The reliability of teacher evaluations of reading skills of primary school pupils

Wieke Harmsen; Martijn Bentum; Ferdy Hubers; Roeland van Hout; Catia Cucchiarini; Helmer Strik
Radboud University Nijmegen, Netherlands

Primary school teachers need to monitor the reading proficiency of their pupils. This can be assessed with a word reading task: A pupil reads a list of words aloud, while a teacher scores the read words as correct or incorrect (van Til et al., 2018). This is a laborious task, and the teacher could be aided by an automatic system that evaluates a pupil's reading proficiency (decoding skills) and can quickly assess which child needs remedial training. To develop such an automatic system, we need teacher evaluations (correct/incorrect) of pupils learning to read. Because the system can only be as good as the data it was trained on, these teacher evaluations should ideally have a high inter- and intra-rater reliability at both the pupil and the individual word level. It is therefore crucial to gain insight in the current practice of teachers with respect to the assessment of pupils' reading proficiency.
As part of the Dutch Automatic Reading Tutor (DART) project, pupils (aged 6 – 7) were recorded in their first year of learning to read. These recordings contain read aloud word lists of 24 words. For six recordings each read aloud word was rated as correct or incorrect by 51 teachers. Another set of 377 recordings of read aloud word lists was evaluated by two or three teachers. Based on these data we examined the inter-rater reliability of teachers on the pupil level, aggregating the binary evaluations of each teacher for a single recoding (correct/incorrect) into a percentage correct. Furthermore, we investigated the reliability at the word level by comparing the ratings from different teachers for a specific word. We present the findings, discuss how they can contribute to developing systems for automatically assessing reading proficiency and indicate possibilities for future research.

van Til, F. Kamphuis, J. Keuning, M. Gijsel, and A. de Wijs, "Wetenschappelijke Verantwoording LVS-toetsen AVI," Arnhem, 2018.