Measuring Intrisic and Extraneous Cognitive Load in Elementary School Students Using Subjective Ratings and Smartpen Data

Sarah Malone[1], Kristin Altmeyer[1], Michael Barz[2,4], Luisa Lauer[3], Daniel Sonntag[4], Markus Peschel[3], & Roland Brünken[1]

[1]Department of Education, Saarland University, Saarbrücken, Germany

[2]German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany

[3]Department of Physics, Saarland University, Saarbrücken Germany

[4]Applied Artificial Intelligence, Oldenburg University, Oldenburg, Germany

New methods are constantly being developed to optimize and adapt cognitive load measurement to different contexts (Korbach et al., 2018). It is noteworthy, however, that research on cognitive load measurement in elementary school students is rare. Although there is evidence that they might be able to report their cognitive load (Ayres, 2006), there are also reasons to doubt the quality of children's self-reports (e.g., Chambers & Johnson, 2002). To avoid these issues, objective online-measures are promising. A novel approach – the use of smartpen data generated by natural use of a pen during task completion – seems particularly encouraging as these measures proved to be predictive of cognitive load in adults (e.g., Yu, Epps, & Chen, 2011). Moreover, Barz et al. (2020) demonstrated the predictive power of smartpen data for performance in children. The present research addressed two prevailing gaps in research on cognitive load assessment in elementary school students. We developed a subjective rating scale and investigated whether this instrument can provide valid measurements of ICL and ECL (Research Question 1). Moreover, we researched whether smartpen data can be used as a valid process measurement of cognitive load (Research Question 2).

Methods

In a within-subjects design, N=36 elementary school children (61% female) used the Neo Smartpen M1 to solve two types of standardized sketching tasks. First, they completed two versions of the Trail Making Test for Children (Reitan, 1992), which were expected to evoke different levels of ICL as they differed in complexity. Subsequently, subjects performed two versions of the subtest "drawing patterns" from the SON-R 51/2-17 (Snijders, Tellegen, & Laros, 2005), which required them to complete omitted parts of multiple reference patterns. The two versions were expected to trigger different levels of ECL as in one version, the reference patterns were provided on the front side of the drawing sheet (low ECL) and in the other version, on the back of the sheet (high ECL), causing split attention. Barz et al. (2020) had confirmed that in fact those versions that were intended to elicit more ICL resp. more ECL resulted in lower test performance in children. After each part of the two tasks, children filled out an adapted version of the Cognitive Load Scale by Klepsch et al. (2017). The adaptations consisted of simplifying the items, relating them strongly to each particular task, reducing the number of response categories, and labelling them verbally and graphically.

Results

Concerning Research Question 1, results revealed that the subjective ratings corresponded to the intended manipulation of ICL and ECL (d_{ICL}=.84; d_{ECL}=.85) and could be validated by performance (r_{ICL}=.23, r_{ECL}=-.48). Regarding Research Question 2, particular smartpen measures could be confirmed as indicators for both, ICL and ECL variation (e.g., number of single strokes; d >.48). Smoothness of pressure (d=.26) and velocity (d=.62) specifically indicated the variation in ICL. Accordingly, some smartpen measures were closely related to performance measures (e.g., number of single strokes in the low ICL (r=.62) and low ECL (r=.42) task versions).

Discussion

Results indicate that even young children can introspectively assess their cognitive load and differentiate it on an appropriate scale. Moreover, smartpen data has revealed to be a promising tool for cognitive load measurement, which is worth investigating further.

Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction*, *16*(5), 389-400. https://doi.org/10.1016/j.learninstruc.2006.09.001

Barz, B., Altmeyer, K., Malone, S., Lauer, S., & Sonntag, D. (2020). Digital Pen Features Predict Task Difficulty and User Performance of Cognitive Tests. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, Genoa, Italy. https://doi.org/10.1145/3340631.3394839

Chambers, C. T., & Johnston, C. (2002). Developmental differences in children's use of rating scales. *Journal of Pediatric Psychology*, *27*(1), 27-36. https://doi.org/10.1093/jpepsy/27.1.27

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Frontiers in Psychology*, *8*(1997). https://doi.org/10.3389/fpsyg.2017.01997

Korbach, A., Brünken, R. & Park, B. (2018). Differentiating Different Types of Cognitive Load: a Comparison of Different Measures. *Educational Psychology Review 30, 503–529.* https://doi.org/10.1007/s10648-017-9404-8

Reitan R. M. (1992), *Trail Making Test: Manual for Administration and Scoring.* Tucson, AZ: Reitan Neuropsychology Laboratory.

Snijders, P., Tellegen, P. J. & Laros, J. A. (2005). SON-R 5½-17. *Snijders-Oomen-Non-Verbal Intelligence Test.* 3rd edition. Hogrefe, Göttingen.

Yu, K., Epps, J., & Chen, F. (2011). Cognitive load evaluation of handwriting using stroke-level features. In *Proceedings of the 16th international conference on Intelligent user interfaces* (pp.423-426). https://doi.org/10.1145/1943403.1943481