How to assess higher order thinking skills: theory and practice

Caroline Jongkamp
Nico Dieteren
January 2021

Topics

1. Higher Order Thinking Skills
2. Assessment in contexts
3. Ways to categorise e-assessment item types
   - by type
   - by assessment objective
4. Q&A

HOTS and eAssessment: how we started

- 2015 Innove Estonia Workshop eAssessment
- 2016 IAEA Cape Town: paper presentation
- 2017 IAEA Batumi: pre-conference workshop
- 2018 AEA-E Arnhem-Nijmegen: pre-conference workshop*
- 2019 IB workshop for examiners*
- 2019 AEA-E Lisbon: pre-conference workshop*
- *2020 IAEA Accra: paper presentation: planned**

* Caroline with Rebecca Hamer (IBO)

Higher Order Thinkings Skills (HOTS)

- Higher order thinking skills in high stakes testing: how to develop valid and reliable assessment instruments in 21st century contexts?

- Examples from

  Final exams the Netherlands (source: CvTE & Cito), from

  PISA (source: OECD) and exams from IB

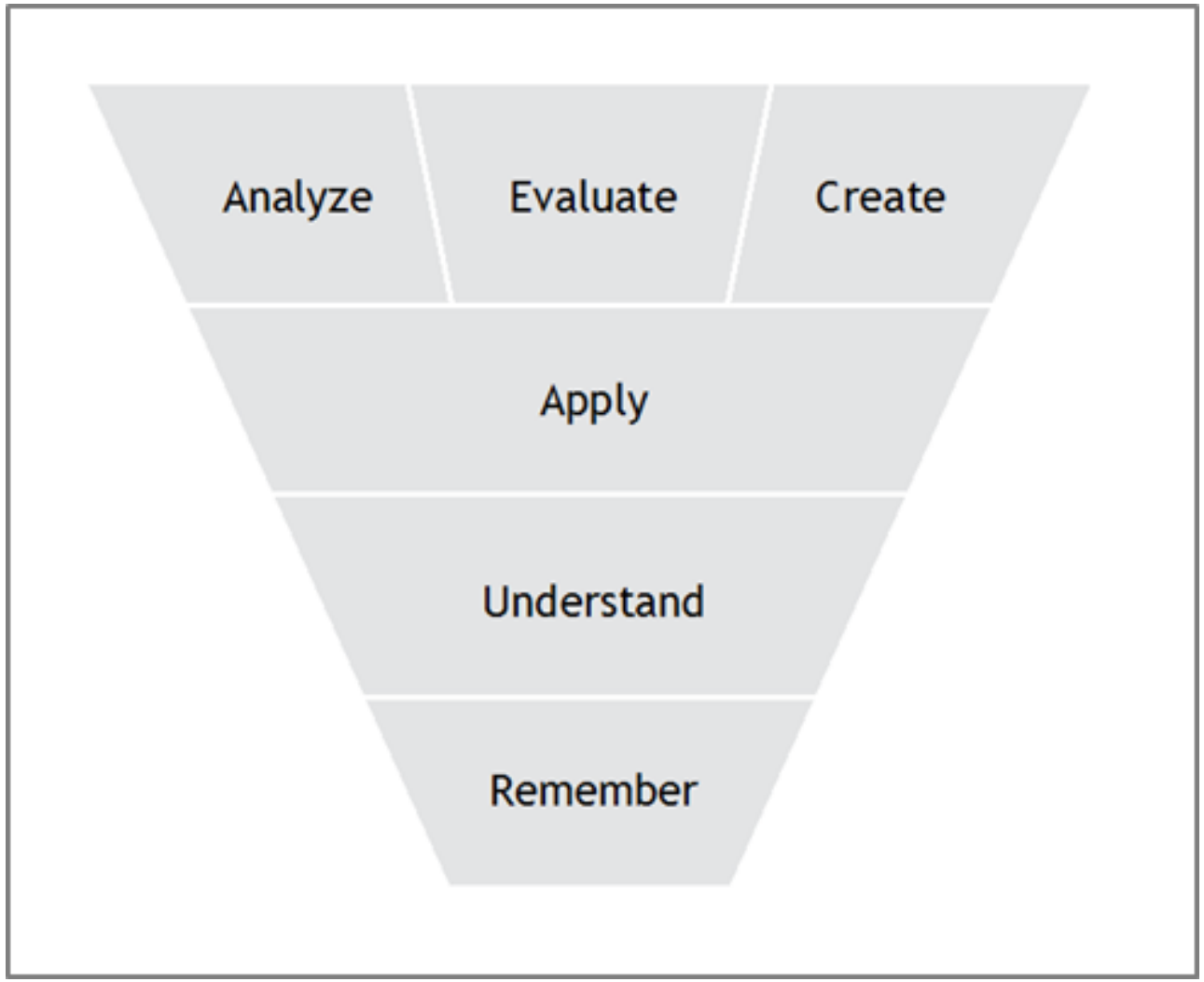
Shifting goals of education

- “The meaning of knowing has shifted from being able to remember and reproduce information, to being able to find, select, judge and use information”

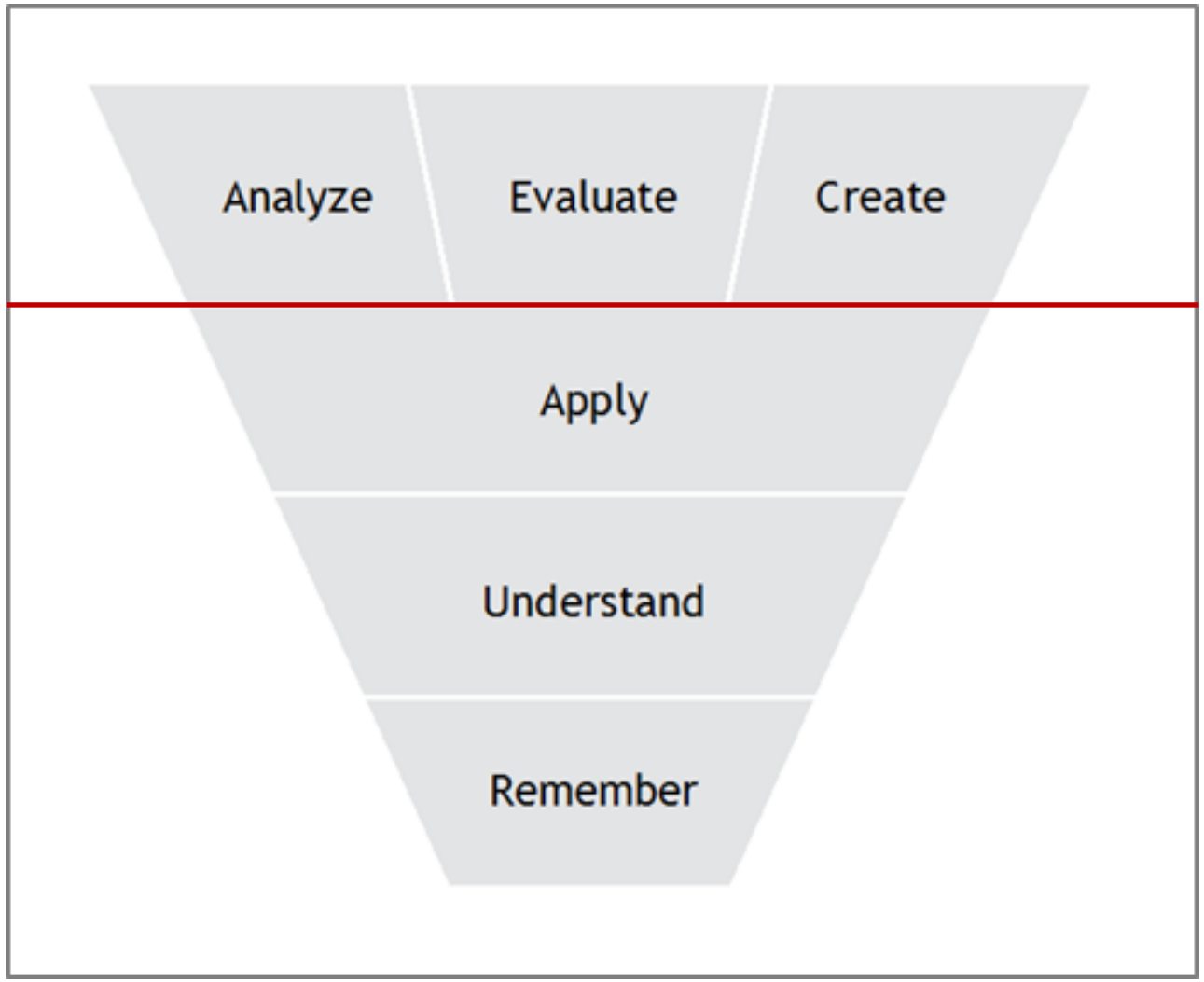
  (free after Herbert Simon)

- “Students should be prepared for life skills in 21st Century contexts: more application, analysis, evaluation, creation” (OECD, PISA frameworks)

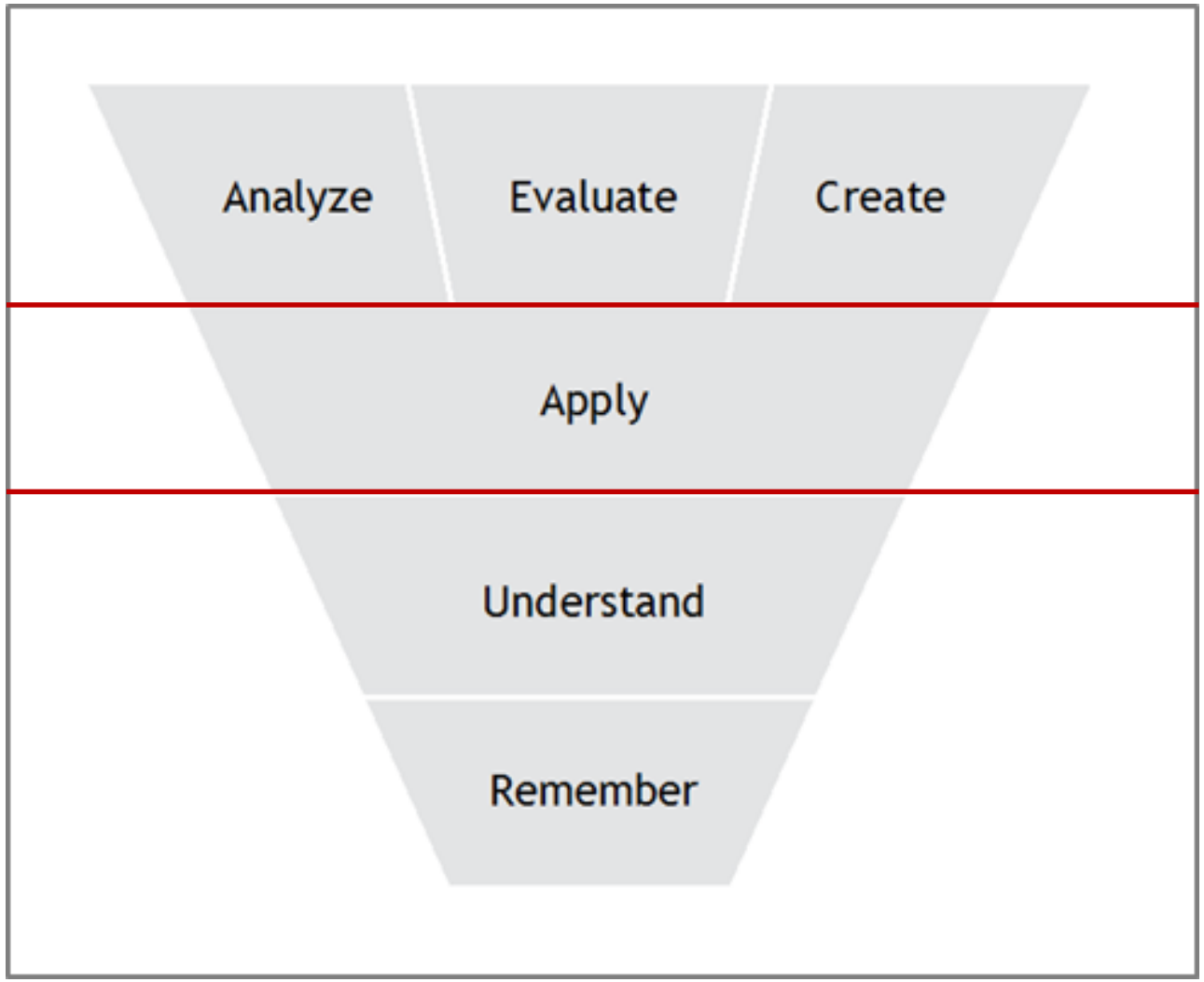
HOTS in a taxonomy : what is high?



HOTS in a taxonomy : what is high?



HOTS in a taxonomy : what is high?



HOTS : our definition

“Higher order thinking occurs when a person takes new information and information stored in memory and interrelates and/or rearranges and extends this information to achieve a purpose or find answers in perplexing situations.

