Fontys
UNIVERSITY OF APPLIED SCIENCES

Open Universiteit
www.ou.nl

Balancing test construction: the influence of pictures on answering test items

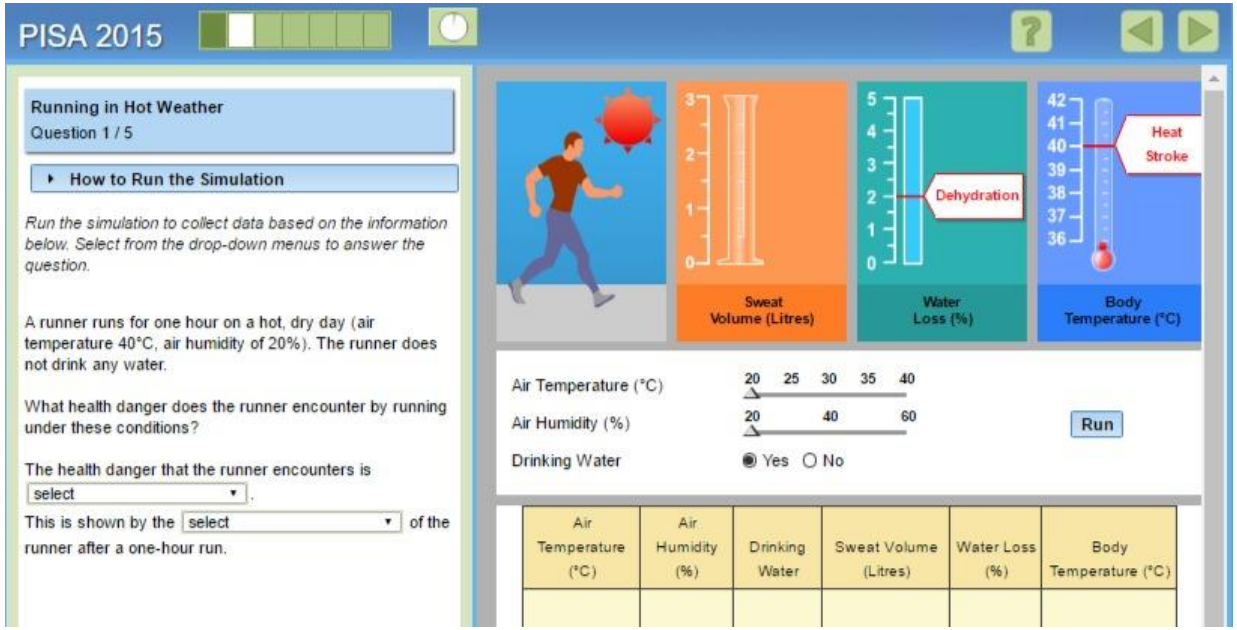
Jorik Arts, Kim Dirkx, Wilco Emons, Desirée Joosten-ten Brinke, Halszka Jarodzka

Lectoraat : Technology enhanced assessment

ORD 2023
AMSTERDAM 5–7 JULI
KANS OP BALANS

Multimedia in test items

- Increased use of multimedia (pictures) in test conditions (Dirkx *et al.*, 2021)
- …, but no clear guidelines on how to use them… (Dirkx et al., 2021; Kirschner et al., 2017)



Effects of multimedia in test tems

- Multimedia effect in testing:
  - The combination of text and picture make test items easier for students (e.g. Lindner, Ihme, Saβ, and Köller, 2018; Lindner, 2020 Saβ, Wittwer, Senkbeil, and Köller, 2012)

Multimedia Effect in Problem Solving: A Meta-Analysis

Liru Hu[1] · Gaowei Chen[1] · Pengfei Li[2] · Jing Huang[3]

Accepted: 22 February 2021/Published online: 13 March 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

The results showed a significant small-to-medium multimedia effect size on response accuracy (Hedges's $g = 0.32$)

Effects of multimedia in test tems

- Multimedia effects in testing (Hu, 2021):
  - Enhanced response accuracy
  - Lower mental effort
  - No effect on time-on-task

Functionality of the picture

- Decorative – representational – informational – organizational

Decorative Picture

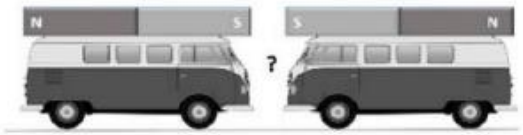
Anton has two identical toy buses and two equally sized rod magnets. He sets up the buses opposite each other and places one of the two rod magnets on each of the buses. The south pole of the magnet on one bus points to the south pole of the magnet on the other bus. Anton pushes the buses close together and then lets them go.

