Characteristics of academic writing: Is what we can count all that counts?

Sara Cushing Weigle
Georgia State University
ALTE, May 2015
Overview

- Automated scoring of writing: history of counting things
- What we can count, and what that tells us about writing proficiency
- What we can’t count yet but is important to know
GRADE

ALL THE ESSAYS!
The beginnings of automated essay evaluation
Project Essay Grader (PEG)

- Developed by Ellis Page in 1966
- Uses regression analysis of surface features of essay to predict human scores
  - Fluency: Essay length
  - Sentence structure: Prepositions, relative pronouns, other parts of speech
  - Diction: Variation in word length
- Essay scores from PEG correlate with human raters as well as or better than pairs of human raters (.87)
- Disadvantage: Does not consider content; specific to each set of essays; no instructional feedback.
Latent Semantic Analysis

- Intelligent Essay Assessor (Thomas Landauer, 1997)
- Compares semantic content of words used in essays to model answer or text
- Based on words only, not word order or other syntactic features
- Useful for evaluating content-based expository writing (e.g., psychology) but not factual knowledge
- Can be used to detect plagiarism
- Adjacent agreement with human raters of 85-91% on GMAT essays
E-rater (ETS)

About the e-rater® Scoring Engine

What Is the e-rater® Engine?
The e-rater® engine is an ETS capability that identifies features related to writing proficiency in student essays so they can be used for scoring and feedback. Among other applications, e-rater is used within the Criterion® Online Writing Evaluation Service.

Who Uses the e-rater Engine and Why?
In the context of the Criterion application, students use the e-rater engine's feedback to evaluate their essay-writing skills as well as to identify areas that need improvement. Teachers use the Criterion service to help their students to develop their writing skills independently and receive automated, constructive feedback.

How Does the e-rater Engine Grade Essays?
The e-rater engine provides a holistic score for an essay as well as real-time diagnostic feedback about grammar, usage, mechanics, style and organization, and development. This feedback is based on natural language processing research specifically tailored to the analysis of student responses, and is detailed in ETS’s research publications.
Lots of commercial products now:

What is WriteToLearn?

WriteToLearn™ is a fully automated online literacy tool that accurately assesses writing and returns targeted instruction and grammar feedback within seconds, providing students with more opportunities to practice writing across subject areas.

Grade less, teach more, improve scores.

Try WriteToLearn
Immediate, Actionable Feedback

In a world of instant gratification, language assessments can now live up to our expectations. Our automated scoring solutions accelerate learning for corporations, government agencies, and learning institutions around the world.

Bayesian Essay Test Scoring System - BETSY

MyACCESS! School Edition

BECAUSE WRITING MATTERS

MY Access!® is our award-winning writing development solution that utilizes artificial intelligence and linguistic technologies to enhance student learning. Students experience a fun, interactive interface that encourages them with instant feedback for improvement.

OpenEd Solutions
A recent competition

AUTOMATED STUDENT ASSESSMENT PRIZE:
Phase 1 & Phase 2

A Case Study to Promote Focused Innovation in Student Writing Assessment

Jaison Morgan
Mark D. Shermis
Lynn Van Deventer
Figure 7. Line Chart for Vendor Performance on the Pearson Product Moment Correlation across the Eight Essay Data Sets
Figure 4. Line Chart for Vendor Performance on Exact+Adjacent Agreements across the Eight Essay Data Sets
From the report:

• “A predictive model may do a good job of matching human scoring behavior, but do this by means of features and methods which do not bear any plausible relationship to the competencies and construct that the item aims to assess. To the extent that such models are used, this will limit the validity argument for the assessment as a whole.”
How does automated scoring work?
E-rater in depth:

- Uses Natural Language Processing (NLP) techniques
- Standardized set of features include:
  - Errors in grammar, usage, mechanics, style
  - Organization & development
  - Lexical complexity
  - Prompt-specific vocabulary usage
Errors in grammar, usage, mechanics, & style (Quinlan, Higgens & Wolfe, 2009)

• Grammar: verbs, pronouns, possessives, wrong/missing words, sentence structure (fragments, run-ons, garbled sentences, subject-verb agreement)

