Sentence Combining Task: Scoring Methods for Differentiating Abilities

Michelle Magalski, B.S. and Nickola W. Nelson, Ph.D., CCC-SLP
Department of Speech Pathology and Audiology
Western Michigan University, Kalamazoo, MI

ABSTRACT
Sentence combining can serve as a standardized test for capturing written language growth and differentiating performance by students with typical language (TL) and language-learning disability (LLD). Results for 112 school-age students (58 TL and 54 LLD) showed medium to large effect sizes for the sentence-level measure of the number of content units per T-unit at different grade levels (grades 3-5, 6-8, and 9-12). The best discourse-level measure was percent content maintained, and the best word-level measure was percent of words without error.

METHODS
•Sentence combining task on the Test of Integrated Language and Literacy Skills (TILLS; beta research version; Nelson, Helm-Estabrooks, Hotz, & Plante, 2007)
•112 TL (n = 58) and LLD (n = 54) students
•LLD students were classified based on a score of less than 85 on CELF-4. Some in the LLD group (n = 16) were receiving language services but scored < 85 on CELF-4.
•The TILLS includes 4 stories to correspond to 4 age groups:
  - Story B: “The Principal’s Daughter,” Grades 3-5
  - Story C: “When the School Closed,” Grades 6-8
  - Story D: “The Building,” Grades 9-12
•The student first reads a short practice story aloud.
•The examiner then shows the example of how the story “The Little Dog” has been written to “sound better.”

RESULTS
The 3 “best” measures were selected based on which differentiated TL from LLD students at most of the 4 age levels. Effect sizes for nonparametric Mann-Whitney [1-tailed] tests are shown in Table 1 and p values are shown in the figures below. Reliability and Kendall’s Tau correlations with age-in-months for each measure by group are:
•Word Level: Percentage of words without spelling or grammatical error (99.5% intercorcer reliability) (Tau: TL = .277***; LLD = .039)
•Sentence Level: Number of content units per T-unit (96% intercorcer reliability) (Tau: TL = .411***; LLD = .295**)
•Discourse Level: Proportion of content units maintained from original story (100% intercorcer reliability) (Tau: TL = .035; LLD = .176*)
Significant at the .05 level (1-tailed) **Significant at the .01 level (1-tailed)

CONCLUSIONS BY GRADE
Strengths of research design: matched groups, age range 6-18, evidence for comparing written language to norms
Limitations of research design: small sample sizes especially at age extremes, scoring of content/T-unit requires specialized knowledge (T-unit division)
Conclusions regarding significant differences between groups:
•Grades 1-2: no measure showed significance differences
•Grades 3-5: percent words without error, content units per T-unit, and proportion of content maintained
•Grades 6-8: content units per T-unit and proportion of content maintained
•Grades 9-12: content units per T-unit

REFERENCES/ACKNOWLEDGMENTS
Thanks to research committee of this master’s research thesis, Dr. Yvette Hyter, Dr. Nickola Nelson, and Dr. Amy Curtis.

ACKNOWLEDGMENTS
Michelle Magalski, B.S. and Nickola W. Nelson, Ph.D., CCC-SLP

DATA SOURCES

ACKNOWLEDGMENTS
Michelle Magalski, B.S. and Nickola W. Nelson, Ph.D., CCC-SLP