(Lewis, A., & Smith, D.C., 1993)

➔ situations ➔ contexts ➔ assessment in contexts

Assessment in contexts

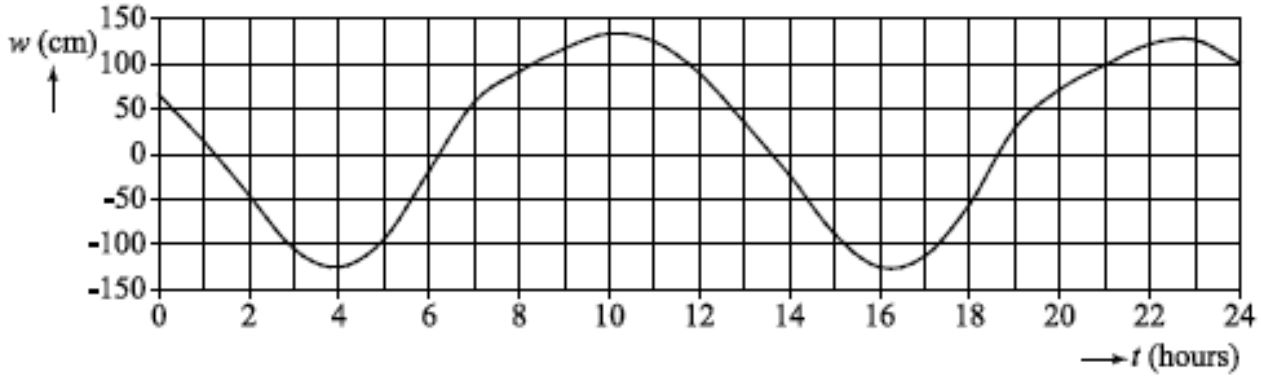
- Assessing knowledge and skills in real world contexts
- Depending on the format of assessment – paper, computer, simulator, practice – contexts will be more or less modeled

Example 1: mathematics – open question – paper based

High and low tide

Rijkswaterstaat publishes tide predictions for a number of places along the Dutch coast. These are calculated using a mathematical model based on measurements over a long period of time. Figure 1 gives the expected water level on 14 November 2012 for Schiermonnikoog.

figure 1



The values in the graph of figure 1 can be approximated by the formula $w = 4 + 128\sin(0.51(t + 5.4))$. Here, w is the water level in cm and t is the time in hours with $t = 0$ at 00:00. The moment of highest water level in the evening according to the formula differs from that moment in the graph in figure 1.

4p 17 Determine this difference in minutes.

Example 1 : marking scheme

Maximum score 4

- high tide occurs at 22:40 1
- $t \approx 2$Calculated the correct maximum of the graph at $t \approx 22.3$ 1
- Read the graph correctly: 2.3 corresponds to 22:18 or 22:19 1
- The difference is 21 or 22 (minutes) 1

Note

A margin of 10 minutes is allowed for the time of high tide.

Example 2: geography - closed and computer based

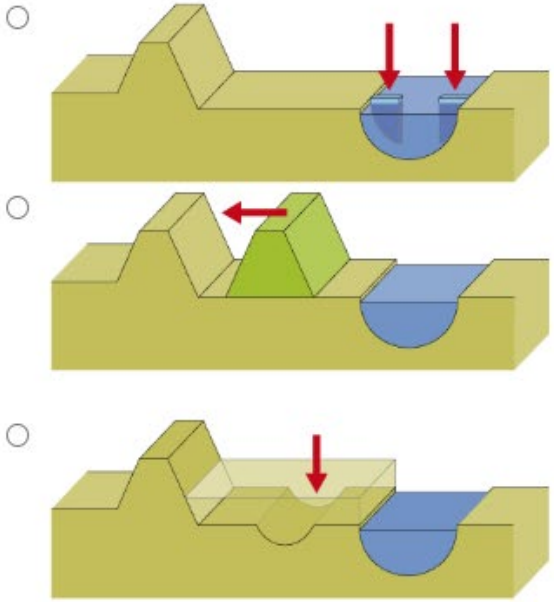
Vraag 23 van 45

1p

Room for the river

The images represent measures that correspond with the “Room for the river” project.

Which measure applies to Deventer, according to the information in the video?



Example 3: mathematics – selected response – computer based

DTT wiskunde vmbo voorbeeldopgaven

Vraag 2 van 11

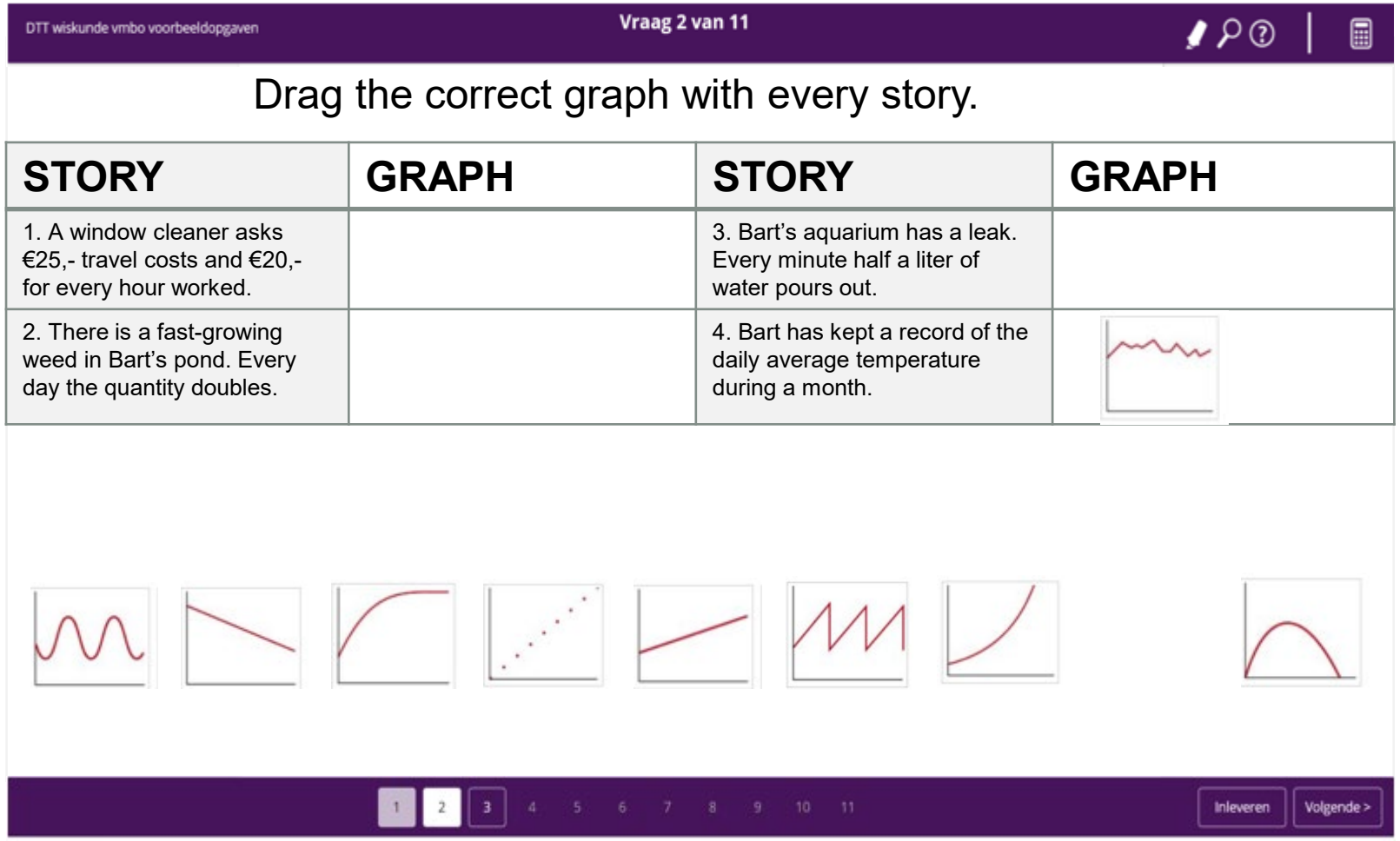
Drag the correct graph with every story.

STORY	GRAPH	STORY	GRAPH
1. A window cleaner asks €25,- travel costs and €20,- for every hour worked.		3. Bart's aquarium has a leak. Every minute half a liter of water pours out.	
2. There is a fast-growing weed in Bart's pond. Every day the quantity doubles.		4. Bart has kept a record of the daily average temperature during a month.	

