Table 2. Reaction times in milliseconds for the baseline and dual tasks on the narrative, map, and decision-making tasks

			Base	line		Sim	pple	Complex			
	N	M	SD	95% CI	M	SD	95% CI	M	SD	95% CI	
Narrativ	e										
NS	24	347.05	48.99	[326.81, 367.29]	480.67	53.89	[452.77, 508.56]	513.08	90.93	[476.51, 549.66]	
ESL	23	326.45	48.21	[306.04, 346.86]	555.85	79.88	[527.36, 584.35]	560.11	86.86	[522.75, 597.48]	
Total	47	336.97	49.19	[322.47, 351.03]	517.46	77.12	[322.47, 351.03]	536.10	91.15	[322.47, 351.03]	
Map											
NS	24	347.05	48.98	[326.81, 367.29]	570.53	107.53	[527.19, 613.86]	547.77	70.18	[517.23, 578.32]	
ESL	23	326.45	48.21	[306.04, 346.86]	578.95	103.14	[534.68, 523.22]	587.84	78.36	[556.65, 619.04]	
Total	47	336.97	49.19	[322.47, 351.03]	574.65	104.34	[543.76, 605.71]	567.38	76.22	[545.98, 589.64]	
Decision	n-maki	ng								_	
NS	24	347.05	48.98	[326.81, 367.29]	578.88	89.94	[544.92, 612.84]	543.70	77.61	[508.35, 579.06]	
ESL	23	323.04	49.51	[302.36, 343.72]	552.99	74.15	[518.30, 587.69]	569.46	93.95	[533.35, 605.57]	
Total	47	335.05	50.19	[320.58, 349.52]	566.21	82.74	[541.67, 590.21]	556.58	86.03	[531.31, 581.85]	

Table 3. Self-ratings of perceived mental effort and task difficulty for the narrative, map, and decision-making tasks

				Perceived m	iental e		Perceived task difficulty							
			Sim	ple	Complex				Sim	ple		Complex		
	N	M	SD	95% CI	M	SD	95% CI	M	SD	95% CI	M	SD	95% CI	
Narrative	2													
NS	24	4.42	1.95	[3.69, 5.14]	4.96	1.52	[4.24, 5.67]	4.17	1.83	[3.50, 4.83]	4.04	1.71	[3.34, 4.75]	
ESL	24	4.08	1.56	[3.36, 4.81]	4.88	1.94	[4.16, 5.59]	3.79	1.38	[3.12, 4.46]	4.46	1.72	[3.76, 5.16]	
Total	48	4.25	1.75	[3.74, 4.76]	4.92	1.72	[4.41, 5.42]	3.98	1.62	[3.51, 4.45]	4.25	1.71	[3.75, 4.75]	
Map														
NS	24	5.08	1.84	[4.33, 5.84]	6.25	1.62	[5.52, 6.98]	5.54	1.38	[4.92, 6.16]	6.33	1.40	[5.66, 7.00]	
ESL	24	4.96	1.85	[4.20, 5.72]	5.88	1.92	[5.15, 6.61]	4.46	1.61	[3.84, 5.08]	6.08	1.82	[5.42, 6.75]	
Total	48	5.02	1.83	[4.48, 5.56]	6.06	1.77	[5.55, 6.58]	5.00	1.58	[4.56, 5.44]	6.21	1.61	[5.74, 6.68]	
Decision-1	making	3												
NS	24	6.00	1.82	[5.34, 6.66]	6.54	1.41	[5.91, 7.17]	5.58	1.74	[4.91, 6.26]	6.29	1.57	[5.63, 6.95]	
ESL	24	5.29	1.37	[4.63, 5.95]	6.17	1.63	[5.54, 6.79]	5.17	1.52	[4.94, 5.84]	6.29	1.63	[5.63, 6.95]	
Total	48	5.65	1.63	[5.18, 6.11]	6.35	1.53	[5.91, 6.80]	5.38	1.63	[4.90, 5.85]	6.29	1.58	[5.83, 6.78]	

The values represent ratings on a 9-point Likert scale, 1 = lowest, 9 = highest.

Table 4. Expert judgments of mental effort and task difficulty for the narrative, map, and decision-making tasks

			J	udged/Expecte	d Menta	al effort	Judged/Expected Task difficulty						
			Sim	ple		Comp	olex		Simp	ole	Complex		
	N	M	SD	95% CI	M	SD	95% CI	M	SD	95% CI	M	SD	95% CI
Narrative	61	4.46	1.88	[3.97,4.92]	5.87	1.81	[5.43,6.31]	3.95	1.99	[3.43,4.46]	5.33	2.05	[4.84,5.79]
Map	61	5.28	2.03	[4.74,5.79]	7.05	1.80	[6.59,7.49]	5.30	2.10	[4.75,5.82]	7.13	1.68	[6.70,7.54]
Decision	61	5.75	1.86	[5.25,6.20]	7.37	1.31	[7.02,7.69]	5.64	1.92	[5.13,6.10]	7.15	1.35	[6.80,7.46]

The values represent ratings on a 9-point Likert scale, 1 = lowest, 9 = highest.

