

Automated Writing Evaluation: College Student Perceptions and Future Intentions

**Rod D. Roscoe¹, Joshua Wilson²,
Adam C. Johnson¹, & Chris R. Mayra¹**

¹Arizona State University

²University of Delaware

Automated Writing Evaluation

- **Natural language processing (NLP)** tools detect linguistic, syntactic, structural, semantic, and rhetorical properties of text
- NLP data can be **statistically modeled** to emulate or predict human judgments of writing quality or specific writing traits
- Statistical models (e.g., regression, DFA, machine learning) guide **scoring and feedback algorithms**



Dikli (2006), Shermis & Burstein (2013)

A Bit of Debate...

Proponents

Scale Up



Speed

Reduce Workload



**More Practice
with Feedback**



Critics

**Incomplete
Writing
Construct**



Formulaic Writing

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

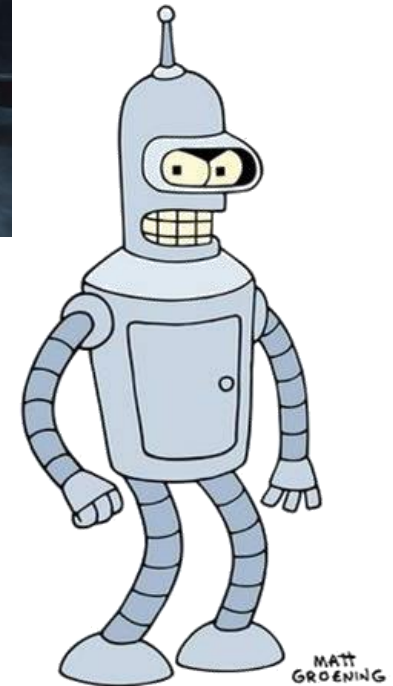


**Inhuman
Audience**

A Bit of Debate...

Proponents

Critics



Consequences of Perceptions?

- **Technology acceptance** may be mediated by perceptions of usefulness and ease of use
Vinkatesh & Davis (2000)
- Beliefs and perceptions about technology can introduce **barriers to implementation**
Ertmer (1999); Ertmer et al. (2012); Koehler & Mishra (2009)
- Educators' beliefs and attitudes can **influence classroom culture** and student behavior
Li et al. (2015); Webb et al. (2006); Yeager & Dweck (2012)

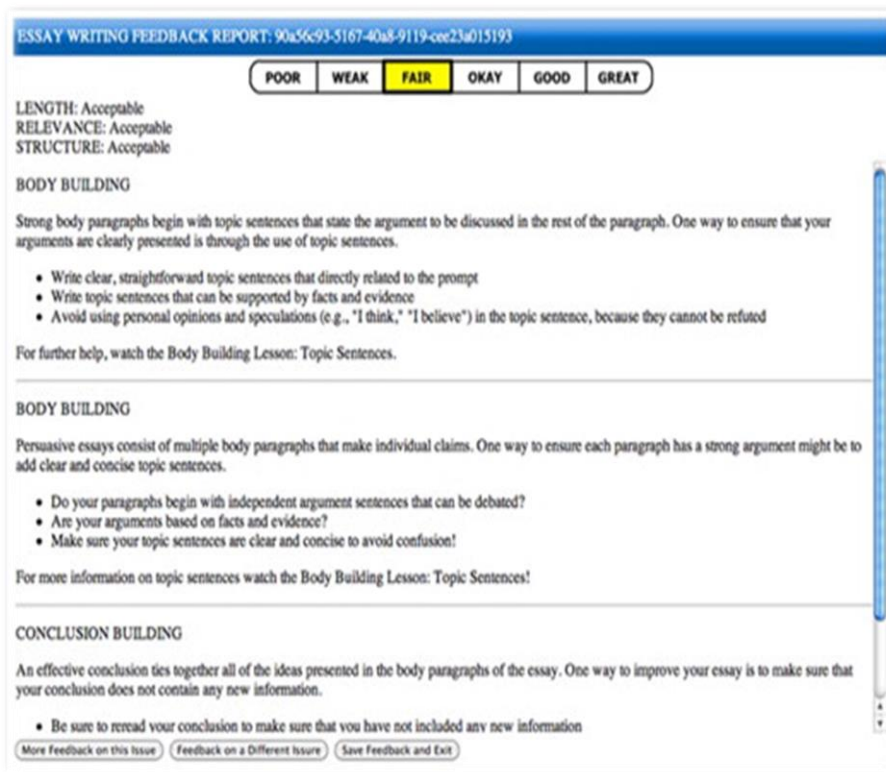


Current Study

- Explore college student perceptions of AWE (using The Writing Pal tutoring system)
 - **expectations** about scoring and feedback
 - **immediate perceptions** of scores and feedback received
 - **change in perceptions** (better or worse than expected)
- Impact of perceptions on revising (in the paper)
- Impact of perceptions on **future intentions**