What happens to the toy buses when Anton lets them go?

A The buses attract each other.
B The buses attract each other first and then repulse each other.
C The buses are repulsing each other.
D Nothing, the buses remain in place.

Representational Picture

Anton has two identical toy buses and two equally sized rod magnets. He sets up the buses opposite each other and places one of the two rod magnets on each of the buses. The south pole of the magnet on one bus points to the south pole of the magnet on the other bus. Anton pushes the buses close together and then lets them go.



What happens to the toy buses when Anton lets them go?

A The buses attract each other.
B The buses attract each other first and then repulse each other.
C The buses are repulsing each other.
D Nothing, the buses remain in place.

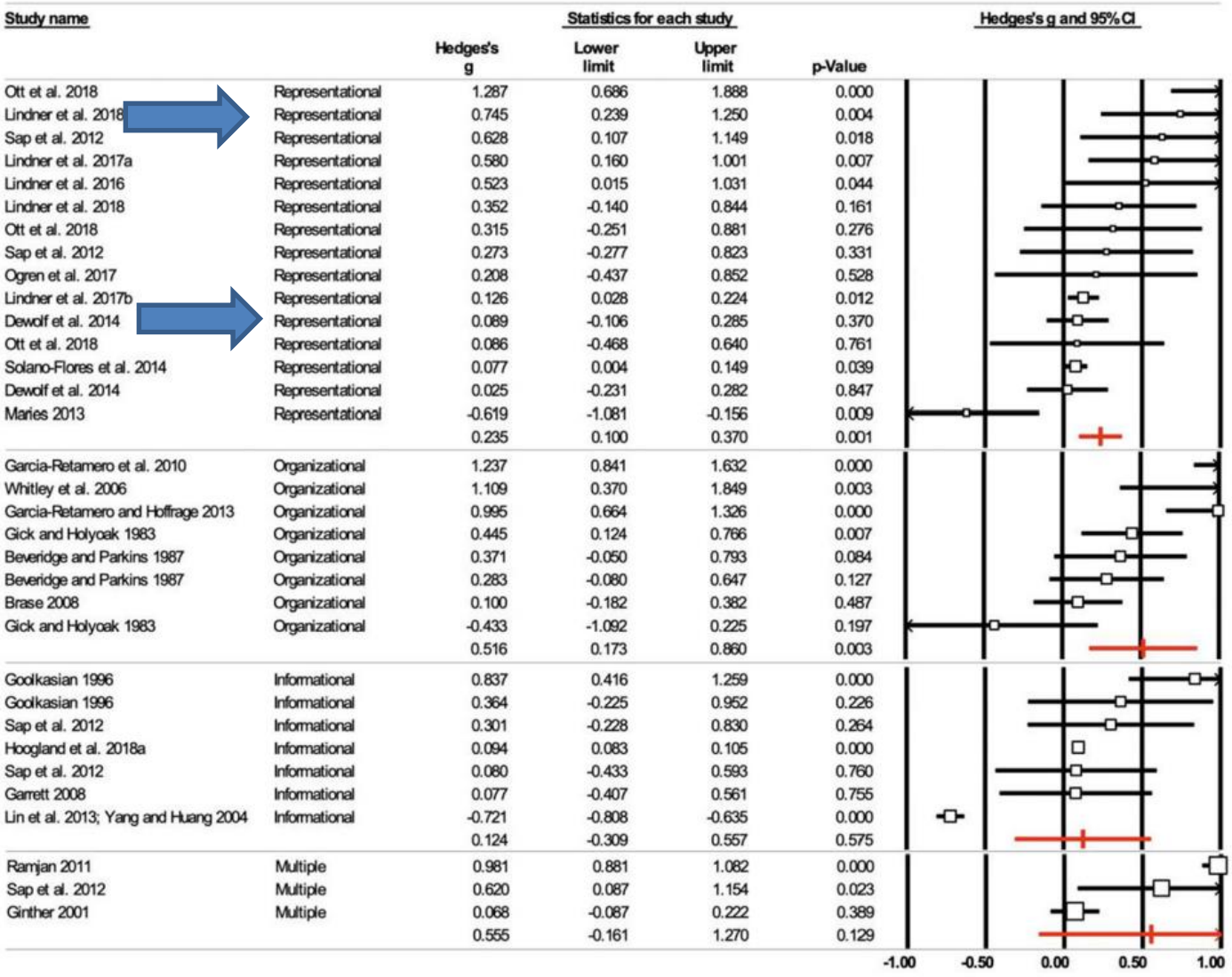
Lindner, 2020

Typical outcomes

		Scientific Items		
		Text-Only	DP	RP
Solution Success	M^a *(SD)*	.52 *(0.15)*	.52 *(0.17)*	.57 *(0.16)*
	$\eta^2_{partial}$ [b]	-	.000	.183
	EST[c] *(SE)*	**.51 *(0.15)***	**.52 *(0.17)***	**.58 *(0.16)***
Perceived Item Ease	*M (SD)*	2.91 *(0.21)*	2.93 *(0.23)*	3.03 *(0.20)*
	$\eta^2_{partial}$	-	.022	.493
	EST *(SE)*	**2.88 *(0.06)***	**2.91 *(0.06)***	**3.00 *(0.06)***
Time on Task	*M (SD)*	35.43 *(8.70)*	35.77 *(8.49)*	35.28 *(8.58)*
	$\eta^2_{partial}$	-	.017	.002
	EST *(SE)*	**34.89 *(1.87)***	**35.39 *(1.87)***	**34.81 *(1.87)***
Item-Solving Satisfaction	*M (SD)*	2.86 *(0.14)*	2.86 *(0.18)*	2.94 *(0.16)*
	$\eta^2_{partial}$	-	.002	.419
	EST *(SE)*	**2.83 *(0.05)***	**2.85 *(0.02)***	**2.92 *(0.02)***

Lindner, 2020

What are we missing?



Study name		Hedges's g	Lower limit	Upper limit	p-Value
Ott et al. 2018	Representational	1.287	0.686	1.888	0.000
Lindner et al. 2018	Representational	0.745	0.239	1.250	0.004
Sap et al. 2012	Representational	0.628	0.107	1.149	0.018
Lindner et al. 2017a	Representational	0.580	0.160	1.001	0.007
Lindner et al. 2016	Representational	0.523	0.015	1.031	0.044
Lindner et al. 2018	Representational	0.352	-0.140	0.844	0.161
