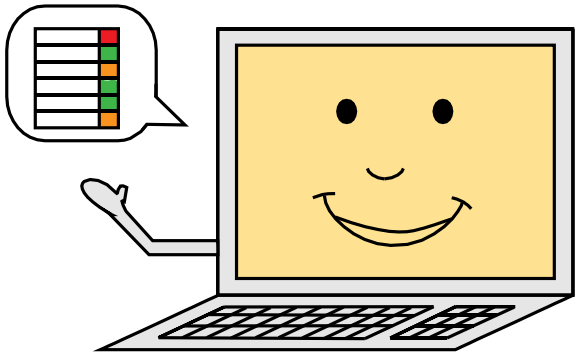
Do the student models that we have designed make sense?

Yes, after redesign of student models and tasks

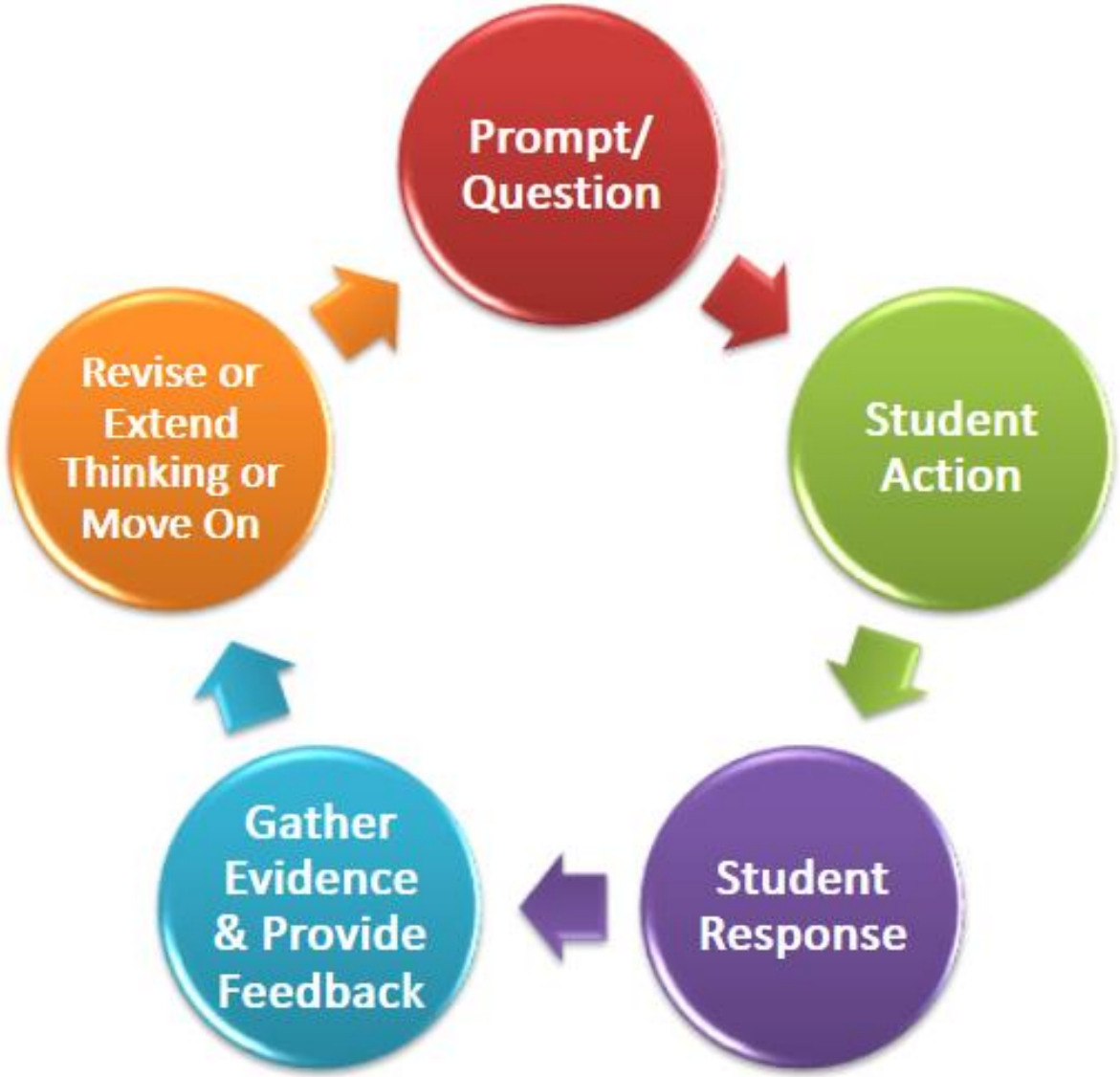
Question 2

How can computers be used to provide intelligent feedback about statistical proficiency?