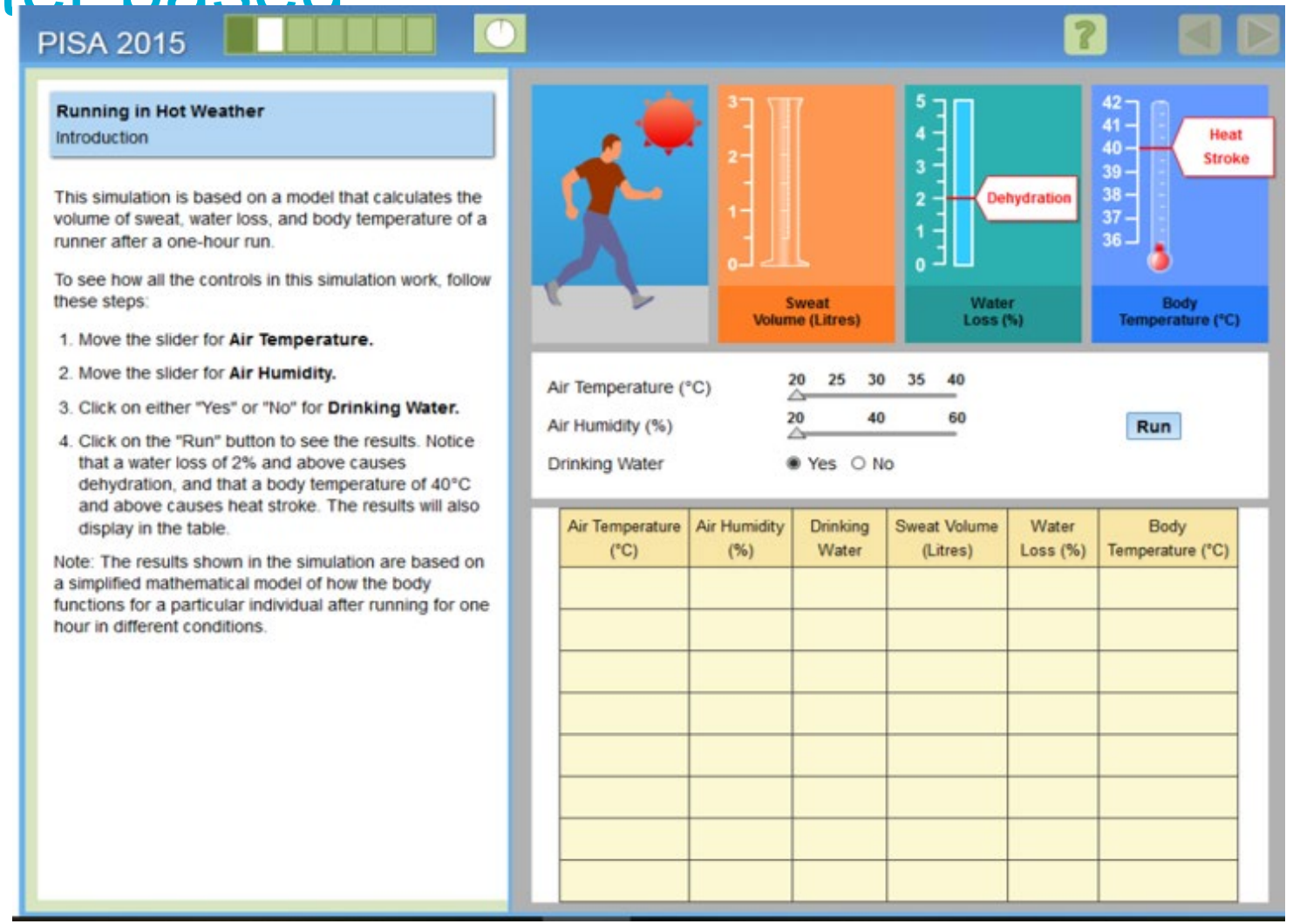
1 2 3 4 5 6 7 8 9 10 11

Inleveren Volgende >

Example 3: mathematics – selected response – computer based



Example 4: science – use of simulations – computer based



Assessing in contexts and on screen: pitfalls

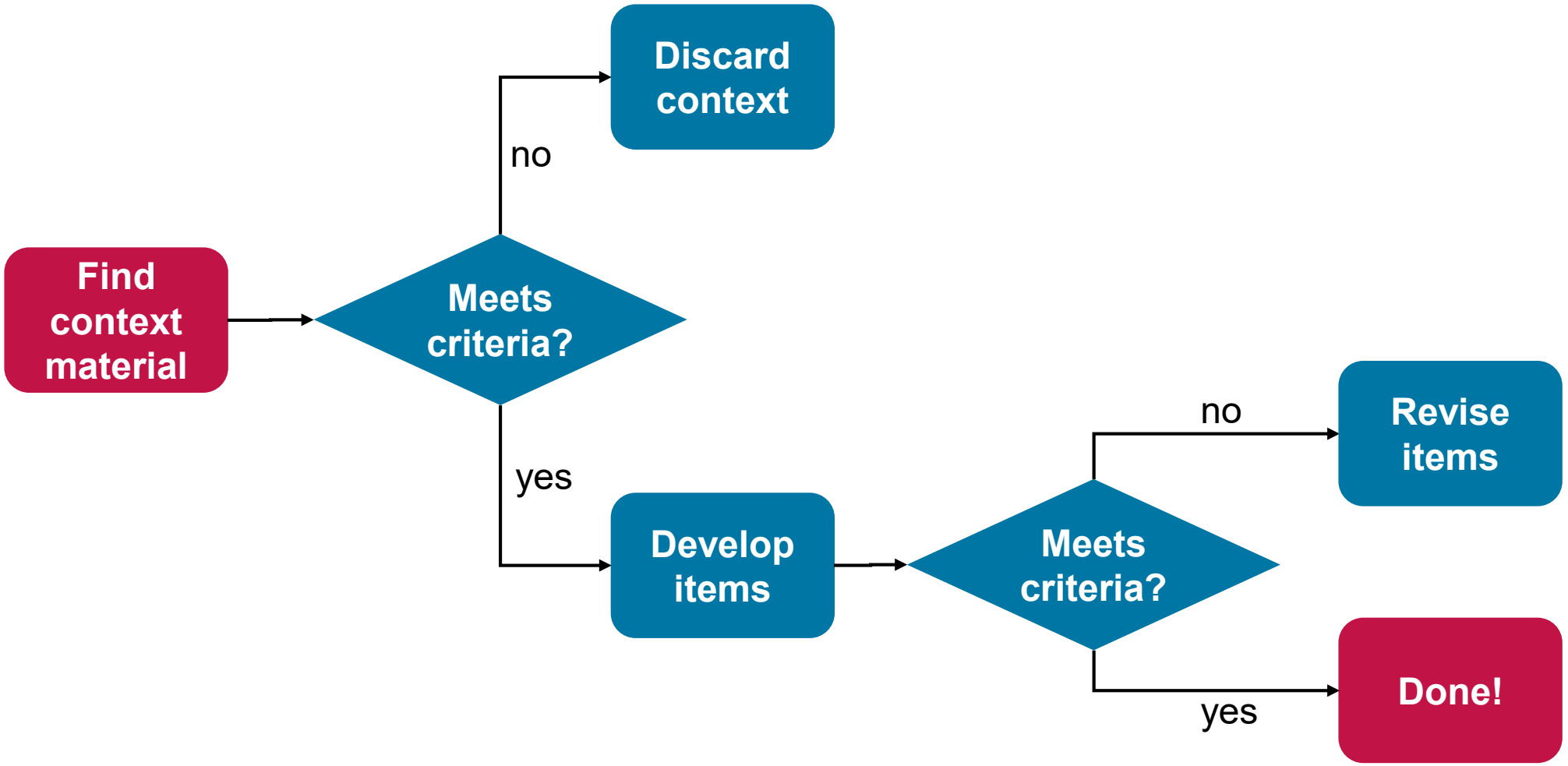
Keep in mind that it is not the context that is assessed but the knowledge and skills that are purposely related to that context!

➔ Quality criteria : how to select good contexts for assessing HOTS?

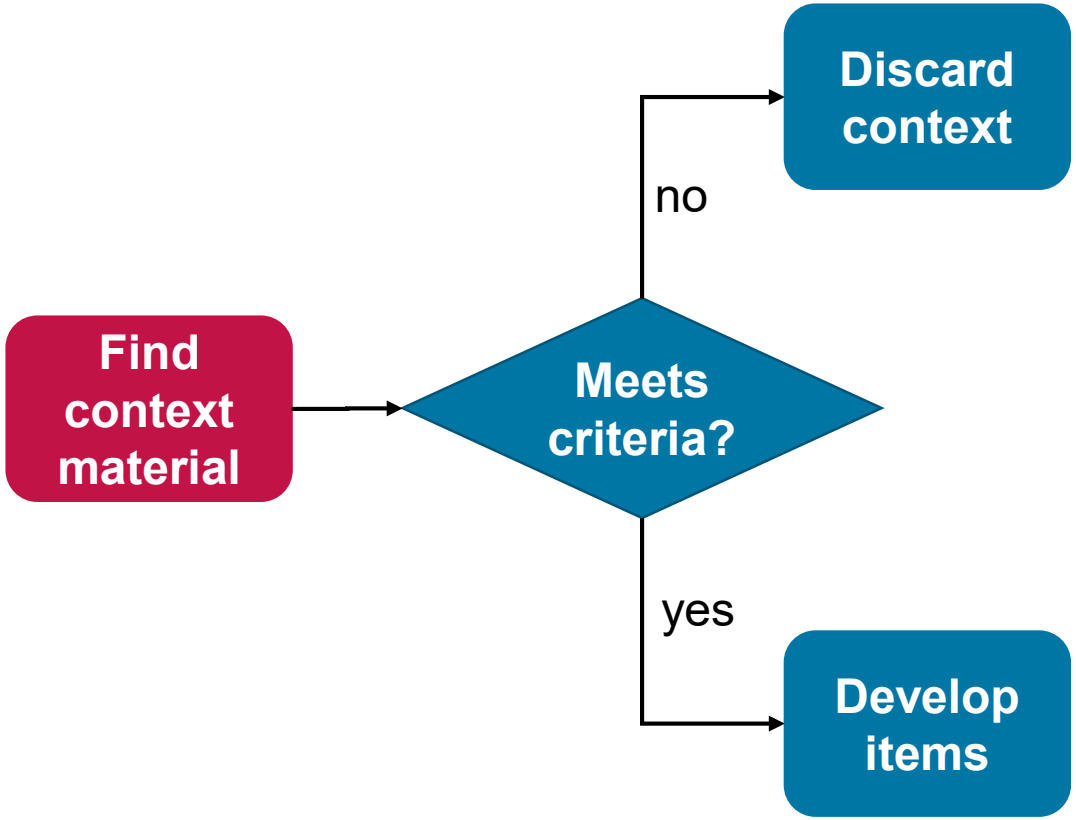
Keep in mind not to assess IT-skills but to assess the knowledge and skills that are intended!

➔ Quality criteria : how to select good item types for assessing on screen?

Item development assessment in context



Item development assessment in context



Quality criteria for selecting contexts

Context must be

- Realistic, but not necessarily real

Quality criteria for selecting contexts

Context must be

- Realistic, but not necessarily real
- Up-to-date / not outdated

Quality criteria for selecting contexts

Context must be

- Realistic, but not necessarily real
- Up-to-date / not outdated
- Functional

Quality criteria for selecting contexts

Context must be

- Realistic, but not necessarily real
- Up-to-date / not outdated
- Functional
- Natural

Quality criteria for selecting contexts

Context must be

- Realistic, but not necessarily real
- Up-to-date / not outdated
- Functional
- Natural
- Efficient

Quality criteria for selecting contexts

Context must be

- Realistic, but not necessarily real
- Up-to-date / not outdated
- Functional
- Natural
- Efficient
- Fit the platform

Quality criteria for selecting contexts

Context must be

- Realistic, but not necessarily real
- Up-to-date / not outdated
- Functional
- Natural
- Efficient
- Fit the platform
- Subject or domain related

Quality criteria for selecting contexts

Context must be

- Realistic, but not necessarily real
- Up-to-date / not outdated
- Functional
- Natural
- Efficient
- Fit the platform
- Subject or domain related
- Neutral

Quality criteria for selecting contexts

Context must be

- Realistic, but not necessarily real
- Up-to-date / not outdated
- Functional
- Natural
- Efficient
- Fit the platform
- Subject or domain related
- Neutral
- Acceptable for all (including legal rights)

Quality criteria for selecting contexts

Context must be

- Realistic, but not necessarily real
- Up-to-date / not outdated
- Functional
- Natural
- Efficient
- Fit the platform
- Subject or domain related
- Neutral
- Acceptable for all (including legal rights)
- Correspond with circle of interests of the target group

These criteria contribute to the validity and reliability of a test.

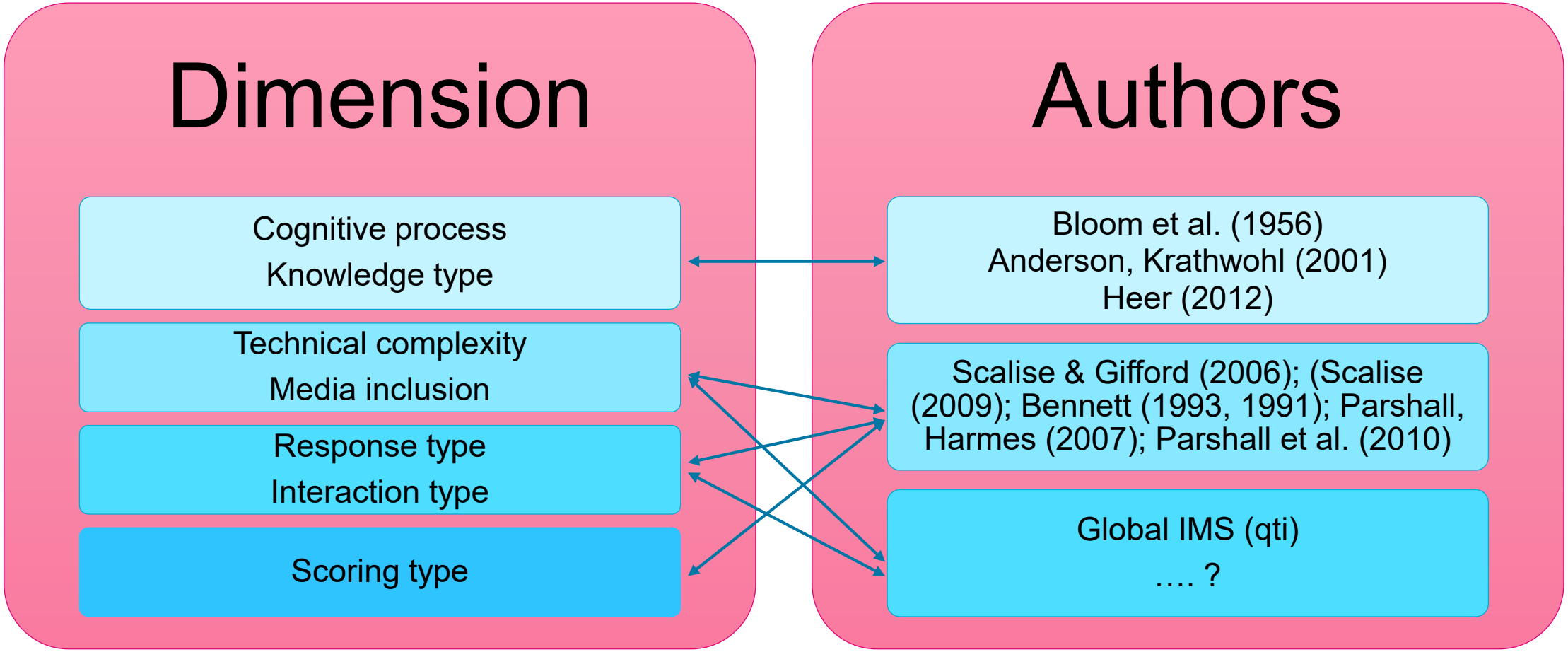
How to work with these criteria?