Ott et al. 2018	Representational	0.315	-0.251	0.881	0.276
Sap et al. 2012	Representational	0.273	-0.277	0.823	0.331
Ogren et al. 2017	Representational	0.208	-0.437	0.852	0.528
Lindner et al. 2017b	Representational	0.126	0.028	0.224	0.012
Dewolf et al. 2014	Representational	0.089	-0.106	0.285	0.370
Ott et al. 2018	Representational	0.086	-0.468	0.640	0.761
Solano-Flores et al. 2014	Representational	0.077	0.004	0.149	0.039
Dewolf et al. 2014	Representational	0.025	-0.231	0.282	0.847
Maries 2013	Representational	-0.619	-1.081	-0.156	0.009
		0.235	0.100	0.370	0.001
Garcia-Retamero et al. 2010	Organizational	1.237	0.841	1.632	0.000
Whitley et al. 2006	Organizational	1.109	0.370	1.849	0.003
Garcia-Retamero and Hoffrage 2013	Organizational	0.995	0.664	1.326	0.000
Gick and Holyoak 1983	Organizational	0.445	0.124	0.766	0.007
Beveridge and Parkins 1987	Organizational	0.371	-0.050	0.793	0.084
Beveridge and Parkins 1987	Organizational	0.283	-0.080	0.647	0.127
Brase 2008	Organizational	0.100	-0.182	0.382	0.487
Gick and Holyoak 1983	Organizational	-0.433	-1.092	0.225	0.197
		0.516	0.173	0.860	0.003
Goolkasian 1996	Informational	0.837	0.416	1.259	0.000
Goolkasian 1996	Informational	0.364	-0.225	0.952	0.226
Sap et al. 2012	Informational	0.301	-0.228	0.830	0.264
Hoogland et al. 2018a	Informational	0.094	0.083	0.105	0.000
Sap et al. 2012	Informational	0.080	-0.433	0.593	0.760
Garrett 2008	Informational	0.077	-0.407	0.561	0.755
Lin et al. 2013; Yang and Huang 2004	Informational	-0.721	-0.808	-0.635	0.000
		0.124	-0.309	0.557	0.575
Ramjan 2011	Multiple	0.981	0.881	1.082	0.000
Sap et al. 2012	Multiple	0.620	0.087	1.154	0.023
Ginther 2001	Multiple	0.068	-0.087	0.222	0.389
		0.555	-0.161	1.270	0.129

Fig. 6 Forest plot of the effect sizes (Hedges's *g*) of response accuracy across different function types of pictures

Hu et al., 2021

Categorization of functionality

Item	Non-coherent representational picture

Carl has 5 friends and George has 6 friends. Carl and George decide to give a party together. They invite all their friends. All friends are present. How many friends are at the party?