By providing inspectable student models



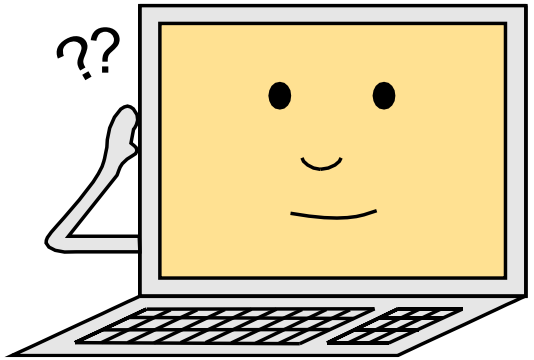
Formative assessment cycle



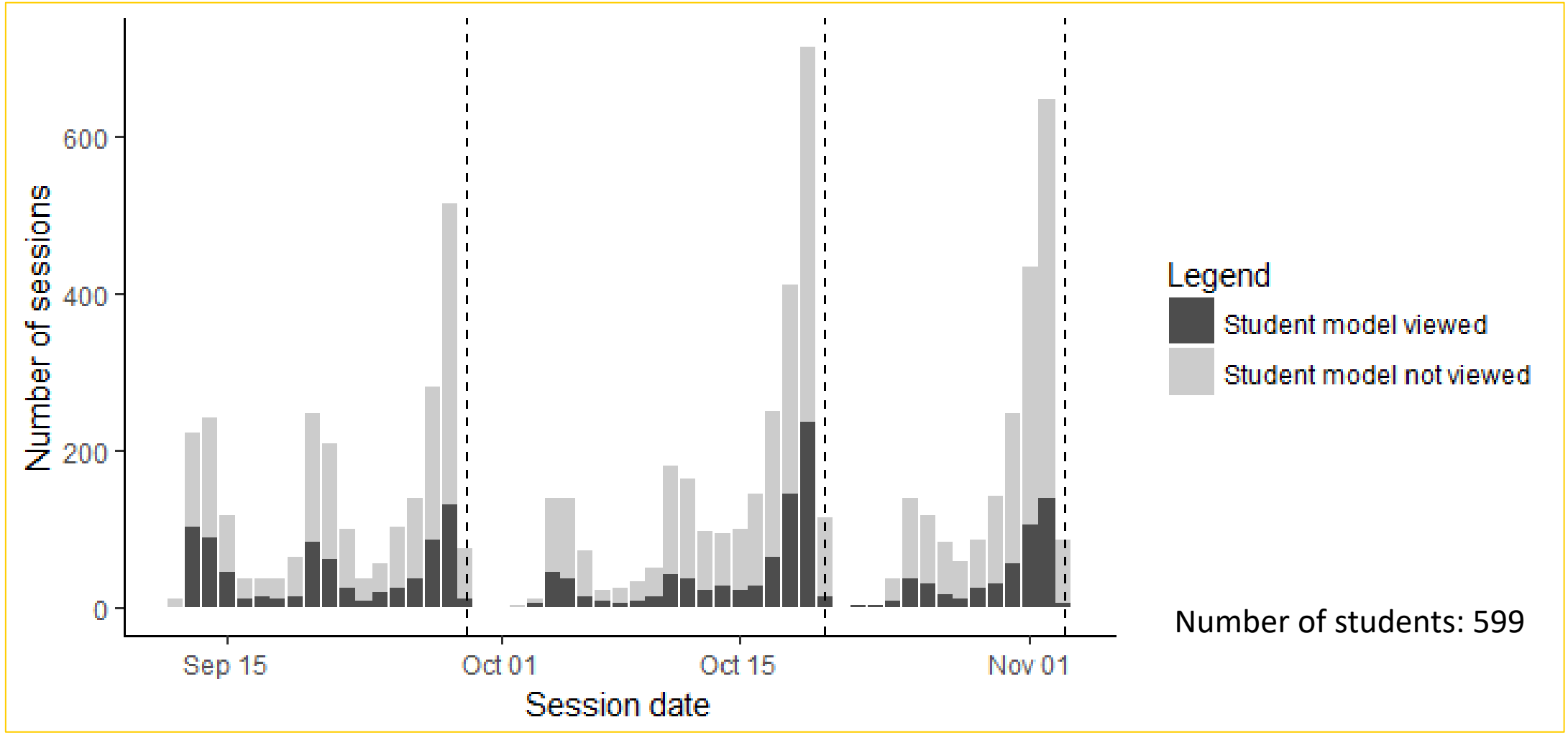
Source: https://slideplayer.com/slide/5907347/

