A number of experimental methods have been used to elicit metalinguistic judgments about syllable division. The underlying assumption is that each of these methods taps the same syllabification strategies. However the syllabification literature is largely silent on the issue of whether this assumption is valid. Côté and Kharlamov (2011) addressed this issue by gathering data from Russian speakers who syllabified nonce words in several different experimental conditions. When the results of each task were compared with each other, significant correlations were only obtained between the results of a few of the conditions. This suggests that syllabification preferences are highly influenced by the particular method used to elicit them, which in turn could cast doubt on the validity of many syllabification studies. However, the experimental design and statistical methods employed by Côté and Kharlamov are somewhat problematic, which weakens any conclusions based on their outcome.

Nevertheless, the question of whether different experimental methods provide similar or disparate results is one that needs to be answered. Since a good deal of the experimental literature deals with the syllabification of English words, the present study focuses on intratask reliability in English. Subjects syllabified 120 English words using four written response and four oral response tasks. The results indicate that there were some significant differences between the experimental tasks. However, analysis of which variables influenced the syllabifications in each task (e.g. stress, vowel quality, cluster legality) showed a high degree of consistency, which suggests a good deal of intratask reliability. Reasons why these results differ from those of Côté and Kharlamov are discussed.