The Writing Pal

- **Intelligent tutor for writing strategies.** Includes educational games and formative feedback on student writing (AWE).



ESSAY WRITING FEEDBACK REPORT: 90a56c93-5167-40a8-9119-ccc23a015193

POOR WEAK **FAIR** OKAY GOOD GREAT

LENGTH: Acceptable
RELEVANCE: Acceptable
STRUCTURE: Acceptable

BODY BUILDING

Strong body paragraphs begin with topic sentences that state the argument to be discussed in the rest of the paragraph. One way to ensure that your arguments are clearly presented is through the use of topic sentences.

- Write clear, straightforward topic sentences that directly related to the prompt
- Write topic sentences that can be supported by facts and evidence
- Avoid using personal opinions and speculations (e.g., "I think," "I believe") in the topic sentence, because they cannot be refuted

For further help, watch the Body Building Lesson: Topic Sentences.

BODY BUILDING

Persuasive essays consist of multiple body paragraphs that make individual claims. One way to ensure each paragraph has a strong argument might be to add clear and concise topic sentences.

- Do your paragraphs begin with independent argument sentences that can be debated?
- Are your arguments based on facts and evidence?
- Make sure your topic sentences are clear and concise to avoid confusion!

For more information on topic sentences watch the Body Building Lesson: Topic Sentences!

CONCLUSION BUILDING

An effective conclusion ties together all of the ideas presented in the body paragraphs of the essay. One way to improve your essay is to make sure that your conclusion does not contain any new information.

- Be sure to reread your conclusion to make sure that you have not included any new information

[More Feedback on this Issue](#) [Feedback on a Different Issue](#) [Save Feedback and Exit](#)

- **Feedback topics include:**
 - length/elaboration
 - structure
 - introductions
 - bodies
 - conclusions
 - paraphrasing
 - cohesion
 - revising

Current Study

- 110 undergraduates wrote (20 min.) an essay on “psychology in the media” and revised (10 min.) after receiving a score and feedback from W-Pal
- **Presentation Conditions** (no deception)
 - manipulated whether **scoring system** was presented as “well tested” vs. “in progress”
 - manipulated whether **feedback system** was presented as “well tested” vs. “in progress”



Perceptions

- **Expectations of scoring and feedback**
 - *after* system was introduced
 - *before* any writing or revising
- **Immediate perceptions of scoring and feedback**
 - *after* writing, receiving feedback, and revising
 - i.e., the “full experience”
- **Change in perceptions of scoring and feedback**
 - at the *end* of the study
 - whether final perceptions “better” or “worse”

Future Intentions

- “Would you **use this software again** to help you improve your writing?”
 - “Yes” or “No”
- “Would you **recommend this software** to a friend who needed writing help?”
 - “Yes” or “No”

Effects of Presentation

Mean Ratings of Expectations, Immediate Perceptions, and Final Perceptions by Condition

	Advertised Scoring and Feedback Quality Conditions				Main Effects	
	Strong Scoring, Strong Feedback (n = 28)	Strong Scoring, Weak Feedback (n = 29)	Weak Scoring, Strong Feedback (n = 28)	Weak Scoring, Weak Feedback (n = 29)	Presented Scoring Accuracy <i>F</i> (1, 106)	Presented Feedback Quality <i>F</i> (1, 106)
Initial Expectations	3.8 (0.8)	3.8 (0.7)	3.8 (0.8)	3.8 (0.8)	7.86 ^b	2.32
Immediate Perceptions	3.8 (0.6)	3.1 (0.6)	3.2 (0.9)	3.1 (0.8)	1.42	< 1.00
Final Perceptions	3.8 (1.0)	3.8 (0.8)	3.8 (1.1)	3.8 (1.1)	1.13	< 1.00

Presentation only slightly influenced expectations, immediate perceptions, or final perceptions. Perhaps "low dosage" or lack of authority?

