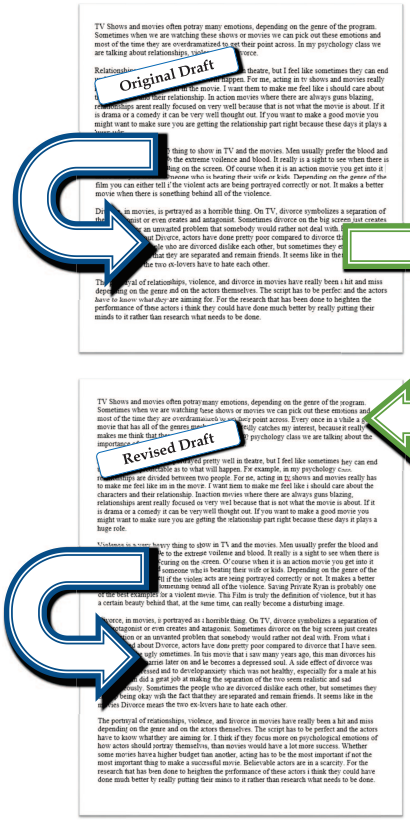
Toward Revision-Sensitive Feedback in Automated Writing Evaluation

Rod D. Roscoe Rod.Roscoe@asu.edu
Matthew E. Jacovina Matthew.Jacovina@asu.edu
Laura K. Allen LauraKAllen@asu.edu
Adam C. Johnson acjohn17@asu.edu
Danielle S. McNamara Danielle.McNamara@asu.edu
Arizona State University

Background

- Automated writing evaluation (AWE) systems provide summative scores and formative feedback on students' writing
- Natural language processing (NLP) tools extract linguistic, syntactic, semantic, and rhetorical features of text
- Statistical models leverage relationships between text data and quality to drive automated scoring and feedback
- Typically assesses discrete drafts, which may limit feedback to writing products rather than writing processes (e.g., revising)

Standard Approach



Revision-Sensitive Approach

- AWE tools "ignore" changes from draft to draft
- Can we automatically detect revisions in a way that guides useful feedback?
- How does automated revision detection map on to students' revising behaviors (e.g., adding, deleting, reorganizing, substituting)?

Method and Results

- 85 high school students wrote and revised one essay on "fame" via the Writing Pal ITS
- Coh-Metrix and related NLP tools used to extract a variety of text properties
- Essay revisions were annotated by human coders (κ = .81-.92)
- Examined linguistic changes from original to revised drafts and their relationship to changes in essay score
- Examined relationships between linguistic changes and annotated revising behaviors

	Linguistic Change	*r* with Score	*r* with Revision Actions			
Measure	*t*(84)	*r*(84)	Add	Delete	Subst.	Reorg.
Basic						
Words	**6.24**[a]	.06	**.29**[b]	**-.36**[a]	-.18	-.10
Sentences	**4.33**[a]	-.09	**.37**[a]	-.18	-.16	.05
Lexical						
Diversity	-0.28	.17	.01	**.26**[c]	-.04	.07
Concreteness	0.83	**.34**[b]	.00	**.29**[b]	.08	.06
Familiarity	-0.74	-.01	-.04	**-.28**[c]	.15	-.09
Hypernymy	0.80	**.24**[c]	-.10	.11	.02	-.18
1st Person	**2.09**[c]	-.07	.04	-.11	.11	.07
2nd Person	-1.06	**-.22**[c]	-.09	-.03	-.05	-.04
3rd Person	-0.23	-.10	-.01	**-.26**[c]	-.07	.00
Cohesion						
Connectives	**1.67**[d]	.03	-.07	.16	.09	-.03
LSA Given/New	**2.98**[b]	.08	-.02	**-.32**[c]	-.07	-.07
LSA Sentences	0.58	**.24**[c]	-.20	-.09	.06	-.12
Deep Cohesion	**1.86**[d]	-.08	.07	**-.24**[c]	-.05	.04
Narrativity	0.52	.01	-.10	**-.25**[c]	.12	-.03

Note. [a]$p \leq .001$. [b]$p \leq .01$. [c]$p \leq .05$. [d]$p \leq .10$

Conclusions and Future Work

- Linguistic changes in essays from original to revised draft were detectable via NLP tools
- The most common linguistic changes were not well aligned to essay quality; indicates that students were not skilled revisers
- Most revision actions also had little impact on the quality or linguistic profile of the essay
- Most meaningful actions were deletions
- Current NLP-based analyses may not be sufficient to capture revising behaviors (and to drive feedback on those behaviors)
- Keylogging tools may be a useful resource

Supported in part by the Institute of Education Sciences (R305A120707).