• Usage: articles, confused words, incorrect word forms, faulty comparisons, nonstandard verbs or word forms

• Mechanics: spelling, capitalization, punctuation, fused words, compound words, duplicated words

• Style: repetition of words, inappropriate words or phrases, too many short or long sentences, too many sentences beginning with conjunctions
Development & organization

• Organization: number of discourse elements (introduction, conclusion, main points, supporting information)
• Development: length of discourse elements
Figure 1: A Student Essay With Annotated Discourse Elements

<Introductory Material> “You can’t always do what you want to do,” my mother said. She scolded me for doing what I thought was best for me. It is very difficult to do something that I do not want to do. </Introductory Material> <Thesis> But now that I am mature enough to take responsibility for my actions, I understand that many times in our lives we have to do what we should do. However, making important decisions, like determining your goal for the future, should be something that you want to do and enjoy doing. </Thesis>

<Introductory Material> I’ve seen many successful people who are doctors, artists, teachers, designers, etc. </Introductory Material> <Main Point> In my opinion they were considered successful people because they were able to find what they enjoy doing and worked hard for it. </Main Point> <Irrelevant> It is easy to determine that he/she is successful, not because it’s what others think, but because he/she have succeed in what he/she wanted to do. </Irrelevant>

<Introductory Material> In Korea, where I grew up, many parents seem to push their children into being doctors, lawyers, engineers etc. </Introductory Material> <Main Point> Parents believe that their kids should become what they believe is right for them, but most kids have their own choice and often doesn’t choose the same career as their parent’s. </Main Point> <Support> I’ve seen a doctor who wasn’t happy at all with her job because she thought that becoming doctor is what she should do. That person later had to switch her job to what she really wanted to do since she was a little girl, which was teaching. </Support>

<Conclusion> Parents might know what’s best for their own children in daily base, but deciding a long term goal for them should be one’s own decision of what he/she likes to do and want to do </Conclusion>

Attari & Burstein (2006, p. 9)
Lexical variables

• Lexical complexity
  • Average word length
  • Sophistication of word choice
• Topic-specific vocabulary usage:
  • Scores assigned to essays with similar vocabulary
  • Similarity to essays receiving highest scores
Use of e-rater

- Used as second rater for both independent and integrated writing tasks on TOEFL
- Anomalous essays go to two human raters
- If human rater and e-rater disagree, essay is sent to another human rater
- Criterion® (feedback and scoring tool sold primarily to schools or school districts) uses e-rater engine (more on this later!)
A few additional thoughts (Enright & Quinlan, 2010)

- Human raters attend to more features than e-rater but differ from each other; e-rater more consistent
- E-rater has potential for providing useful analytic information about specific aspects of writing
- Considerations for future include having humans rate essays on content & meaning, computers on more quantifiable aspects of essays
- Further research needed on impact of automated scoring on students and on teaching and learning
Getting computers to replicate human scores is the easy part.

- Can they measure aspects of writing that are important to raters and teachers?
- Can they provide useful feedback to students?
Welcome to Holt Online Essay Scoring, an Internet site designed to help you improve your writing skills and prepare for standardized writing tests. After you write a response to one of the prompts, you will receive an essay score and some written feedback about your paper from our computer scoring system. Don't forget to take advantage of these helpful features of our site:

- Graphic organizers that let you do your prewriting online
- Tips and reminders for writing and revising your essay
- Specific feedback on several features of your essay
- Suggestions for improving your writing
- A model paper based on each prompt
On a 4-point scale, here's your score: 2

2 This response demonstrates limited success with the persuasive writing task. The essay may:

- include some loosely related ideas that distract from the writer's position
- show some organization, with noticeable gaps in the logical flow of ideas
- offer routine, predictable ideas and reasons
- support ideas with uneven reasoning and elaboration
- exhibit limited control of written language

Dear Principle Adams,
Analytic Feedback for Your Essay

Our system has analyzed your essay for five important writing traits:

- Content and Development
- Focus and Organization
- Effective Sentences
- Word Choice
- Grammar, Usage, and Mechanics

Study the statements that describe each trait to help you improve your writing.