Lindner, 2020

Decorative Picture

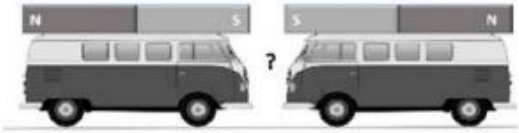
Anton has two identical toy buses and two equally sized rod magnets. He sets up the buses opposite each other and places one of the two rod magnets on each of the buses. The south pole of the magnet on one bus points to the south pole of the magnet on the other bus. Anton pushes the buses close together and then lets them go.

What happens to the toy buses when Anton lets them go?

A The buses attract each other.
B The buses attract each other first and then repulse each other.

Coherent representational picture

Anton has two identical toy buses and two equally sized rod magnets. He sets up the buses opposite each other and places one of the two rod magnets on each of the buses. The south pole of the magnet on one bus points to the south pole of the magnet on the other bus. Anton pushes the buses close together and then lets them go.



What happens to the toy buses when Anton lets them go?

A The buses attract each other.
B The buses attract each other first and then repulse each other.

This project

- Multimedia effect in testing
- Coherence as a moderator?
- *Universal effects? Or effects on item-level?*

Methodology

	Item 1	Item 2	Etc.
Student #1	Text-only	Text-Picture	Text-only
Student #2	Text-Picture	Text-only	Text-Picture
Etc.	…	…	…

Experiment *In Duplo*

Two different tests on different subjects with coherent, and non-coherent representational pictures: test A and test B

Tests were answered by two different groups of participants:

- Test A 1st year students
- Test B 2nd year students

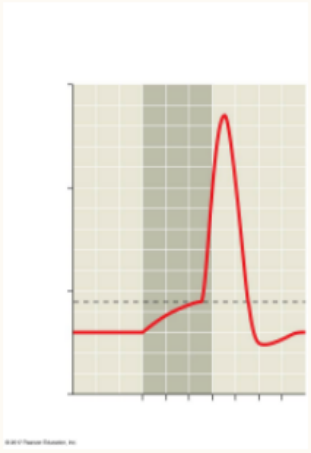
Methodology

- Participants
  - Students from teacher training institute (Fontys and HRO)
    - Test A : n = 88 (28 males), M_{age} 20,8 yrs
    - Test B : n = 70 (29 males), M_{age} 22,1 yrs

Methodology

- Testvision as test taking application

Een actiepotentiaal wordt gevolgd door een absoluut refractaire periode, waarin geen nieuwe actiepotentiaal mogelijk is in het desbetreffende neuron.



Wat is een van de oorzaken voor deze refractaire periode?

- ○ Het duurt enige tijd voordat de energie voor een actiepotentiaal geleverd kan worden.
- ○ Het duurt enige tijd voordat de ionenverdeling voldoende is hersteld.
- ○ Het duurt enige tijd voordat de neurotransmitter van de receptor loslaat.
- ○ Het duurt enige tijd voordat de natrium/kaliumpomp kan gaan werken.

Methodology

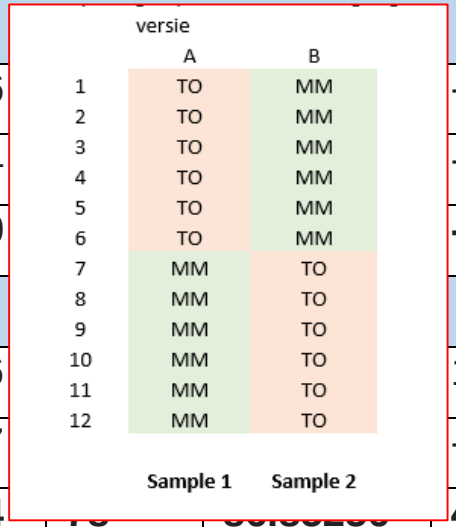
- After each item, students indicated subjective mental effort