Table 5. Teacher comments on sources of difficulty/mental effort for narrative task

-	N ²	% ²	Example
Category/subcategory Mara reasoning required	52	85%	Example
More reasoning required Interpretation needed	32 16	26%	on the second one, the students would not only explain what is seen in the pictures, but they
interpretation needed	10	2070	would also try to explain what is implicit in the story
			would also if y to explain what is implicit in the story
Lack of clear storyline	15	25%	the second set was much more complex, with not as clear a narrative
Each of clear story line	10	2370	the second set was mach more complex, with not as clear a narrative
Lack of predictable storyline	5	8%	[in narrative 1] the changes between the various phases are less common and predictable than
1			in narrative 2
Intentional reasoning needed	5	8%	in the second task it seems that the student would need to explain what happens in the
			pictures adding how the person felt and why he did so students would have to put on the
			other person's shoes and explain what he did, why and the consequences
Open-ended	5	8%	in the second story there is a somewhat open end
~			
Creativity required	4	7%	the first is less creatively demanding the second is creatively more stimulating thus more
			challenging and demanding deeper cognitive processes
Causal reasoning required	2	3%	Fin the good day I course and congruence has to be taken into account
Causal reasoning required	2	370	[in the second story] cause and consequence has to be taken into account
Greater number of elements	26	43%	
Greater number of elements	20	7370	
More characters	11	18%	the first task seems simpler as students only need to describe one person the second is
112010 011111100020		10,0	more complex. first of all, there are multiple people
			r · · · · · · · · · · · · · · · · · · ·
More actions	10	16%	the second task is more difficult as there are multiple actions to describe
More places	3	5%	task 2 has different locations
In general	2	3%	second story contains more elements

Greater linguistic demands	16	26%	
More varied lexis needed	13	21%	the second one more difficult because the range of vocabulary involved in it is richer
In general	2	3%	The language required to explain the actions in the second task appears to be a bit more complex.
Complex morphosyntax needed	1	2%	[in the second one] grammar and syntax being more difficult
Story not interesting	4	7%	while judging these two tasks I had my students in mind and I guess they would be more attracted by the second one the story is "funnier"
Lack of clarity in pictures	3	5%	some pictures not clear in first story
Sense of irony involved	1	2%	in second set of pictures the narrative itself requires perhaps a sense of irony
Culture-specific info included	1	2%	[first story] is also very culturally specific

¹Value in general categories may be higher than the sum of comments in the subcategories, as some teachers only referred to the more general category.

2N and % refer to the number and percentage of teachers who mentioned a certain category/subcategory.

Table 6. Teacher comments on sources of difficulty/mental effort for map task

Category/subcategory ¹	$\frac{N^2}{N^2}$	% ²	Example
Lack of clarity in visual input	42	69%	the first picture is more difficult to understand visually
Lack of clarity in map	17	28%	the first map is very confusing, the second map is nice and clear
3- versus 2-dimensional	15	25%	the second task requires the speaker to think in 3 dimensions
Greater number of elements	13	21%	the second task is pretty straight forward - just by glancing at the map you can see how to get to town. the first task confuses me! there are lots of elements, stairs, different levels, two different destinations, two buildings.
More difficult route	12	20%	the route is simple and not confusing because the Australian friend will only pass by the same place once. besides, the only way to get to the destinations is by walking for map 2, but not for map 1 which requires the Australian friend to walk, take the elevator and the escalator.
Greater linguistic demands	7	11%	second one is less difficult, because the range of prepositions and language structures involved in are at a lower level.
More difficult instructions	1	2%	for the first one I was able to understand the instructions straight away. for the second one I had to refer back several times to be clear about what was being asked of me (should you start from the same place every time? for example)
Cultural familiarity	1	2%	if for a course book, I'd be surprised if this [first task] got past the editors. it requires a knowledge of shopping malls

Value in general categories may be higher than the sum of comments in the subcategories, as some teachers only referred to the more general category.

²N and % refer to the number and percentage of teachers who mentioned a certain category/subcategory.

Table 7. Teacher comments on sources of difficulty/mental effort for decision-making task

Category/subcategory ¹	N^2	% ²	Example
More reasoning required	24	39%	the second task is more difficult students need to engage higher order thinking skills to complete it successfully
More difficult decisions to make	6	10%	the second situation seems to be more complex and that makes it more difficult to decide on actions and sequences
More creativity required	4	7%	more inventive answer [needed in] second task
Need to prioritize	1	2%	the second task is more complex it requires the student to prioritise
More value judgement needed	2	3%	I feel that the second situation is much harder for the student to describe they're not just asked to describe the situation but to make value judgments on who to save or at least the sequence in which to do it.
Greater number of elements	20	33%	the second task has more factors to consider when looking at the image
More people	11	18%	there seem to be more people in danger in the second one so it would require a more detailed description of how to save the people
More actions	2	3%	in the second picture there are too many contrasting actions to describe
Lack of clarity in visual prompt	6	10%	the second is more difficult the images are not as straightforward
Higher level of danger	8	13%	task 2 presents greater cognitive demands as the visual presents a more perilous context
More difficult solution Fewer resources	10 6	16% 10%	second one more difficult: in this situation, there are many different aspects to pay attention to (children playing in the roof, old people and/or with reduce mobility, pregnant woman with kids, wind direction, lift in fire and people trapped in it, no helicopter, just one fire-truck, no easy access to the extinguishers)

Wind	4	7%	fire-chief 2 is a lot more difficult than Fire-chief 1 because the wind blows toward the stairs, making it difficult for people to get out of the building.
Greater linguistic demands	7	11%	[in second task] the speaker will need to draw upon a wider range of linguistic resources to explain their decision
More varied lexis needed	4	7%	the second one would require more vocabulary (e.g. "stuck in the lift").

¹Value in general categories may be higher than the sum of comments in the subcategories, as some teachers only referred to the more general category.

²N and % refer to the number and percentage of teachers who mentioned a certain category/subcategory.