- Checklist
- Learning by doing
- Team work

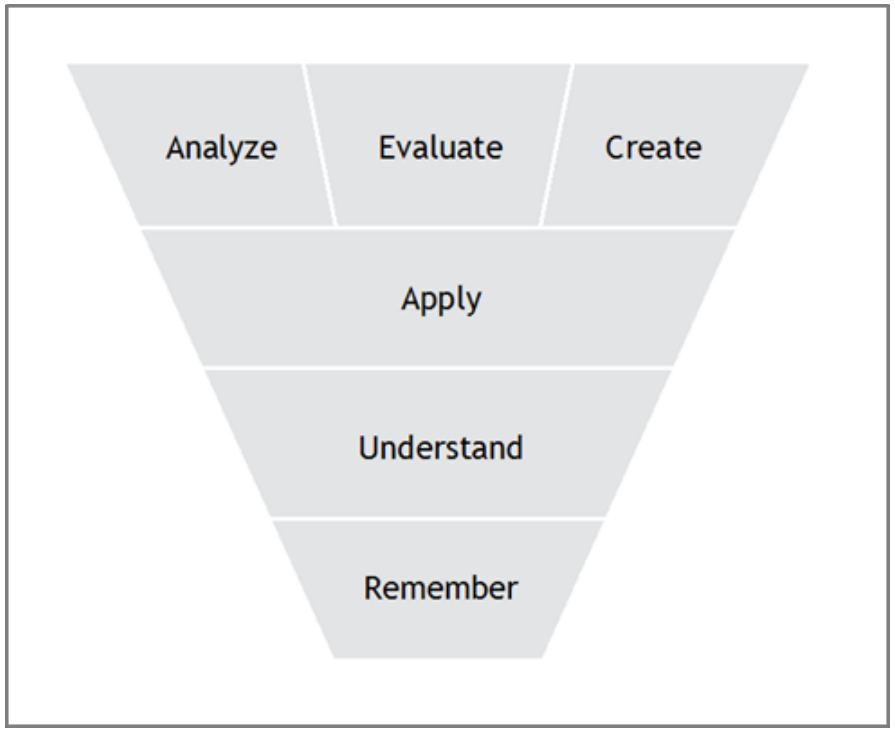
Topics

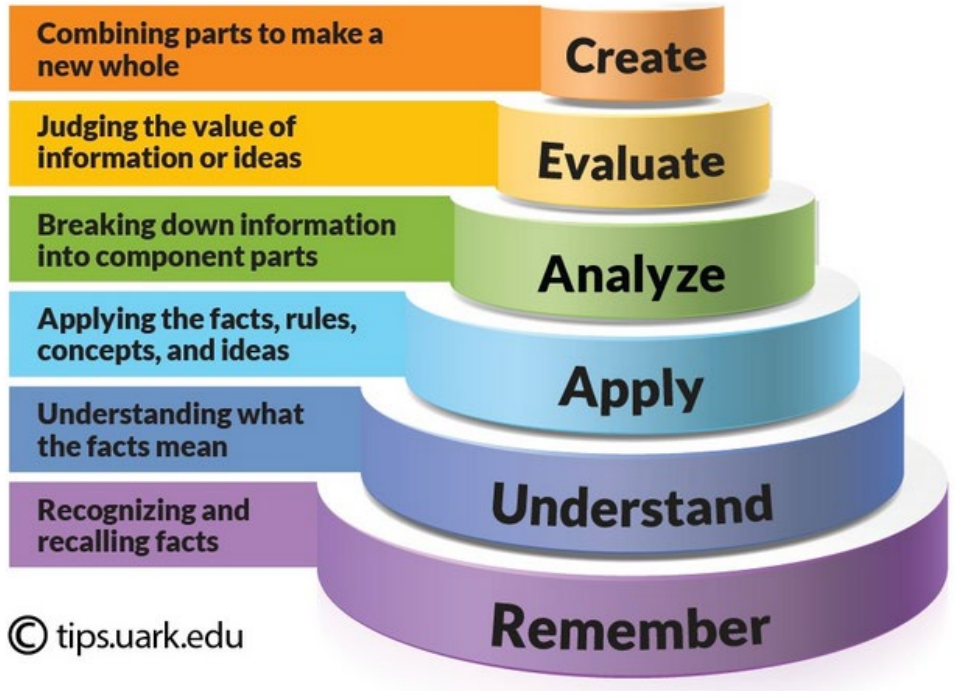
1. Higher Order Thinking Skills
2. Assessment in contexts
3. Ways to categorise e-assessment item types
   - by type
   - by assessment objective
4. Q&A

Test item types – how to categorise?



Categorising by cognitive process



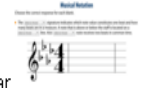
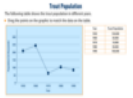
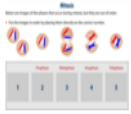
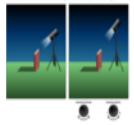


Categorising by technology & response type

Most Constrained → *Least* Constrained

Fully Selected — *Intermediate Constraint Item Types* — *Fully Constructed*

Less Complex ↓ *More Complex*

1. Multiple Choice	2. Selection/ Identification	3. Reordering/ Rearrangement	4. Substitution/ Correction	5. Completion	6. Construction	7. Presentation
1A. True/False	2A. Multiple True/False	3A. Matching	4A. Interlinear 	5A. Single Numerical Constructed	6A. Open-Ended Multiple Choice 	7A. Project
1B. Alternate Choice	2B. Yes/No with Explanation	3B. Categorizing	4B. Sore-Finger	5B. Short-Answer and Sentence Completion	6B. Figural Constructed Response	7B. Demonstration, Experiment, Performance
1C. Conventional Multiple Choice	2C. Multiple Answer	3C. Ranking and Sequencing 	4C. Limited Figural Drawing 	5C. Cloze-Procedure	6C. Concept Map	7C. Discussion, Interview
1D. Multiple Choice with New Media Distractors	2D. Complex Multiple Choice	3D. Assembling Proof	4D. Bug/Fault Correction	5D. Matrix Completion	6D. Essay and Automated Editing	7D. Diagnosis, Teaching

http://pages.uoregon.edu/kscalise/taxonomy/taxonomy.html, Kathleen Scalise, University of Oregon, June 2009

Categorising by scoring type & interaction

Parshall et al. focus on 7 types of item characteristics

dimension	Focus	example
Format	Item type	Select response, situated task, MPG
Complexity	Number of variables to consider in response	
Response action	Required manipulation for response	Click, drag and drop, write, draw
Media inclusion	Type of media	Written, visual, audio, simulation
Interactivity	Level of interactivity	
Scoring	How responses are scored	Automated, human
Fidelity	Authenticity	Task simulation

Looking at what is out there....

AEA-E workshop, Jongkamp & Hamer

Introduced participants to a range of existing digital assessment items		
Cito	International Baccalaureate	Various others

Categorise existing item types using		
Scalise (2009)	Parshall et al. (2010)	Revised Bloom

Results AEA-E Workshop (2018)

Chain of reasoning

Essay response

Create table and graph

Drawing

Most Constrained

Fully Selected

Intermediate Constraint Item Types

Less Complex

More Complex

1. Multiple Choice	2. Selection/ Identification	3. Reordering/ Rearrangement	4. Substitution/ Correction	5. Completion	6. Construction	
1A. True/False	2A. Multiple True/False	3A. Matching	4A. Interlinear	5A. Single Numerical Constructed	6A. Open-Ended Multiple Choice	Project
1B. Alternate Choice	2B. Yes/No with Explanation	3B. Categorizing	4B. Sore-Finger	5B. Short-Answer and Sentence Completion	6B. Figural Constructed Response	7B. Demonstration, Experiment, Performance
1C. Conventional Multiple Choice	2C. Multiple Answer	3C. Ranking and Sequencing	4C. Limited Figural Drawing	5C. Cloze-Procedure	6C. Concept Map	7C. Discussion, Interview
1D. Multiple Choice with New Media Distractors	2D. Complex Multiple Choice	3D. Assembling Proof	4D. Bug/Fault Correction	5D. Matrix Completion	6D. Essay and Automated Editing	7D. Diagnosis, Teaching

http://pages.uoregon.edu/kscalise/taxonomy/taxonomy.html, Kathleen Scalise, University of Oregon, June 2009

Findings

Scalise

was easy to use and new to many participants

BUT

- There are intermediate (combined) item types
- Scalise is incomplete (it is 10 years old)
- The jump from 6 to 7 is “huge”
- Many digital item types cannot be allocated

AND

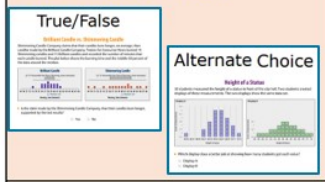
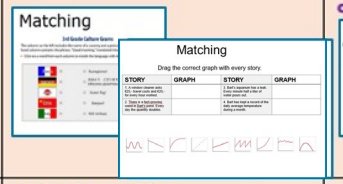
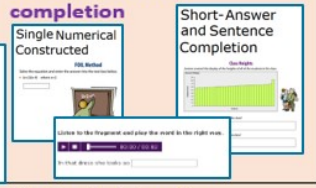
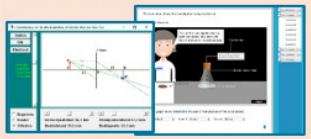
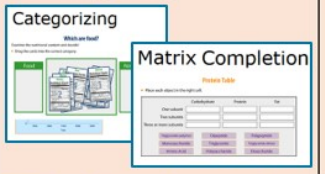
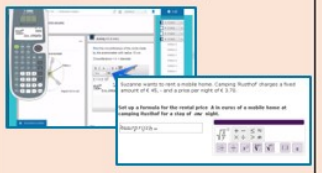
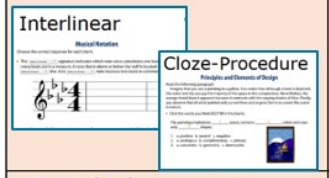
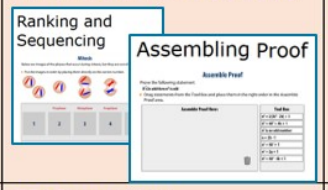
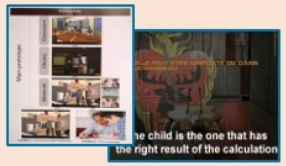
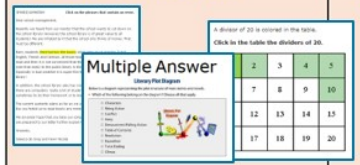
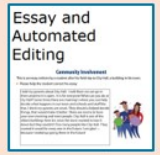
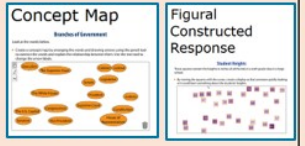
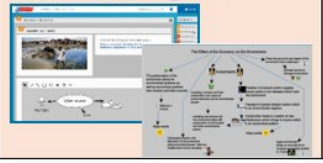
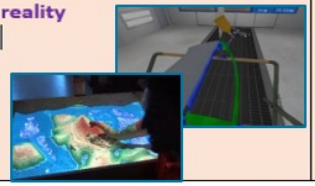
There is no connection to the assessment objective of the item

A new framework

(Hamer and Jongkamp, 2019, inspired by Scalise 2009)

Closed response → Open response

Least complex ↓ Most complex

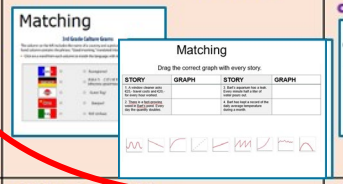
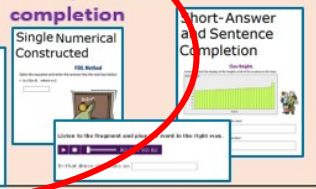
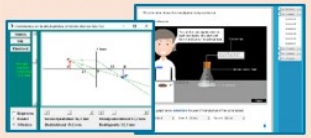
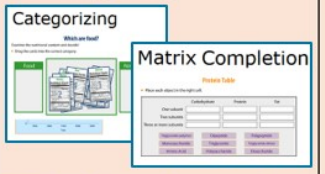
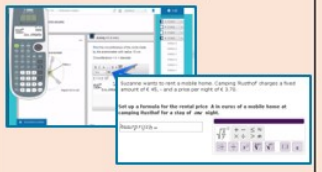
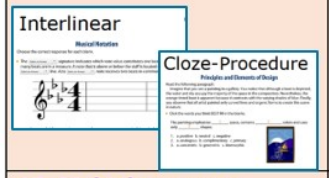
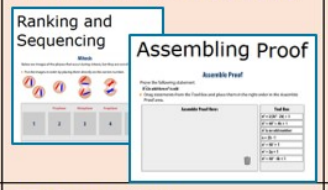
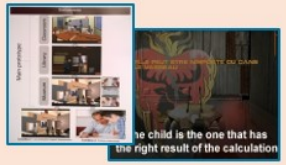
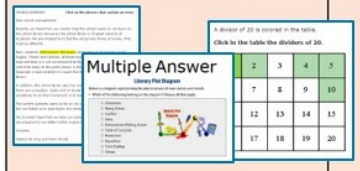
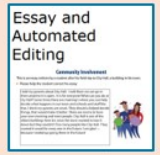
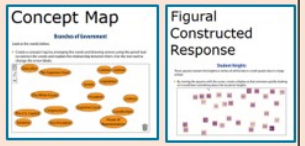
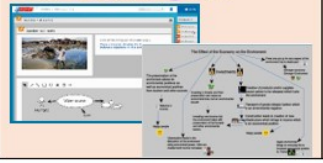
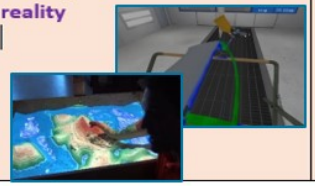
1. Choice	2. Match & order	3. Pre-structured completion	4. Custom interactive: Scaffolded/scripted	5. Free construction	6. Upload	7. Not yet allocated
1A. Anternate choice 	2A. One-on-one matching 	3A. Alphanumeric completion 	4A. Simulations & experiments 	5A. Short response with chain of reasoning	6A. Demonstration Experiment/Exhibition	
1B. Multiple choice	2B. Categorizing 	3B. Limited drawing	4B. Tailored tooling 	5B. Scaffolded open response	6B. Project report/ Paper	
1C. Inline choice 	2C. Ranking and sequencing 	3C. Limited graphic completion	4C. Avatar interaction 	5C. Essay response	6C. Audiovisual presentation; Performance;	
1D. Multiple response 	2D. Structuring 	3D. Construction using image menu/data	4D. Game-based assessment & collaborative work;	5D. Structured Oral	6D. Discussion, debate, open interview	
1E. Composite choice	2E. Arrange and re-arrange 	3E. Drawing with image menu & drawing tool 	4E. Augmented & virtual reality 	5E. Construction free drawing tool, self generated data	6E. Teaching, coaching	

A new framework

(Hamer and Jongkamp, 2019, inspired by Scalise 2009)

Closed response → Open response

Least complex → Most complex

1. Choice	2. Match & order	3. Pre-structured completion	4. Custom interactive: Scaffolded/scripted	5. Free construction	6. Upload	7. Not yet allocated
1A. Anternate choice	2A. One-on-one matching 	3A. Alphanumeric completion 	4A. Simulations & experiments 	5A. Short response with chain of reasoning	6A. Demonstration Experiment/Exhibition	
1B. Multiple choice	2B. Categorizing 	3B. Limited drawing	4B. Tailored tooling 	5B. Scaffolded open response	6B. Project report/ Paper	
1C. Inline choice 	2C. Ranking and sequencing 	3C. Limited graphic completion	4C. Avatar interaction 	5C. Essay response	6C. Audiovisual presentation; Performance;	
1D. Multiple response 	2D. Structuring 	3D. Construction using image menu/data	4D. Game-based assessment & collaborative work;	5D. Structured Oral	6D. Discussion, debate, open interview	
1E. Composite choice	2E. Arrange and re-arrange 	3E. Drawing with image menu & drawing tool 	4E. Augmented & virtual reality 	5E. Construction free drawing tool, self generated data	6E. Teaching, coaching	

2A. One-on-one matching

DTT wiskunde vmbo voorbeeldopgaven

Vraag 2 van 11

Drag the correct graph with every story.