In solving or studying the preceding problem I invested

1. very, very low mental effort
2. very low mental effort
3. low mental effort
4. rather low mental effort
5. neither low nor high mental effort
6. rather high mental effort
7. high mental effort
8. very high mental effort
9. very, very high mental effort

Results – test level

Item Set	n_TO	M_TO	n_MM	M_MM	delta	df	*t*	P	Cohen's d
Response accuracy									
Set 1	41	0.4674797	37	0.4234234	-0.044	76	0.738	0.463	0.167
Set 2	37	0.4144144	41	0.4065041	0.008	76	0.156	0.877	0.035
Total	**78**	**0.4423077**	**78**	**0.4145299**	**-0.028**	**154**	**0.712**	**0.478**	**0.161**
Subjective mental effort									
Set 1	36	5.925926			-0.435	70	1.789	0.078	0.422
Set 2	36	5.824074			-0.111	70	0.472	0.638	0.111
Total	**72**	**5.875000**			**-0.162**	**142**	**0.952**	**0.343**	**0.224**
Time-on-task									
Set 1	41	81.59756			12.812	76	1.859	0.067	0.422
Set 2	37	83.46847			-3.436	76	0.538	0.592	0.122
Total	**78**	**82.48504**			**4.368**	**154**	**0.923**	**0.358**	**0.209**



Categorization of functionality

Item	Non-coherent representational picture

Carl has 5 friends and George has 6 friends. Carl and George decide to give a party together. They invite all their friends. All friends are present. How many friends are at the party?

Lindner, 2020

Decorative Picture

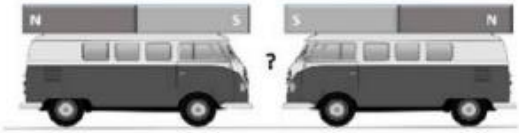
Anton has two identical toy buses and two equally sized rod magnets. He sets up the buses opposite each other and places one of the two rod magnets on each of the buses. The south pole of the magnet on one bus points to the south pole of the magnet on the other bus. Anton pushes the buses close together and then lets them go.

What happens to the toy buses when Anton lets them go?

A The buses attract each other.
B The buses attract each other first and then repulse each other.

Coherent representational picture

Anton has two identical toy buses and two equally sized rod magnets. He sets up the buses opposite each other and places one of the two rod magnets on each of the buses. The south pole of the magnet on one bus points to the south pole of the magnet on the other bus. Anton pushes the buses close together and then lets them go.



What happens to the toy buses when Anton lets them go?

A The buses attract each other.
B The buses attract each other first and then repulse each other.

Results – coherent pictures

Item Set	n_TO	M_TO	n_MM	M_MM	delta	df	t	p	Cohen's d
Response accuracy multimedia items with **coherent** representational pictures									
Set 1	41	0.4573171	37	0.4594595	0.002	76	0.032	0.974	0.007
Set 2	37	0.4414414	41	0.4878049	-0.046	76	0.677	0.500	0.154
Total	**78**	**0.4497863**	**78**	**0.4743590**	**0.025**	**154**	**0.518**	**0.605**	**0.118**
Subjective **mental effort** multimedia items with **coherent** representational pictures									
Set 1	37	6.162162	36	5.590278	-0.572	71	2.266	**0.027**	0.530
Set 2	36	5.564815	39	5.572650	-0.008	73	0.029	0.977	0.007
Total	**73**	**5.867580**	**75**	**5.581111**	**-0.286**	**146**	**1.524**	**0.130**	**0.357**
Time-on-task multimedia items with **coherent** representational pictures									
Set 1	41	85.03659	37	94.52703	9.490	76	1.231	0.222	0.279
Set 2	37	88.69369	41	86.30894	-2.385	76	0.322	0.749	0.073
Total	**78**	**86.77137**	**78**	**90.20727**	**3.436**	**154**	**0.644**	**0.520**	**0.146**