Note. ^a *F* composites computed by averaging individual feedback ratings (see Method). Perceptual Change ratings were given on a scale of -2 to +2 and are not difference scores.

Predicting Scoring Perceptions

Linear Regression Predicting Perceptual Change for Scoring Accuracy

Predictor	Coefficients			Model Fit		
	β	t	p	R^2	F	p
Pres. Scoring Accuracy	-0.08	-0.98	.332	.37	9.99	< .001
Pres. Feedback Quality	0.03	0.37	.710			
Exp. Scoring Accuracy	0.24	2.29	.024			
Exp. Feedback Quality	-0.38	-3.60	< .001			
Imm. Scoring Accuracy	0.47	5.54	< .001			
Imm. Feedback Quality	0.26	2.96	.004			

Note. "Pres." refers to presented system capabilities (dichotomous coding, weak

Expectations and immediate perceptions influenced judgments of the scoring system as "better" or "worse" than expected.

Predicting Feedback Perceptions

Linear Regression Predicting Perceptual Change for Feedback Quality

Predictor	Coefficients			Model Fit		
	β	t	p	R^2	F	p
Pres. Scoring Accuracy	-0.07	-0.88	.383	.44	13.26	< .001
Pres. Feedback Quality	0.10	1.22	.224			
Exp. Scoring Accuracy	-0.01	-0.07	.947			
Exp. Feedback Quality	-0.06	-0.57	.570			
Imm. Scoring Accuracy	0.18	2.25	.027			
Imm. Feedback Quality	0.56	6.69	< .001			

Note. "Pres." refers to presented system capabilities (dichotomous coding, weak =

Only immediate perceptions influenced judgments of the feedback system as "better" or "worse" than expected.

Willingness to Use Again

Logistic Regression Predicting Willingness to Use W-Pal in the Future

Predictor	Coefficients				
	<i>B</i>	<i>SE</i>	<i>Wald</i>	<i>p</i>	e^B
Exp. Scoring Accuracy	0.49	0.46	1.09	.296	1.63
Exp. Feedback Quality	0.97	0.63	2.37	.124	2.65
Imm. Scoring Accuracy	0.88	0.61	2.10	.147	2.42
Imm. Feedback Quality	0.42	0.62	0.46	.499	1.52
Scoring Accuracy Change	0.39	0.49	0.62	.432	1.48
Feedback Quality Change	2.10	0.60	12.38	< .001	8.18

Note. "Exp." refers to students' expectations, and "Imm." refers to students' immediate perceptions. All Feedback Quality ratings are composites computed by averaging

Willingness to use again predicted by perceiving the feedback as "better than expected."