STORY	GRAPH	STORY	GRAPH
1. A window cleaner asks €25,- travel costs and €20,- for every hour worked.		3. Bart's aquarium has a leak. Every minute half a liter of water pours out.	
2. There is a fast-growing weed in Bart's pond. Every day the quantity doubles.		4. Bart has kept a record of the daily average temperature during a month.	

1 2 3 4 5 6 7 8 9 10 11

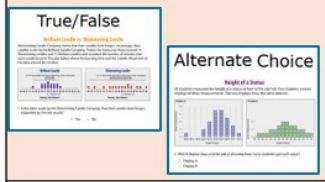
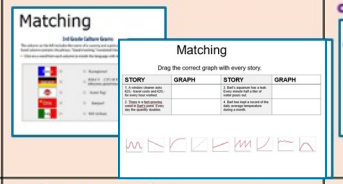
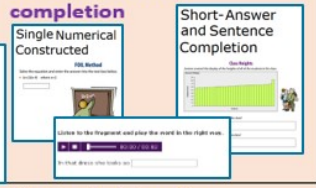
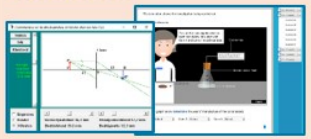
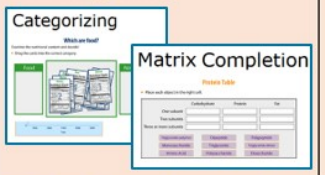
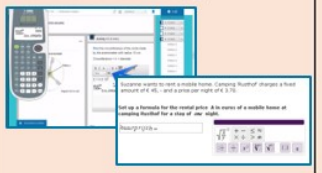
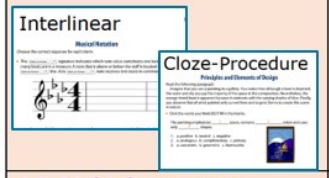
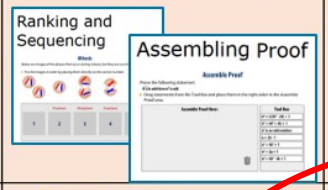
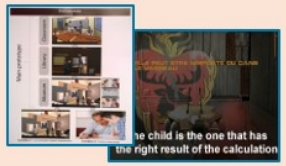
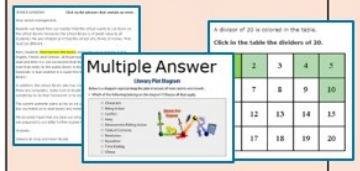
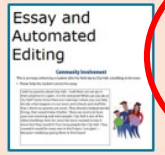
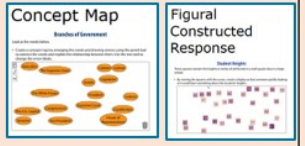
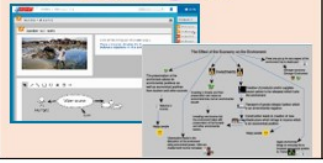
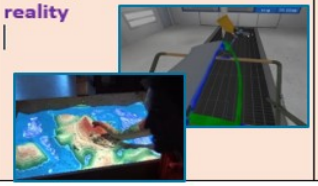
Inleveren Volgende >

A new framework

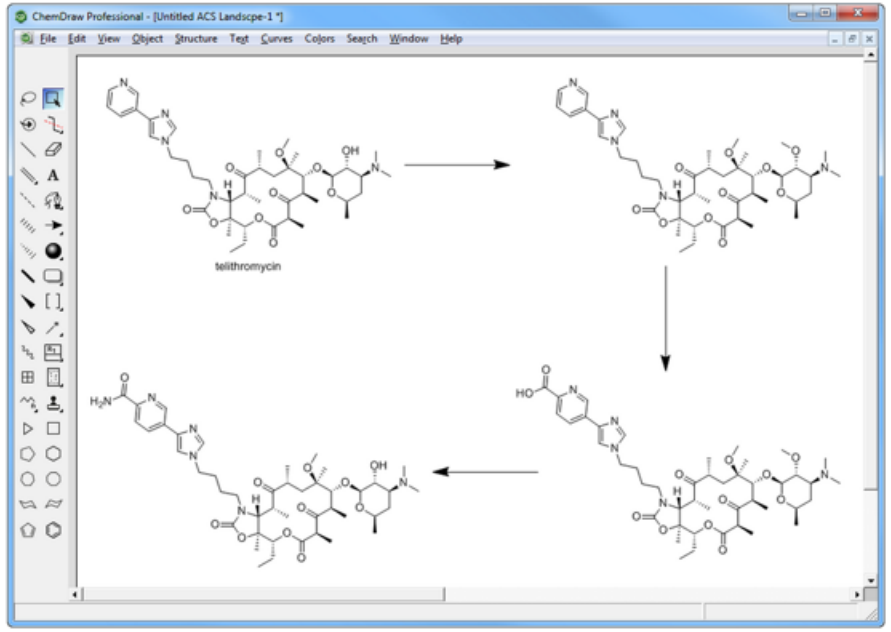
(Hamer and Jongkamp, 2019, inspired by Scalise 2009)

Closed response → Open response

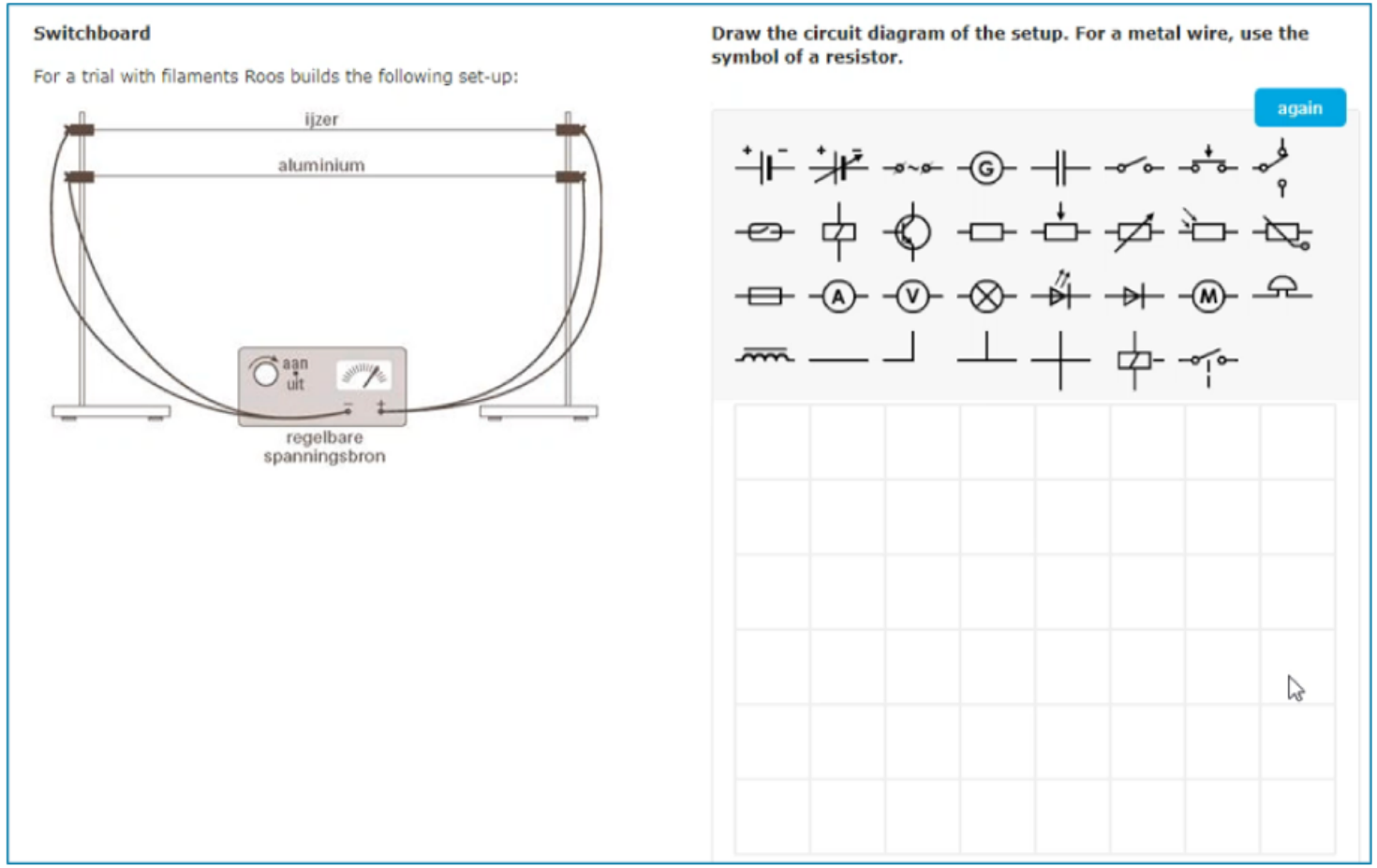
Least complex → Most complex

1. Choice	2. Match & order	3. Pre-structured completion	4. Custom interactive: Scaffolded/scripted	5. Free construction	6. Upload	7. Not yet allocated
1A. Anternate choice 	2A. One-on-one matching 	3A. Alphanumeric completion 	4A. Simulations & experiments 	5A. Short response with chain of reasoning	6A. Demonstration Experiment/Exhibition	
1B. Multiple choice	2B. Categorizing 	3B. Limited drawing	4B. Tailored tooling 	5B. Scaffolded open response	6B. Project report/ Paper	
1C. Inline choice 	2C. Ranking and sequencing 	3C. Limited graphic completion	4C. Avatar interaction 	5C. Essay response	6C. Audiovisual presentation; Performance;	
1D. Multiple response 	2D. Structuring 	3D. Construction using image menu/data	4D. Game-based assessment & collaborative work;	5D. Structured Oral	6D. Discussion, debate, open interview	
1E. Composite choice	2E. Arrange and re-arrange 	3E. Drawing with image menu & drawing tool 	4E. Augmented & virtual reality 	5E. Construction free drawing tool, self generated data	6E. Teaching, coaching	

3D. Construction with image menu/data



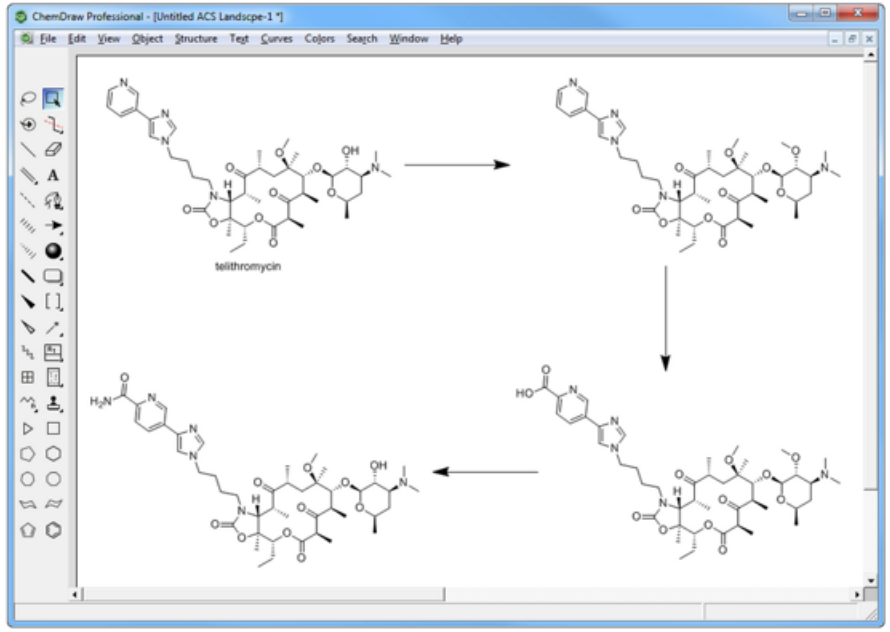
To see a video, click the image (11 min)
http://blog.cambridgesoft.com/post/2017/10/24/chemdraw-video-how-to-use-chemdraw-without-a-mouse.aspx