Results – non-coherent pictures

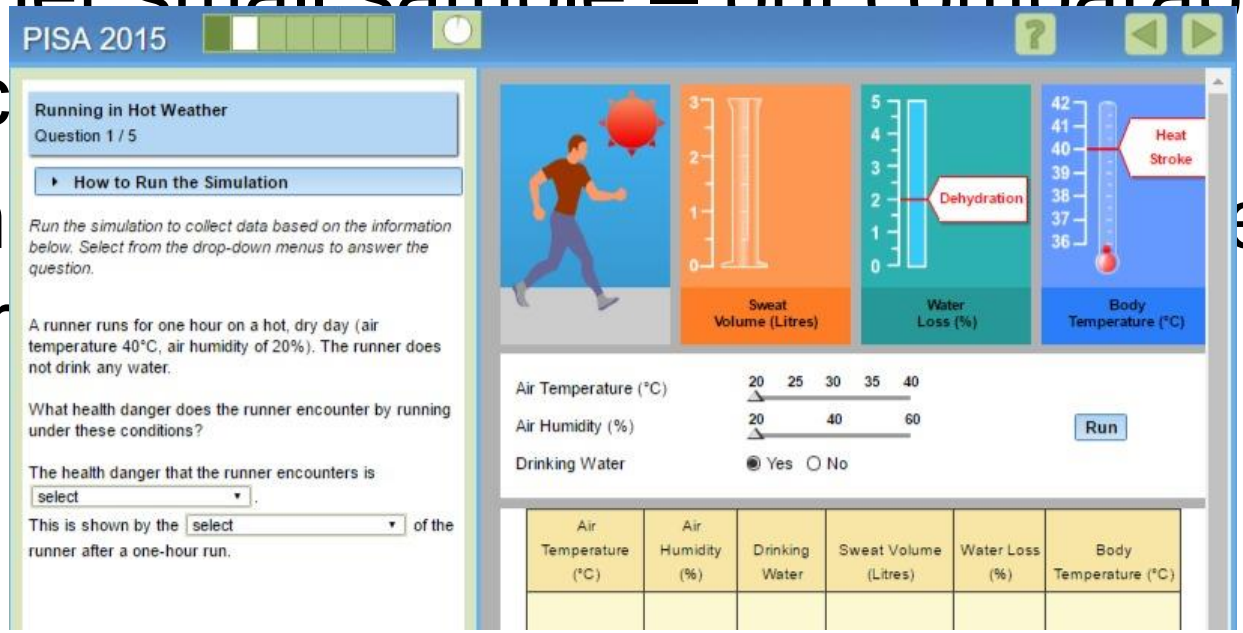
Item Set	n_TO	M_TO	n_MM	M_MM	delta	df	t	p	Cohen's d
Response accuracy multimedia items with **non-coherent** representational pictures									
Set 1	41	0.4878049	37	0.3513514	-0.136	76	1.606	0.113	0.364
Set 2	37	0.3873874	41	0.3252033	0.062	76	0.976	0.332	0.221
Total	**78**	**0.4401709**	**78**	**0.3376068**	**-0.103**	**154**	**1.934**	**0.055**	**0.439**
Subjective **mental effort** multimedia items with **non-coherent** representational pictures									
Set 1	39	5.346154	37	5.310811	-0.035	74	0.115	0.909	0.026
Set 2	37	6.054054	37	6.279279	-0.225	72	0.899	0.372	0.206
Total	**76**	**5.690790**	**74**	**5.795045**	**0.104**	**148**	**0.498**	**0.620**	**0.114**
Time-on-task multimedia items with **non-coherent** representational pictures									
Set 1	41	74.71951	37	94.17568	19.456	76	2.267	**0.026**	0.514
Set 2	37	78.24324	41	73.75610	-4.487	76	0.621	0.537	0.141
Total	**78**	**76.39103**	**78**	**83.44231**	**7.051**	**154**	**1.239**	**0.217**	**0.281**

Conclusion

- Contrary to expectation: No clear evidence for the multimedia effect in testing
  - No enhanced response accuracy
- Coherence seems to moderate some of the outcomes:
  - Some effects on mental effort for coherent representational pictures
  - Negative effect on response accuracy and time-on-task for non-coherent representational pictures
- Multimedia effects differ between individual items

Discussion

- Why no multimedia effect in testing?
  - Influence of (semi)authentic situation?
  - Influence of prior knowledge?
- Rather small sample – but comparable outc
- Non e distr