Content and Development Your essay shows limited ability for this trait. For the most part, the essay:

- uses routine, predictable ideas
- provides limited or uneven elaborations and support of ideas

Focus and Organization Your essay shows limited ability for this trait. For the most part, the essay:

- attempts to address the prompt but frequently loses focus
- shows little awareness of audience
- displays basic organization, with noticeable lapses in the logical flow of ideas and few, if any, transitions
- demonstrates minimal unity and completeness
Feedback on grammar

Grammar Checker

“Write like a native speaker with Ginger’s grammar checker!”
In the other way, people have time to do that also result in the improvement of financial conditions, which make sure people don't need to consider whether they can survive or not and then you have energy and time to consider other thinks, for instance, their appearance.
In the other way, people have time to do that also result in the improvement of financial conditions, which make sure people don't need to consider whether they can survive or not and then you have energy and time to consider other thinks, for instance, their appearance.

On the other way, people have time to do that also result in the improvement of financial conditions, which make sure people don't need to consider whether they can survive or not and then you have the energy and time to consider other thinks, for instance, their appearance.

No alternative sentences.
View Question

Subject-Verb Agreement

The **essence** of an old proverb are **Monkey see, Monkey do**. All to often many feel the necessity to do what others do, be it choosing cars, clothes, colors, restaurants, refreshments, and recreation. The most obtrusive factor is that the driver of the SUV can not see smaller cars well out the back and they have many mor blind spots. This poses a safety hazard to all others on the road. Another point is that they do not handle as well as smaller passenger cars that can be corrected in a emergency situation faster without the risk of rolling over this is obtrusive of the drivers of them to endanger passengers of smaller cars as well as pedestrians the stopping capabilities of these vehicles has been proven to be longer then passenger cars.

The SUV craze is obnoxious because it pollutes our atmosphere more then smaller cars. A second reason is that they take up lots of space in a parking lot. This makes other drivers mad and wastes the earth’s limited amount of space. The last obnoxious attribute is that everyone is always talking about them. This is very bothersome to ones ears to always hear about the fad that we all know will pass eventually. SUVs are destructive the earth in a few ways. One is that they pollute a lot. Another destructive attribute is that then people take them of road they damage our hilly areas and cause...
Questions we might ask:

• Does the system find the errors that are important to teachers and learners?
• How accurately does the system identify errors?
## Categories of ESL errors marked by teachers (Ferris, 2006)