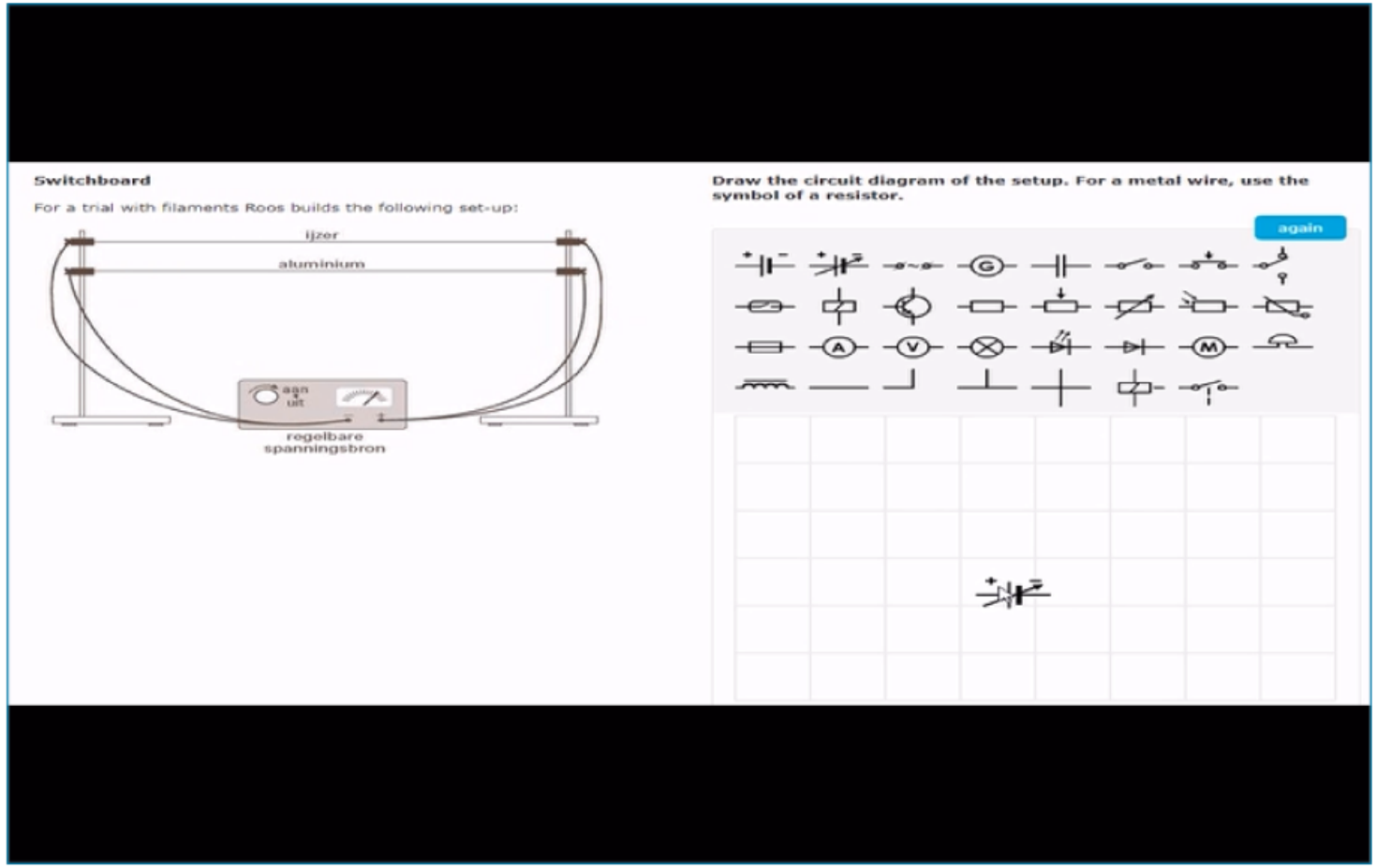
Back to taxonomy

Source: Chemdraw, CvTE, The Netherlands

3D. Construction with image menu/data



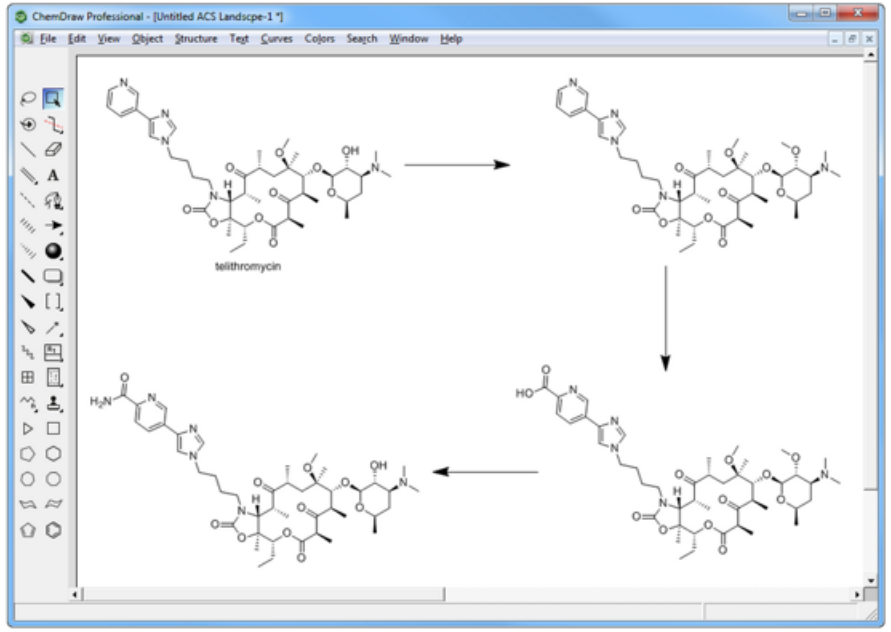
To see a video, click the image (11 min)
http://blog.cambridgesoft.com/post/2017/10/24/chemdraw-video-how-to-use-chemdraw-without-a-mouse.aspx



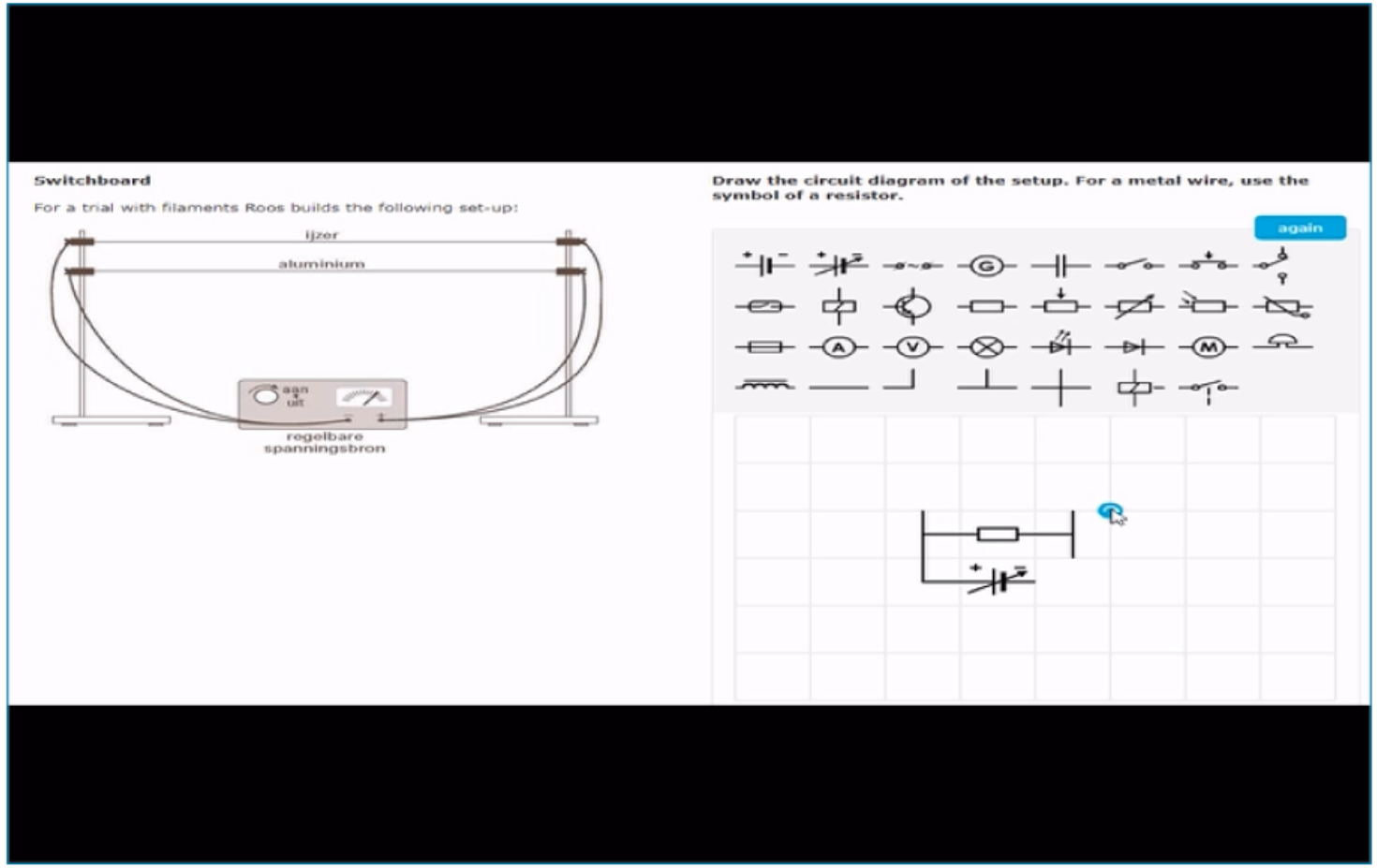
Back to taxonomy

Source: Chemdraw, CvTE, The Netherlands

3D. Construction with image menu/data



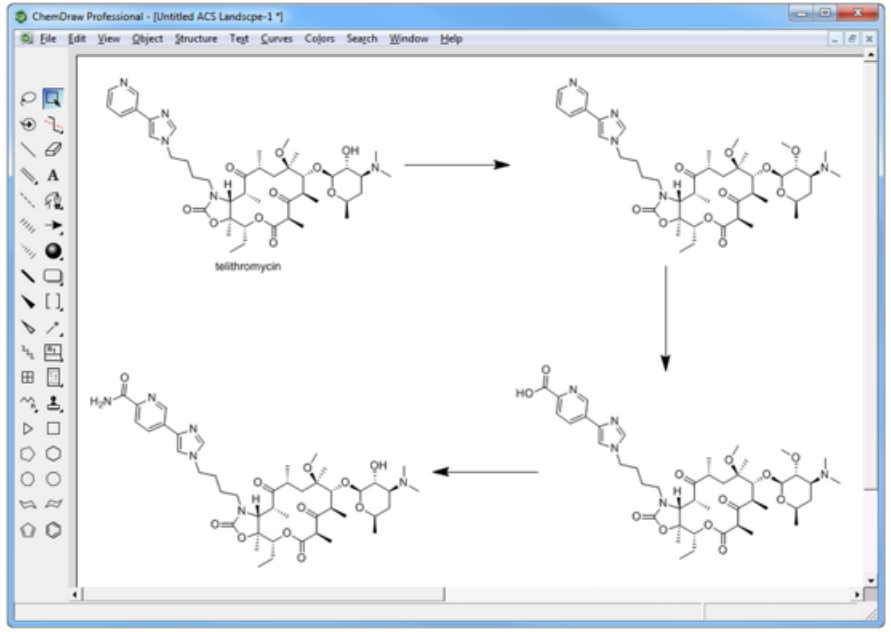
To see a video, click the image (11 min)
http://blog.cambridgesoft.com/post/2017/10/24/chemdraw-video-how-to-use-chemdraw-without-a-mouse.aspx