Thank you

- Contact: jorik.arts@fontys.nl

Literature

- Dewolf, T., Van Dooren, W., Hermens, F., & Verschaffel, L. (2015). Do students attend to representational illustrations of non-standard mathematical word problems, and, if so, how helpful are they? Instructional Science, 43(1), 147-171. doi:10.1007/s11251-014-9332-7
- Dirkx, K. J. H., Skuballa, I. T, Manastirean-Zijlstra, C. S., & Jarodzka, H. (2021). Designing computer-based tests: Design guidelines from multimedia learning studied with eye tracking. Instructional Science, 49(5), 589–605. https://doi.org/10.1007/s11251-021-09542-9
- Hu, L., Chen, G., Li, P., & Huang, J. (2021). Multimedia effect in problem solving: A meta-analysis. In (Vol. 33, pp. 1717-1747): Springer.
- Kirschner, P. A., Park, B., Malone, S., & Jarodzka, H. (2017). Toward a cognitive theory of multimedia assessment (CTMMA). In M. J. Spector, B. B. Lockee, & M. D. Childress (Eds.), Learning, Design, and Technology. An International Compendium of Theory, Research, Practice, and Policy (pp. 1-23). Springer. https://doi.org/10.1007/978-3-319-17727-4_53-1
- Lindner, M. A. (2020). Representational and decorative pictures in science and mathematics tests: Do they make a difference? Learning and Instruction, 68, 101345. doi:https://doi.org/10.1016/j.learninstruc.2020.101345
- Lindner, M. A., Ihme, J. M., Saß, S., & Köller, O. (2018). How representational pictures enhance students' performance and test-taking pleasure in low-stakes assessment. European Journal of Psychological Assessment, 34(6), 376.
- Saß, S., Wittwer, J., Senkbeil, M., & Köller, O. (2012). Pictures in Test Items: Effects on Response Time and Response Correctness. Applied Cognitive Psychology, 26(1), 70-81. doi:https://doi.org/10.1002/acp.1798

Effects on item-level

Hoogland et al., 2018

Table 3. The difference in students' performance on task level.

Item	N		M (SE)		*t* test
	version A	version B	version A	version B	*P* (|*T*|>|*t*|)
1	15,878	15,964	.716 (.004)	.720 (.004)	.424
2	15,986	15,856	.525 (.004)	.483 (.004)	.000***
3	15,785	16,057	.314 (.004)	.290 (.004)	.000***
4	15,835	16,007	.826 (.003)	.833 (.003)	.131
5	16,038	15,804	.720 (.004)	.828 (.003)	.000***
6	15,775	16,067	.631 (.004)	.640 (.004)	.102
7	16,065	15,777	.404 (.004)	.416 (.004)	.042
8	16,298	15,544	.303 (.004)	.299 (.004)	.420
9	16,069	15,773	.221 (.003)	.213 (.003)	.085
10	15,882	15,960	.495 (.004)	.525 (.004)	.000***
11	15,850	15,992	.145 (.003)	.310 (.004)	.000***
12	15,871	15,971	.466 (.004)	.438 (.004)	.000***
13	15,931	15,911	.619 (.004)	.641 (.004)	.000***
14	15,889	15,953	.040 (.002)	.046 (.002)	.080
15	15,793	16,049	.394 (.004)	.388 (.004)	.264
16	15,921	15,921	.803 (.003)	.815 (.003)	.005
17	15,986	15,856	.803 (.003)	.787 (.003)	.000***
18	15,847	15,995	.153 (.003)	.168 (.003)	.000***
19	15,932	15,910	.247 (.003)	.284 (.004)	.000***
20	15,925	15,917	.130 (.003)	.164 (.003)	.000***
21	16,044	15,798	.188 (.003)	.256 (.003)	.000***

Note: N is number of items tested. M is mean score on items (with standard error in parentheses) $P(|T|>|t|)$ is result of *t* test, unpaired, unequal with hypothesis that difference in score is 0; ***$p < .001$. Version A is the word problem; Version B is the image rich numeracy problem; Significant results are underlined.

Results – individual items

Response accuracy

Item	Δ	χ^2	Sig
7 – C	-2,0	,048	,826
8 – NC	**-19,0**	**3,106**	**,078**
9 – NC	-8,0	,496	,481

Time-on-task

Item	Δ	χ^2	Sig
7 – C	8,92	-,953 (76)	,344
8 – NC	13,98	-1,380 (76)	,172
9 – NC	**24,93**	**-2,118 (76)**	**,037**

Results – individual items

Mental Effort

Item	Δ	χ^2	Sig
7 – C	**-0,63**	**1,885 (75)**	**,063**
8 – NC	-0,06	,170 (74)	,866
9 – NC	-0,03	,075 (75)	,941