Willingness to Recommend

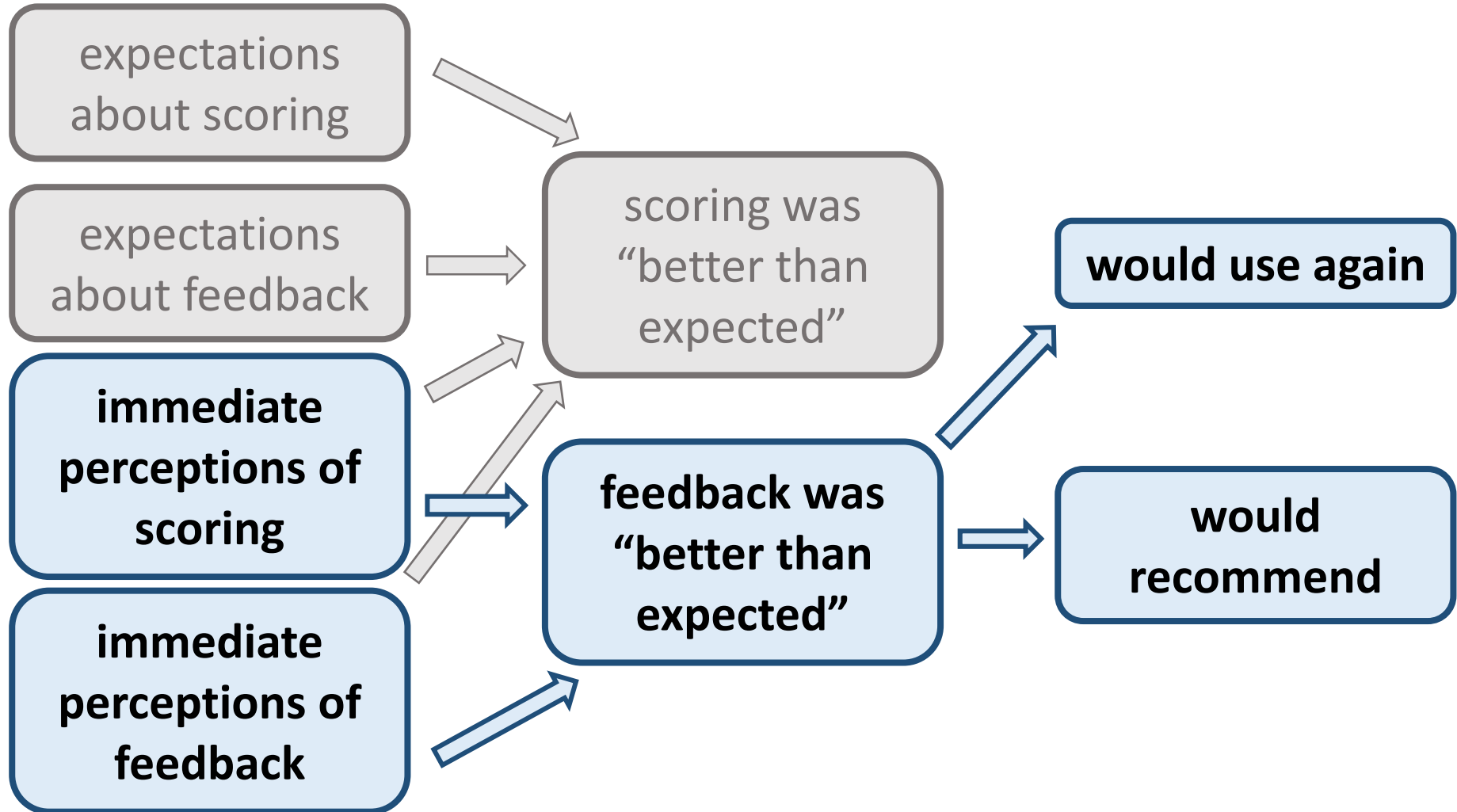
Logistic Regression Predicting Willingness to Recommend W-Pal to a Friend

Predictor	Coefficients				
	<i>B</i>	<i>SE</i>	<i>Wald</i>	<i>p</i>	e^B
Exp. Scoring Accuracy	1.16	0.56	1.25	.039	3.17
Exp. Feedback Quality	0.29	0.68	0.18	.673	1.33
Imm. Scoring Accuracy	1.37	0.71	3.67	.055	3.93
Imm. Feedback Quality	1.12	0.71	2.50	.114	3.08
Scoring Accuracy Change	-0.27	0.58	0.22	.636	0.76
Feedback Quality Change	2.84	0.78	13.37	< .001	17.12

Note. "Exp." refers to students' expectations, and "Imm." refers to students' immediate feedback. The dependent variable is willingness to recommend, measured on a 5-point scale (1 = "not at all willing" to 5 = "very willing"). The independent variables are the average scores on the following items:

Willingness to recommend predicted by perceiving the feedback as "better than expected."

Informal "Path"



Winning them Over?

- Effective automated feedback is not just a “learning sciences” issue (e.g., principles of feedback)...

Hattie & Timperley (2007); Shute (2008)

- ... and not just a “computer science” issue (e.g., better NLP detection algorithms)...

Deane (2013); McNamara et al. (2015); Shermis & Burstein (2013)

- **... might also be a “user science” issue**
 - feedback perceptions, design, classroom integration
 - users’ beliefs about methods and appropriateness
 - **direct, positive user experiences are how these perceptions are formed, reinforced, or overturned**

Automation More Broadly

- Beyond automated feedback, what about overall **automation in educational technology**?
 - e.g., intelligent tutoring systems, pedagogical agents, teachable agents, learner modeling, grading, course assignments and placement, etc.
- How are educational technologies running afoul (or taking advantage) of users' beliefs and doubts about computers, automation, and AI?
- **What is the “effect size” of good design and HCI?**



Questions?

Roscoe, R. D., Wilson, J., Johnson, A. C., & Mayra, C. R. (2017). Presentation, expectations, and experience: Sources of student perceptions of automated writing evaluation. *Computers in Human Behavior, 70*, 207-221.