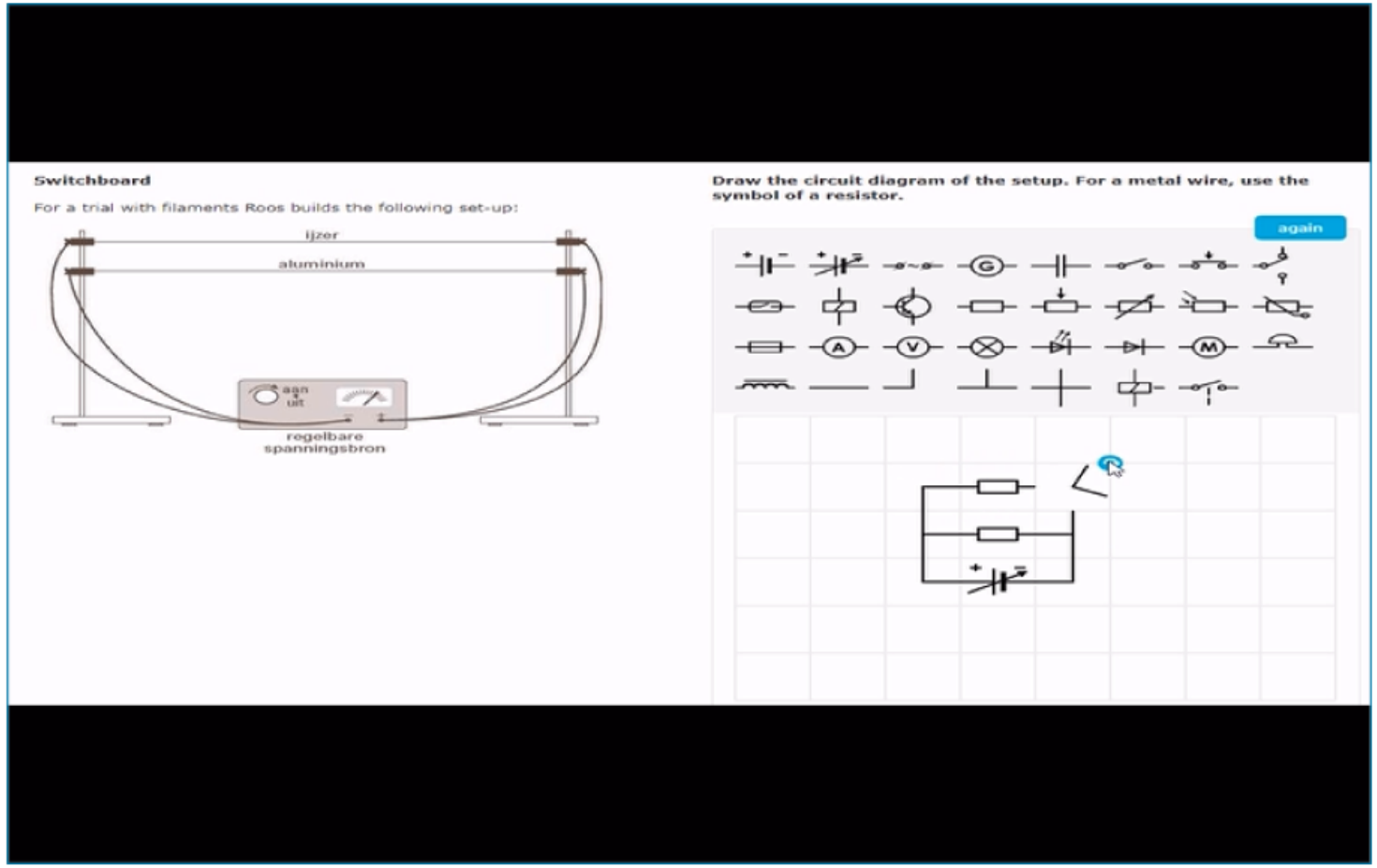
Back to taxonomy

Source: Chemdraw, CvTE, The Netherlands

3D. Construction with image menu/data



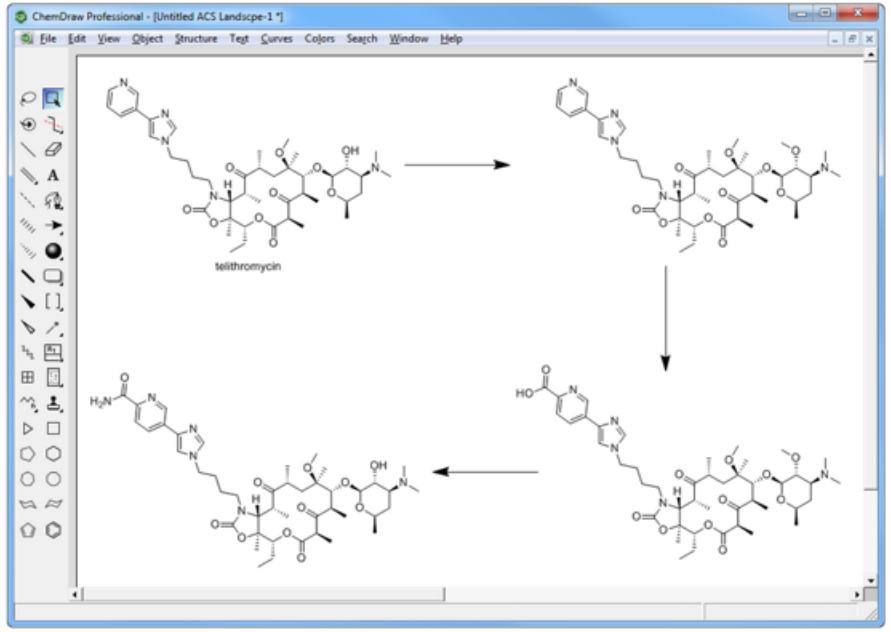
To see a video, click the image (11 min)
http://blog.cambridgesoft.com/post/2017/10/24/chemdraw-video-how-to-use-chemdraw-without-a-mouse.aspx



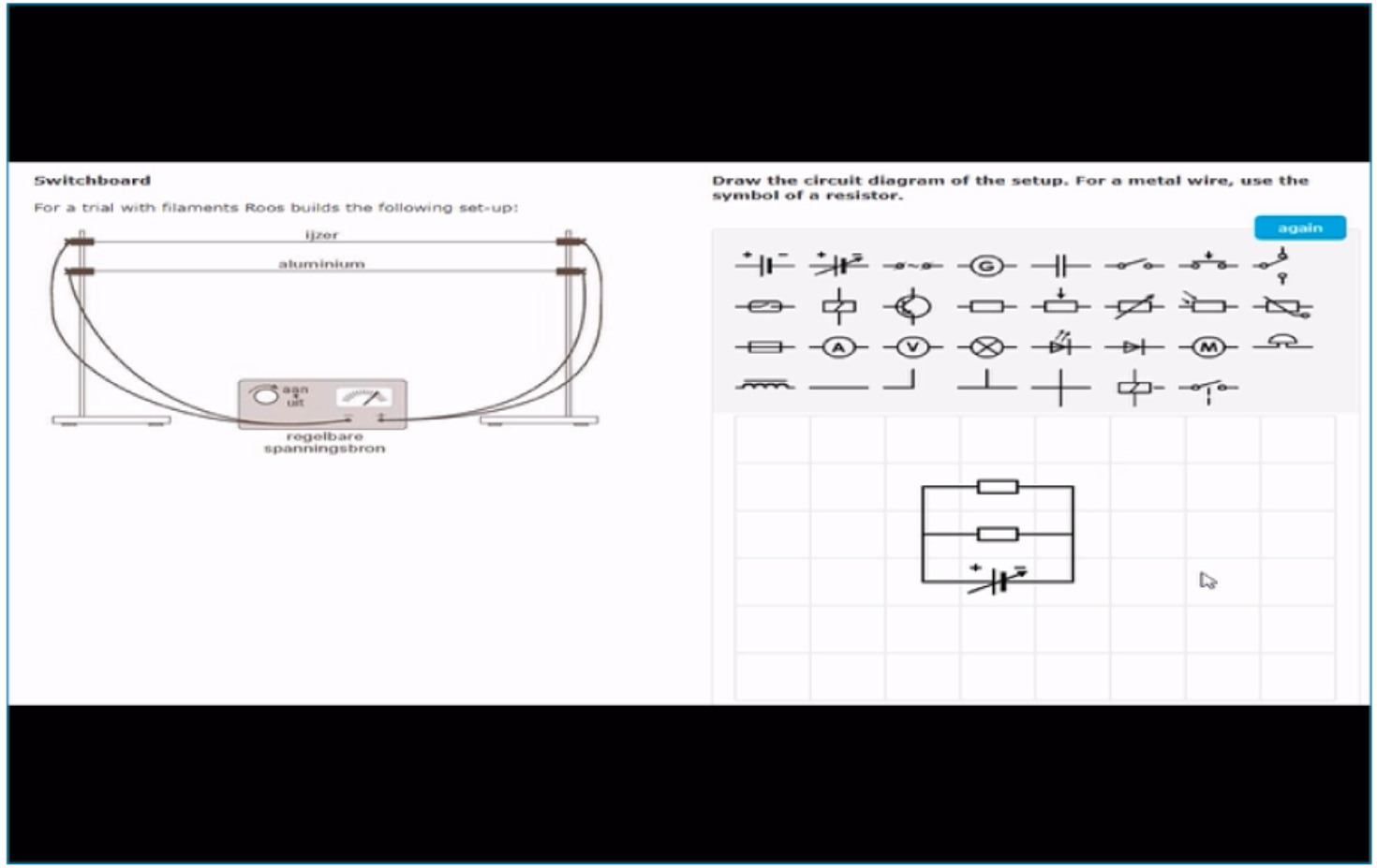
Back to taxonomy

Source: Chemdraw, CvTE, The Netherlands

3D. Construction with image menu/data



To see a video, click the image (11 min)
http://blog.cambridgesoft.com/post/2017/10/24/chemdraw-video-how-to-use-chemdraw-without-a-mouse.aspx



Back to taxonomy

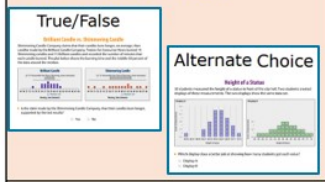
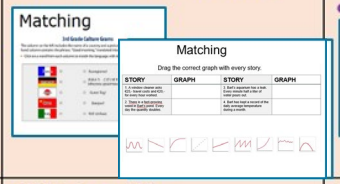
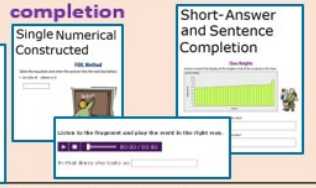
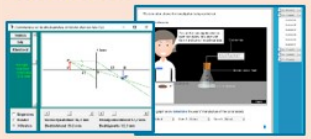
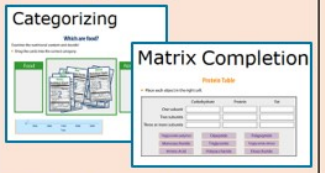
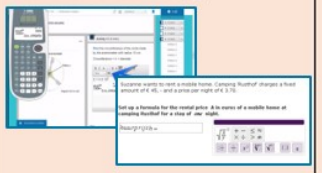
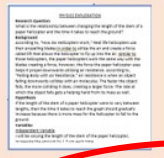
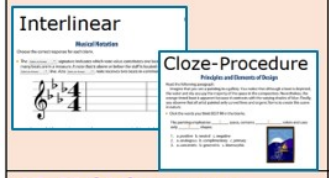
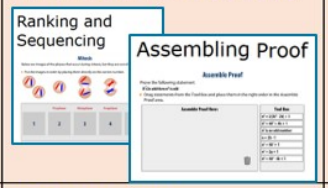
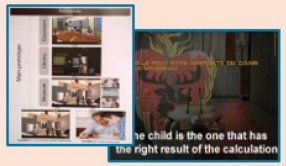
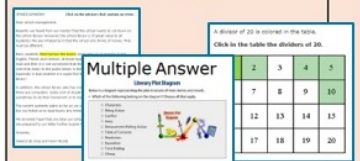
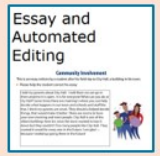
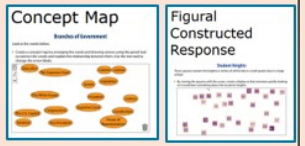
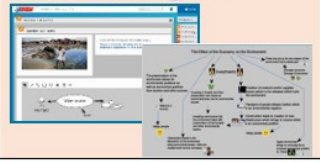
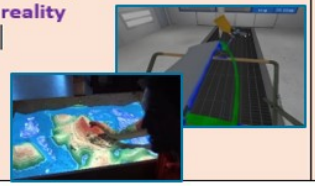
Source: Chemdraw, CvTE, The Netherlands

A new framework

(Hamer and Jongkamp, 2019, inspired by Scalise 2009)

