Exploring the cognitive validity of a computerized writing test using eye-tracking, keystroke logging and stimulated recalls

Benjamin Kremmel, Kathrin Eberharter & Elisa Guggenbichler, Universität Innsbruck

This paper presents a study that investigates the cognitive validity of the new computer-based Linguaskill Writing test in relation to test taker characteristics. It explores the cognitive processes engaged by a mixed-proficiency (CEFR A2-C1) sample of EFL learners (N=30) when responding to two writing tasks, and in how far they elicit a full model of L2 writing processes (Barkaoui, 2019; Révész, Michel, & Lee, 2017; Shaw and Weir, 2007). In addition, it aims to uncover the links between computer writing behaviour (i.e., fluency, pausing and revision behaviours of candidates) and the cognitive processes involved in L2 writing at different levels of test taker ability. For this, the study triangulates data from eye-tracking, keystroke logging, performance ratings, and text analyses of the written output of candidates using measures of text complexity, as well as stimulated recalls of 16 participants. Participants took a typing test as a baseline measure and a standalone receptive proficiency test in addition to responding to both parts of the Linguaskill Writing test. During the writing test, their eye movements and writing behaviour were recorded with a Tobii TX300 eye tracker and Inputlog 8.0. 16 of the participants produced stimulated recalls after each writing task, prompted by a replay of their eye movements together with the recording of their typing. The replays generated rich verbal recalls, which were transcribed and coded for evidence of cognitive processing as described in the literature. Selected metrics from the keystroke logging (pause duration and frequency) as well as eye tracking (fixation duration and number of visits) were used to explore relationships between task, test taker proficiency and writing outcome. The findings provide valuable information regarding test validation and task design in terms of the two prompts being successful in eliciting a substantial sample of cognitive processes and writing behaviours across proficiency levels.