<table>
<thead>
<tr>
<th>Error type</th>
<th>Percentage of errors marked</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Morphological errors</strong></td>
<td></td>
</tr>
<tr>
<td>Verbs (tense, form, subject/verb agreement)</td>
<td>21.6</td>
</tr>
<tr>
<td>Nouns (articles/determiners, noun endings: plural, possessives)</td>
<td>15.6</td>
</tr>
<tr>
<td><strong>Lexical errors</strong> (word choice, word form, informal usage, idiom error, pronoun error)</td>
<td>22.0</td>
</tr>
<tr>
<td><strong>Syntactic errors</strong> (sentence structure, run-ons, fragments)</td>
<td>27.2</td>
</tr>
<tr>
<td><strong>Mechanical</strong> (punctuation, spelling)</td>
<td>12.7</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>0.9</td>
</tr>
<tr>
<td>Criterion error types (Enright et al., 2010)</td>
<td>Errors marked by ESL teachers (Ferris, 2006)</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td><strong>Grammar</strong></td>
<td></td>
</tr>
<tr>
<td>Proofread this!</td>
<td>--</td>
</tr>
<tr>
<td>Ill-formed verbs</td>
<td>Verb tense</td>
</tr>
<tr>
<td></td>
<td>10.9%</td>
</tr>
<tr>
<td></td>
<td>Verb form</td>
</tr>
<tr>
<td></td>
<td>7.8%</td>
</tr>
<tr>
<td>Pronoun errors</td>
<td>Pronoun Error</td>
</tr>
<tr>
<td></td>
<td>2.9%</td>
</tr>
<tr>
<td>Possessive errors</td>
<td>Noun endings (plural/possessive)</td>
</tr>
<tr>
<td></td>
<td>8.9%</td>
</tr>
<tr>
<td>Wrong or missing words</td>
<td>Word choice (includes prepositions)</td>
</tr>
<tr>
<td></td>
<td>11.5%</td>
</tr>
<tr>
<td><strong>Sentence structure</strong></td>
<td></td>
</tr>
<tr>
<td>Fragments</td>
<td>Fragments</td>
</tr>
<tr>
<td></td>
<td>1.8%</td>
</tr>
<tr>
<td>Run-ons</td>
<td>Run-ons</td>
</tr>
<tr>
<td></td>
<td>2.9%</td>
</tr>
<tr>
<td>Garbled sentences</td>
<td>Sentence structure</td>
</tr>
<tr>
<td></td>
<td>22.5%</td>
</tr>
<tr>
<td>Subject-verb agreement</td>
<td>Subject-verb agreement:</td>
</tr>
<tr>
<td></td>
<td>2.9%</td>
</tr>
<tr>
<td><strong>Usage</strong></td>
<td></td>
</tr>
<tr>
<td>Article errors</td>
<td>Articles/determiners</td>
</tr>
<tr>
<td></td>
<td>6.6%</td>
</tr>
<tr>
<td>Confused words</td>
<td>--</td>
</tr>
<tr>
<td>Incorrect word forms</td>
<td>Word form</td>
</tr>
<tr>
<td></td>
<td>6.5%</td>
</tr>
<tr>
<td>Faulty comparisons</td>
<td>--</td>
</tr>
<tr>
<td>Nonstandard words or word forms</td>
<td>Informal usage:</td>
</tr>
<tr>
<td></td>
<td>0.3%</td>
</tr>
<tr>
<td></td>
<td>Idiom error:</td>
</tr>
<tr>
<td></td>
<td>0.8%</td>
</tr>
<tr>
<td>Preposition errors</td>
<td>--</td>
</tr>
<tr>
<td>Double negation</td>
<td>--</td>
</tr>
</tbody>
</table>
Accuracy of error identification

- Chodorow et. al (2010): Criterion intended to minimize false positives (saying something is an error when it’s not)
  - 25% of preposition errors with 80% accuracy
  - 40% of article errors with 90% accuracy
- Ferris (2006): Teachers agreed with trained raters 80-90% of the time on all errors
Table 2

Grammar Errors Identified by Criterion and Instructor

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Instructor Total</th>
<th>Coded Inaccurate</th>
<th>Criterion Total</th>
<th>Coded Inaccurate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wrong or Missing Word</td>
<td>213</td>
<td>5 (2.3%)</td>
<td>3</td>
<td>1 (33.3%)</td>
</tr>
<tr>
<td>Ill-formed Verbs</td>
<td>90</td>
<td>1 (1.1%)</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>Proofread This!</td>
<td>59</td>
<td>0</td>
<td>8</td>
<td>6 (75%)</td>
</tr>
<tr>
<td>Subject-Verb Agreement</td>
<td>56</td>
<td>1 (1.8%)</td>
<td>11</td>
<td>2 (18.2%)</td>
</tr>
<tr>
<td>Pronoun Errors</td>
<td>52</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Garbled Sentences</td>
<td>36</td>
<td>0</td>
<td>3</td>
<td>3 (100%)</td>
</tr>
<tr>
<td>Fragment or Missing Comma</td>
<td>29</td>
<td>0</td>
<td>33</td>
<td>10 (30.3%)</td>
</tr>
<tr>
<td>Possessive Errors</td>
<td>26</td>
<td>0</td>
<td>3</td>
<td>1 (33.3%)</td>
</tr>
<tr>
<td>Run-on Sentences</td>
<td>9</td>
<td>0</td>
<td>17</td>
<td>12 (70.6%)</td>
</tr>
<tr>
<td>Error Type</td>
<td>Instructor Total</td>
<td>Coded Inaccurate</td>
<td>Criterion Total</td>
<td>Coded Inaccurate</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Missing or Extra Article</td>
<td>118</td>
<td>0</td>
<td>76</td>
<td>31 (40.8%)</td>
</tr>
<tr>
<td>Preposition Error</td>
<td>106</td>
<td>0</td>
<td>19</td>
<td>5 (26.3%)</td>
</tr>
<tr>
<td>Wrong Form of Word</td>
<td>96</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Wrong Article</td>
<td>37</td>
<td>0</td>
<td>17</td>
<td>15 (88.2%)</td>
</tr>
<tr>
<td>Nonstandard Word Form</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Confused Words</td>
<td>2</td>
<td>0</td>
<td>11</td>
<td>3 (27.3%)</td>
</tr>
<tr>
<td>Faulty Comparisons</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Negation Error</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>364</td>
<td>0</td>
<td>125</td>
<td>54 (43%)</td>
</tr>
</tbody>
</table>
Is the automated feedback meaningful to students’ learning?