Closed response → Open response

Least complex ↓ Most complex

1. Choice	2. Match & order	3. Pre-structured completion	4. Custom interactive: Scaffolded/scripted	5. Free construction	6. Upload	7. Not yet allocated
1A. Anternate choice 	2A. One-on-one matching 	3A. Alphanumeric completion 	4A. Simulations & experiments 	5A. Short response with chain of reasoning	6A. Demonstration Experiment/Exhibition	
1B. Multiple choice	2B. Categorizing 	3B. Limited drawing	4B. Tailored tooling 	5B. Scaffolded open response	6B. Project report/ Paper 	
1C. Inline choice 	2C. Ranking and sequencing 	3C. Limited graphic completion	4C. Avatar interaction 	5C. Essay response	6C. Audiovisual presentation; Performance;	
1D. Multiple response 	2D. Structuring 	3D. Construction using image menu/data	4D. Game-based assessment & collaborative work;	5D. Structured Oral	6D. Discussion, debate, open interview	
1E. Composite choice	2E. Arrange and re-arrange 	3E. Drawing with image menu & drawing tool 	4E. Augmented & virtual reality 	5E. Construction free drawing tool, self generated data	6E. Teaching, coaching	

6C. Upload – presentation, performance, exe

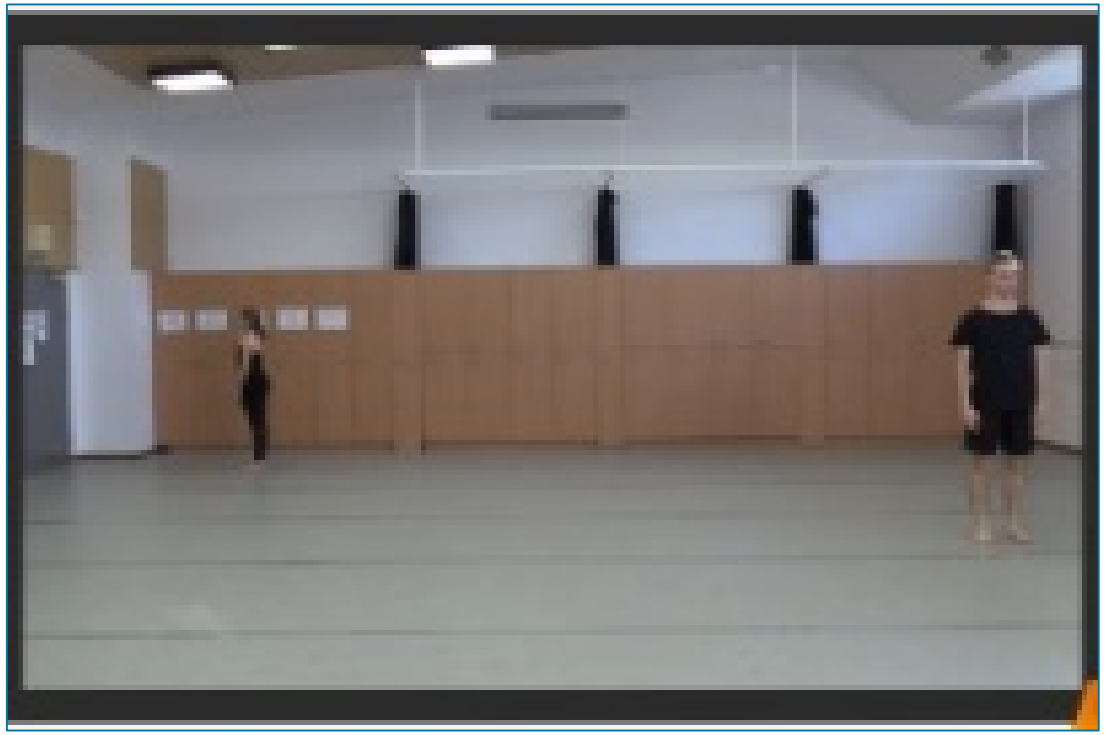
'Battle of the Seasons'

In this work, I wanted to show the battle between Summer and Winter. The female dance is summer and the male dancer is winter. They are both flighting to keep their weather in season.

At the **first section** of the dance the dancers do not connect. They move around the space in a circle as they prepare for the battle. You will see a pushing movement with a twist of the wrist; this is repeated to show the seasons tempting each other into battle. The dancers also do not look at each other often. This is to show they are angry and are not wanting to reason with each other.

In the **second section,** I have choreographed lots of partner work. This is to show the battle. I wanted the lifts and partner work to look like the seasons were in battle and fighting each other. I really tried to challenge myself with the lifts I created with my partner. I had to work hard on building my strength so that I was able to perform the movements safely.

The preparation to perform these two works is similar in many ways, however, there are some differences. The partner work 'Battle of the Seasons' is very intense and I have to ensure my mind and body are prepared, as my partner is relying on me.



Back to taxonomy

Source: IB DP Dance trial, 2019

A new framework

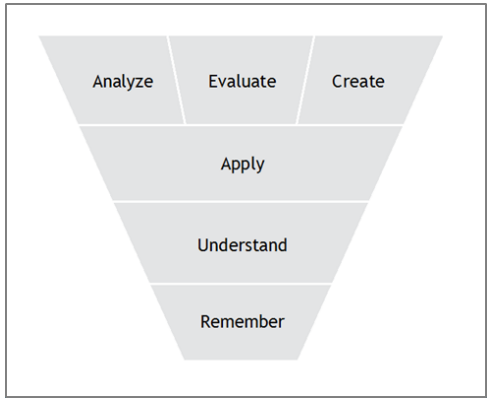
(Hamer and Jongkamp, 2019, inspired by Scalise 2009)

Closed response → Open response

Least → Most complex

	1. Choice	2. Match&Order	3. Pre-structured completion	4. Custom interactive	5. Free construction	6. Upload	7. Not allocated
A	**1A Alternate choice** Items that require the test-taker to select one of two closed response options.	**2A One-on-one match** Items that require the test-taker to pairwise match objects, concepts or item stems with possible responses.	**3A Alphanumerical completion** Items that require the test-taker to provide a response that consists of a number, letters, a couple of words, or a combination. Aka 'short-answer constructed response'	**4A Simulations and experiments** Items that require the test-taker to conduct an onscreen experiment and collect or measure the results.	**5A Short response and chain of reasoning** Items that require the test-taker to provide a series of short-response answers, each building on the previous response, creating an argument or a chain of reasoning.	**6A Demonstration / experiment** Items that require the test-taker - to conduct a real-word experiment / demonstration and - to upload the results in an appropriate digital format.	
B	**1B Multiple choice** Items that require the test-taker to select one correct option out of three or more closed response options	**2B Categorizing** Items that require the test-taker to categorize an object, concept or other construct into an appropriate umbrella class.	**3B Limited drawing** Items that require the test-taker to draw elementary shapes, like lines, circles, triangles.	**4B Tailored tooling** Items that require the test-taker to combine regular input devices such as keyboard and mouse with a given set of symbols and tools in order to produce a response. E.g. Equation editors, capture of onscreen calculator workings	**5B Scaffolded open response** Items that require the test-taker to provide an extended textual response which is guided by a series of prompts or suggestions.	**6B Project report / paper** Items that require the test-taker - to write a report or paper and - to upload the documents in an appropriate digital format.	
C	**1C Inline choice** Items that require the test-taker to - select one correct response option from two or more options, - one or a series of choices embedded within a text passage Aka 'Cloze', 'completion'.	**2C Ranking and sequencing** Items that require the test-taker to arrange objects, concepts or other constructs in a ranked or sorted order in a one-dimensional space.	**3C Limited graphic completion** Items that require the test-taker to move, adjust and rotate elementary shapes.	**4C Avatar interaction** Items that require the test-taker to interact with one or more avatar(s) in order to produce a response. E.g. Chatbot, Conversational Assessment	**5C Essay response** Items that require the test-taker to provide an extended textual response in a constrained (one screen) or unconstrained (scrollable) answer space.	**6C Audiovisual presentation, performance** Items that require the test-taker - to conduct an audiovisual presentation or performance and - to upload the audio or video recording.	
D	**1D Multiple response** Items that require the test-taker to - select out of three or more closed response options - where multiple options are correct.	**2D Structuring** Items that require the test-taker to structure objects, concepts or other constructs in a pattern, other than by (re)ordering or categorizing.	**3D Constructing with image menu / data** Items that require the test-taker to - select, move or rotate elementary shapes from a menu, and in order to construct a scheme, installation, graph, diagram etcetera.	**4D Game-based assessment & collaborative work** Items that require the test-taker to conduct a task which has traits of a game, that is: - A goal - Rules - A feedback system - Voluntary participation	**5D Structured oral** Items that require the test-taker to provide oral response to a series of structured questions.	**6D Discussion, debate** Items that require the test-taker - to perform a task that involves a discussion or debate and - to upload the audio or video recording.	
E	**1E Composite choice** Items that require the test-taker - to select one out of two or more options, and - to provide an explanation or example connected to the chosen option.	**2E Arrange and rearrange** Items that require the test-taker - to structure objects, concepts or other constructs in a pattern, other than by (re)ordering or categorizing, - and to construct relationships between the objects, concepts or other constructs.	**3E Drawing with image menu / drawing tool** Items that require the test-taker - to select, move, rotate or adjust elementary shapes from a menu, - to draw new shapes or connections in order to construct a scheme/flowchart installation, illustration etcetera.	**4E Augmented and virtual reality** Items that require the test-taker to interact in real-time with an environment that - provides multiple sensory information, - simulates a real-world situation in a realistic way, either in combination with physical world perception (AR) or as complete replacement of the physical world (VR)	**5E Construction free drawing, self-generated data** Items that require the test-taker - to draw shapes in a free style manner, and/or - using self-collected data.	**6E Teaching, coaching** Items that require the test-taker - to perform a task that involves teaching or coaching and - to upload the audio or video recording.	

Linking item types to assessment objectives (Revised Bloom)

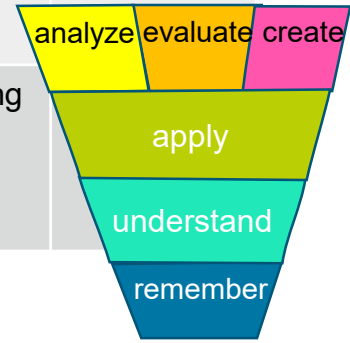


Preliminary results found – link to objectives

Closed response → Open response

Least → Most complex

1. Choice	2. Match&Order	3. Pre-structured completion	4. Custom interactive	5. Free construction	6. Upload	7. Not allocated
1A Alternate Choice	2A One-on-one match	3A Alphanum. Compl.	4A Simulations& experiments	5A Short response & chain of reasoning	6A Demonstration/ experiment/project	
1B Multiple Choice	2B Categorizing	3B Limited drawing	4B Tailored tooling	5B Scaffolded open response	6B Project report/ paper	
1C Inline choice	2C Ranking & sequencing	3C Limited graphic completion	4C Avatar Interaction	5C Essay response	6C Audiovisual presentation, performance	
1D Multiple response	2D Structuring	3D Constructing with image menu/data	4D Educational gaming & collaborative Work	5D Structured oral	6D Discussion	
1E Composite choice	2E Arrange & rearrange	3E Drawing with image menu/drawing tool	4E Augmented & Virtual reality	5E Construction free drawing, self generated data	6E Teaching/coaching	



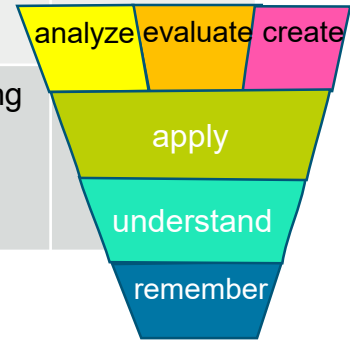
(Hamer and Jongkamp, 2019, inspired by Scalise 2009)

Preliminary results found – link to objectives

Closed response → Open response

Least → Most complex

1. Choice	2. Match&Order	3. Pre-structured completion	4. Custom interactive	5. Free construction	6. Upload	7. Not allocated
1A Alternate Choice	2A One-on-one match	3A Alphanum. Compl.	4A Simulations& experiments	5A Short re & chain of re	6A Demonstration/ experiment/project	
1B Multiple Choice	2B Categorizing	3B Limited drawing	4B Tailored tooling	5B Scaffolded open response	6B Project report/ paper	
1C Inline choice	2C Ranking & sequencing	3C Limited graphic compl.	4C Avatar Interaction	5C Essay response	6C Audiovisual presentation, performance	
1D Multiple response	2D Structuring	3D Constructing with image menu/data	4D Educational gaming & collaborative Work	5D Structured oral	6D Discussion	
1E Composite choice	2E Arrange & rearrange	3E Drawing with image menu/drawing tool	4E Augmented & Virtual reality	5E Construction free drawing, self generated data	6E Teaching/coaching	



(Hamer and Jongkamp, 2019, inspired by Scalise 2009)

Future plans: validation of the framework

Expand with

- larger and more diverse set of items
- More experts

Check on

- Robustness
- Ease of use
- Aptitude as guideline for item development / assessment of HOTS.

Publish in peer-reviewed journal

Interested? Do you want to help us improve the new framework of item types?

How : Share items and/or engage in a classification task

→ please contact caroline.jongkamp@cito.nl or rebecca.hamer@ibo.org

Thank you for your attention!

nico.dieteren@cito.nl

linkedin.com/in/nico-dieteren

caroline.jongkamp@cito.nl

linkedin.com/in/carolinejongkamp

cito.com