Question 4

Do students benefit from the availability of the domain reasoner and inspectable student models?

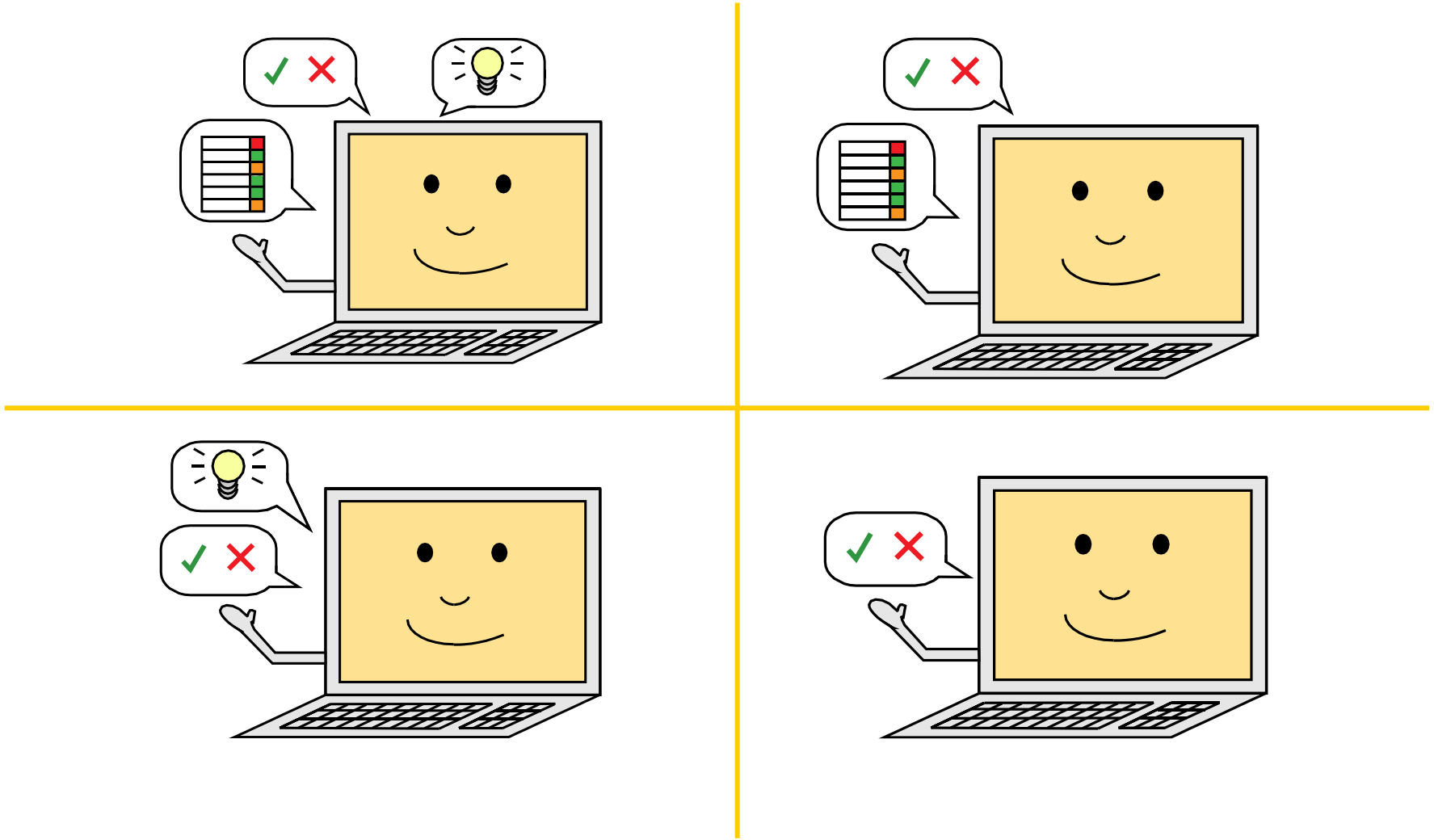


Students' use of inspectable student models



Tacoma, Geurts, Slof, Jeuring, & Drijvers (2020)