- Chen & Cheng (2008): Student learning from automated feedback depended on how teachers used the software (MyAccess!, 3 EFL classrooms in Taiwan)
  - Students found feedback “vague,” “abstract,” and “repetitive”
- Grimes & Warschauer (2010): Software simplified classroom management, increased student motivation to write (MyAccess!, middle schools in California)
  - Amount of feedback overwhelming; teacher support needed to help students prioritize and use feedback
I can’t afford to pay for an automated text analysis tool. What’s out there I can use for free for my research?
What we can count:

- Syntactic variables (e.g. Lu's L2 Syntactic Complexity Analyzer)
- Overall sentence complexity
- Clausal coordination
- Overall T-unit complexity
- Clausal subordination
- Elaboration at clause level
- Non-finite subordination
- Phrasal coordination
- Noun phrase complexity

Mean length of sentence
T-units per sentence
Mean length of T-unit
Dependent clauses per T-unit
Mean length of clause
Non-finite elements per clause
Coordinate phrases per clause
Complex NPs per clause
Variables that predicted high scores on L2 essays: (Yang, Weigle & Lu, 2014)

- Overall sentence complexity
- Clausal coordination
- Overall T-unit complexity
- Clausal subordination
- Elaboration at clause level
- Non-finite subordination
- Phrasal coordination
- Noun phrase complexity

Mean length of sentence
T-units per sentence
Mean length of T-unit
Dependent clauses per T-unit
Mean length of clause
Non-finite elements per clause
Coordinate phrases per clause
Complex NPs per clause
What we can count:

- Lexical variables e.g., Coh-Metrix (Crossley et al., 2011)

- $D$ (Lexical diversity)
- CELEX content word frequency
- Word meaningfulness content words
- Word familiarity content words
- Word imagability all words
- Word hyponymy average
- LSA sentence to sentence adjacent (Semantic co-referentiality)
- Word concreteness all words
- Polysemy standard deviation
- Average syllables per word
These variables predicted essay scores (Crossley et al, 2011):

- $D$ (Lexical diversity)
- CELEX content word frequency
- Word meaningfulness content words
- Word familiarity content words
- Word imagability all words
- **Word hypernymy average**
- LSA sentence to sentence adjacent (Semantic coreferentiality)
- Word concreteness all words
- Polysemy standard deviation
- Average syllables per word
In other words.....

- Essays with longer sentences and more diverse, less frequent words get higher scores.
- Can we improve on this?
Comparing to reference corpora
It is always said that the first impression is the last impression so it automatically follows that appearance is important how you look does matter a lot now some might say that beauty is superficial you need to go beyond it when I refer to appearance I refer to appearance in general not to how beautiful or ugly one is I associate adjectives like neat tidy and untidy with the word appearance and not beautiful or ugly

appearance is of utmost importance not only in the business words but in all walks of life even in personal life appearance is important nobody would like a badly dressed spouse you would not want to be embarrassed because of how someone near and dear to you looks how you look defines your personality when you meet a person for the first time you judge him from his appearance a clean well dressed person with neatly kept hair will appeal to you more than a badly dressed person with unkempt hair your appearance speaks volumes about your personality it tells about the kind of person you are however brilliant you are you will always remain behind if your appearance is not acceptable now why do we dress formally for interviews and presentations when we dress well and bear a decent appearance we feel more confident a good appearance is appealing not only to the other person but also to oneself it helps to boost our confidence

It is often said that you feel the way you look so in order to feel good you need to look good your first impression is many a times the last impression so you need to make sure that your first impression is good at all times and that will happen only if you bear a decent appearance

