Latent Class Analysis – content knowledge test

A latent class analysis was conducted to gain more insight into the heterogeneity of the content knowledge test data. Using the package 'poLCA' in R (Linzer & Lewis, 2011), a 1, 2 and 3-class model were fitted for both versions of the content knowledge test. Fit indices used to determine model fit were the AIC and BIC, of which lower values are preferred. The fit indices are shown in table S3 and S4.

For both versions, the AIC-value was lowest for the 2-class model, while the BIC-value was lowest for the 1-class model. Both models for version 1 and 2 of the content knowledge test were inspected to determine which model was most suited, together with the reliability statistics per item. The class-conditional parameters on each indicator (item) per class for version 1 and 2 are shown in table S5, as well as which items were excluded after reliability analysis.

For the 2-class model of version 1, the predicted class membership was approximately equal with 58% for class 1. For 7 out 10 items, class 1 showed the highest percentages of probabilities for answering an item correctly, while class 2 showed the highest percentages of probabilities for answering an item incorrectly. Thus, class 1 could be identified as pupils with more conceptual knowledge on sound. The 2-class model also showed that the 5 items that were retained after the reliability analysis were most sufficient in distinguishing between class 1 and class 2. In other words, these items were more likely to be answered correctly by the more 'knowledgeable' group than the other items, where differences between the two classes were smaller. Compared to the 1-class model of version 1, this valuable information is missing. For example, in the 1-class model, item 1 and 2 have the same mean and probability of being answered correctly (.60). In the 2-class model, however, it is shown that item 1 does not distinguish well between class 1 and 2, while item 2 does. Therefore, the interpretation of the 2-class model seems more meaningful.

For version 2, class sizes differ more substantially, with a predicted class membership of 76% for class 1. The more knowledgeable class seems to be class 2 for this version, who shows the highest percentages of probabilities for answering an item correctly for 8 out of 10 items. The same pattern is found, however, meaning that items that were retained after reliability analysis seem to be more capable of distinguishing between class 1 and 2.

Further inspecting the items that were retained after reliability analysis (see supplementary material 'Content knowledge items'), provides some more insight into why these items may not distinguish well between knowledgeable and less knowledgeable students. A first observation is that all items that were retained are related to the experiments that students had conducted in class, corresponding to the idea that active processing is more effective in constructing new knowledge than passive listening (Vosniadou et al., 2001). Although some of the items that should be excluded were also related to the experiment students had conducted, some other difficulties with these items may have caused them to be suboptimal.

First, correctly answering item 1 in both versions demands an understanding that sound needs a medium to travel through and can pass through solids such as a table. Students in this age range (around 10 years old) often hold the presupposition of substantiality (i.e. that sound is made of a matter and cannot pass through solids) (Mazens & Lautrey, 2003), which would cause their answer on item 1 to be incorrect. Furthermore, the objective that sound needs a medium to travel was discussed in class (i.e., passive), but it was not an element that students were exposed to with the experiments (i.e. active). Second, while item 4 is also related to the objective that sound needs a medium to travel through, it is not constrained by the presupposition of substantiality, as this item can be answered correctly when assumed that sound is made of a matter and can't move through a bell jar. Third, item 3 involves the absorption/reflection of sound, for which both classes had high probabilities of answering correctly and was thus not good in distinguishing. Fourth, item 8 probably caused confusion because of the diversion answer possibilities. This item referred to difference in volume between the small sound box of a violin and the large sound box of a contrabass. The diversion answer possibilities included "sounds faster" and "sounds slower", which is confusing as a violin is usually played faster than a contrabass. Last, item 9 referred to sound waves, which is also not something that students were exposed to with the experiments.

Based on these results, it seems more reasonable to exclude the 5 items (item 1, 3, 4, 8, 9) that do not seem to distinguish well between knowledgeable and less knowledgeable pupils, and to retain the 5 items (item 2, 5, 6, 7, 10) that do.

Table S3

LCA content knowledge test version 1 (N = 67)

	Maximum log-likelihood	AIC	BIC	Likelihood ratio/deviance statistic	Chi-square goodness of fit
1-class	-421.16	862.32	884.37	296.57	959.47
2-class	-408.52	859.04	905.34	271.30	901.40
3-class	-400.01	864.02	934.57	254.27	875.50

Table S4

LCA content knowledge test version 2 (N = 68)

	Maximum log-likelihood	AIC	BIC	Likelihood ratio/deviance statistic	Chi-square goodness of fit
1-class	-412.85	845.70	867.90	279.57	933.29
2-class	-398.73	839.46	886.07	251.33	847.57
3-class	-393.30	850.61	921.63	240.48	668.74

Table S5

Class-conditional parameters on each indicator per class for version 1 and 2

Item	Class	Prob. incorrect	Prob. correct
1*	1	0.44	0.56
	2	0.38	0.62
2	1	0.17	0.83
	2	0.54	0.46
3*	1	0.17	0.83
	2	0.25	0.75
4*	1	0.08	0.92
	2	0.12	0.88
5	1	0.23	0.77
	2	0.70	0.30
6	1	0	1
	2	0.67	0.33
7	1	0.08	0.92
	2	0.53	0.47
8*	1	0.61	0.38
	2	0.56	0.44
9*	1	0.54	0.46
	2	0.63	0.37
10	1	0.30	0.70
	2	0.61	0.39

** Excluded after reliability analysis*

Item	Class	Prob. incorrect	Prob. correct
1*	1	0.45	0.55
	2	0.26	0.74
2	1	0.54	0.46
	2	0	1
3*	1	0.09	0.91
	2	0	1
4	1	0.19	0.81
	2	0	1
5	1	0.60	0.41
	2	0.06	0.94
6	1	0.72	0.28
	2	0.15	0.85
7	1	0.40	0.60
	2	0.03	0.97
8*	1	0.37	0.63
	2	0.65	0.35
9*	1	0.54	0.46
	2	0.79	0.21
10	1	0.49	0.51
	2	0.12	0.88

* *Excluded after reliability analysis*

References

Linzer, D. A., and Lewis, J. B. (2011). poLCA: an R package for polytomous variable latent class analysis. J. Stat. Softw. 42, 1–29.

Mazens, K., & Lautrey, J. (2003). Conceptual change in physics: Children's naive representations of sound. *Cognitive Development*, *18*(2), 159–176. https://doi.org/10.1016/S0885-2014(03)00018-2

Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction*, *11*(4–5), 381–419. https://doi.org/10.1016/S0959-4752(00)00038-4