Combining inner and outer loop feedback

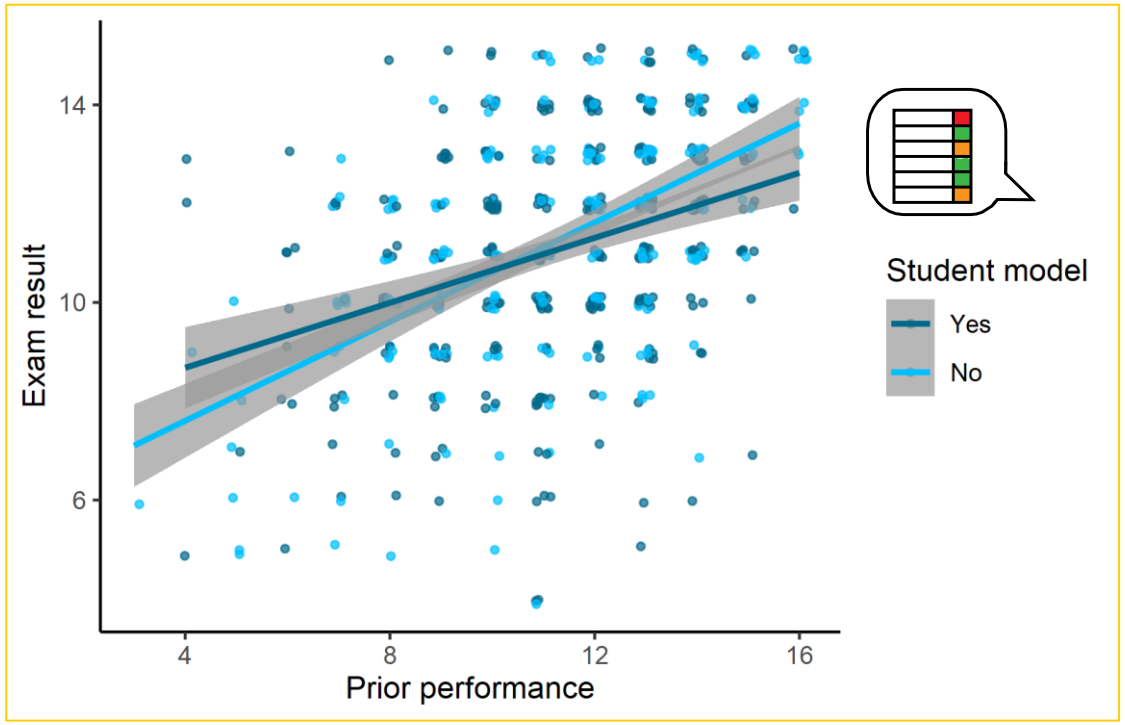


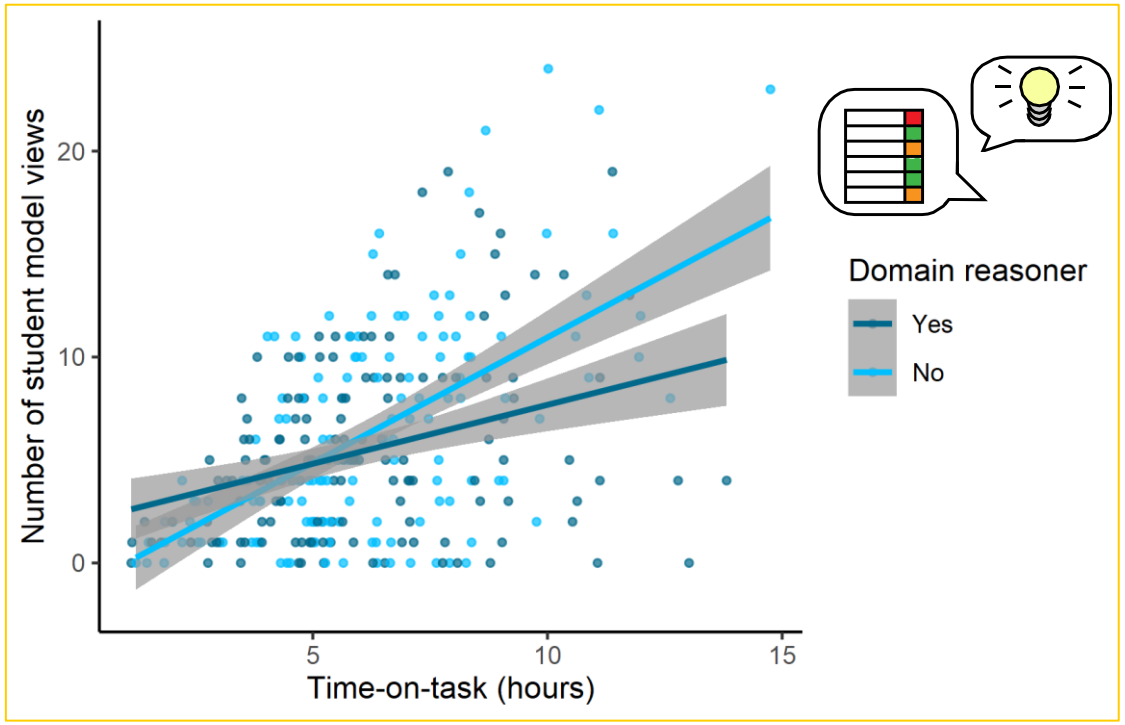
Number of students: 521

Combining inner and outer loop feedback

Results varied widely between students

Inner loop feedback  : Students with prior experience benefited



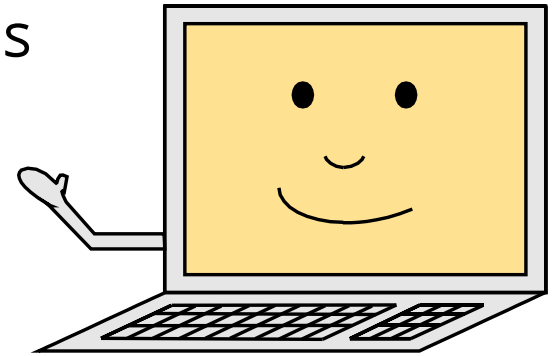


Tacoma, Drijvers, & Jeuring (submitted)

Question 4

Do students benefit from the availability of the domain reasoner and inspectable student models?

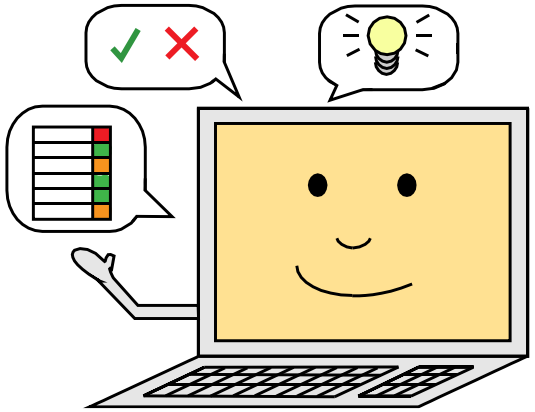
- Students use the feedback
- Feedback effects vary widely
- Weaker students benefit from student models
- Prior experience helps for domain reasoner



Question 2, second try

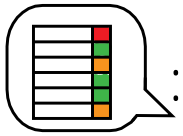
How can computers be used to provide intelligent feedback about statistical proficiency?

- Domain reasoner feedback and inspectable student models
- Give students time to find out what is useful
- Use student data to improve feedback



What's next?

- Inner loop feedback  : domain reasoner and steps ready to use

- Outer loop feedback  :

    - Implementation in Numworx for secondary education
    - Future: offer student model analyses in software

- Sietske  :

    - If all goes well: defense on June 24, 14:30h
    - HU University of Applied Sciences Utrecht
        -Data Driven Business
        -Human Resources Management

Thank you