An essay analyzed by VocabProfile (http://www.lextutor.ca/vp/eng/)
The same essay analyzed in Text Inspector – English Profile
(http://www.englishprofile.org/index.php/wordlists/text-inspector)

<table>
<thead>
<tr>
<th>Word List</th>
<th>Types</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>81</td>
<td>197</td>
</tr>
<tr>
<td></td>
<td>(59.12%)</td>
<td>(71.38%)</td>
</tr>
<tr>
<td>A2</td>
<td>19</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>(13.87%)</td>
<td>(8.70%)</td>
</tr>
<tr>
<td>B1</td>
<td>17</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>(12.41%)</td>
<td>(10.51%)</td>
</tr>
<tr>
<td>B2</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>(8.03%)</td>
<td>(5.80%)</td>
</tr>
<tr>
<td>C1</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>(3.65%)</td>
<td>(1.81%)</td>
</tr>
<tr>
<td>C2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>(2.19%)</td>
<td>(1.45%)</td>
</tr>
<tr>
<td>Unlisted</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>(0.73%)</td>
<td>(0.36%)</td>
</tr>
</tbody>
</table>
Results: Comparison with A-graded papers from MICUSP

Screenshot from Test X-ray (http://www2.gsu.edu/~wwwesl/16027.html)
What we can’t easily count (by machine)

• Claims & backing
• Logical organization
• Effectiveness of supporting details
• Features of audience awareness
• Etc.
• i.e., What learners DO with their linguistic resources to fulfill a communicative goal
Genre-based automated writing assessment (Cotos, 2014)
B2 descriptors of argumentation *(Neff-van Aertselaer, 2013)*

<table>
<thead>
<tr>
<th>Structural features</th>
<th>Qualifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can reword the prompt of a writing assignment incorporating opposing points of view</td>
<td>Proper type and amount of contextualization</td>
</tr>
<tr>
<td>Can establish or progressively refine the goal of the argument(s)</td>
<td>Establishes points of dispute within the argument</td>
</tr>
<tr>
<td>Can present all claims and supporting data in a logically organized way</td>
<td>Few stranded claims or data</td>
</tr>
<tr>
<td>Can clarify claims by offering reasons and evidence through definitions, elaboration, etc.</td>
<td>Issues accurately defined in order to sustain further development</td>
</tr>
<tr>
<td>Can use both prospection and encapsulation to create coherence</td>
<td>Few limitations regarding lexical phrases used</td>
</tr>
<tr>
<td>Can conclude by restating major ideas and placing the arguments in a wider context</td>
<td>Suggestion of future circumstances</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rhetorical features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can calculate which information (points of dispute) is adequate for a given audience</td>
</tr>
<tr>
<td>Can consider a range of viewpoints, adopting a critical stance</td>
</tr>
<tr>
<td>Can report others' views and statements, using discourse verbs which show writer alignment/non-alignment</td>
</tr>
<tr>
<td>Can use a reasonably extensive range of hedges and boosters as well as impersonalization strategies in presenting claims</td>
</tr>
<tr>
<td>Can successfully use a variety of discourse markers (DMs) to indicate the flow of a text</td>
</tr>
</tbody>
</table>

| Exclusion of information not pertinent to the claims                                  |
| Distinguishes the arguments in sources in order to summarize or paraphrase them      |
| Uses evaluative lexical devices to include sources' voices; aided by a wide range of reporting verbs (suggest, claim, show, etc.) and sometimes by reader-in-the-text strategies |
| Shows effective use of passive voice, modalized utterances, abstract rhetors         |
| Effectively deploys lexical cohesive devices (e.g., synonyms, hyponyms,) and discourse markers (DM) |
To summarize

• Automated writing analysis systems can do a good job of predicting human scores but
  • They aren’t always based on a writing construct
  • They don’t always do well with L2 writing (particularly in terms of providing useful feedback)
  • Many are commercial products and deserve close scrutiny (Chapelle & Chung, 2010)
• What counts in writing is not always easy to count; for now, automated systems can only complement, not replace, human judgment.
Thank you!

Comments/questions?
Contact me at sweigle@gsu.edu
References


