

# **Text Understanding Techniques for Automated Assessment**

Claudia Leacock  
Educational Testing Service

# ETS Natural Language Processing Group

Jill Burstein

Martin Chodorow

Lisa Hemat

Karen Kukich

Claudia Leacock

Chi Lu

Susanne Wolff

Daniel Zuckerman

# Scoring Constructed Responses

... is labor intensive, time-consuming and expensive.

- **Uncoachable:** e.g., avoid use of length
- **Defensible:** Use scoring guide criteria
- **Evaluation:** Compare performance with human readers

# Outline

- ***e-rater***: operational essay scoring system
- ***c-rater***: research collaboration for scoring course-based questions.

# ***e-rater*** **(analytic writing skills)**

- holistic scoring**
- high stakes (GMAT)**
- no solo scoring (...yet)**

# Example Prompt

## Analysis of an Issue

[www.gmat.org](http://www.gmat.org)

In some countries, television and radio programs are carefully censored for offensive language and behavior. In other countries, there is little or no censorship.

In your view, to what extent should government or any other group be able to censor television or radio programs? Explain, giving relevant reasons and/or examples to support your position.

# Holistic Scoring Rubric

## Rubric Criteria

- Syntactic Variety
- Vocabulary Usage
- Organization of Ideas

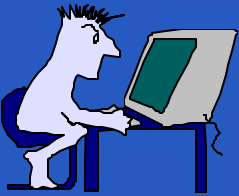
## *e-rater* Variables

- Sentence Structure
- Content Analysis
- Rhetorical Structure
- Content Analysis for Arguments

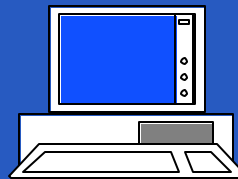
# 50+ Features for Scoring

- **Syntactic Structure Features**
  - Subordinate, Relative, Infinitive, ... clauses
- **Content Features**
  - “score” from content words in essay
- **Rhetorical / Discourse Structure Features**
  - parallel, contrast, evidence, ...argument development

# NLP & Essay Scoring



*“I also assume that shrinking high school enrollment ...”*



Parse:        S        NP |prp I ↓  
                          VP |rb also  
                                  |vbp assume  
                          SC COMP |wdt that ...

Syntactic: *COMPCL*

Discourse: *also = parallel argument*  
*that = claim*

Content: { *assume, shrink, high, school, enrollment ...* }

# Building Models & Scoring

- **Build Essay Models**
  - **Collect** feature information from hand-scored essays
  - **Generate** weighted predictive feature set using regression for each prompt

## Score Essay Responses

- Use weighted predictive feature set in score prediction formula

# ***e*-rater Performance**

## **GMAT:**

**91% agreement between two human readers.**

**91% agreement between e-rater and a human reader.**

# Course-based Short-Answer Questions: *c*-rater

Collaboration between ETS and NYU Virtual College.

- “gold standard” in Teacher’s Guide
- low stakes (quizzes)
- solo scoring
- pass/fail grades

# Example Prompt

## Systems Auditing & Database Management Courses

**Q: Differentiate between triggers and stored procedures.**

**A: Triggers are programs embedded within a table that are automatically invoked by updates to another table. Stored procedures are programs embedded within a table that can be called from an application program.**

# Paraphrase Recognition

- **Syntactic variety:**
  - ...can be called from a program.
  - ...that a program can call.
- **Synonymy**
  - ...can be invoked from a program.
- **Negation**
  - ...are not invoked by updates ...
- **anaphoric reference**
  - Triggers are programs. They are embedded ...

# tuples: Predicate Argument Structure

Triggers are programs embedded within a table that are automatically invoked by updates to another table.

are :obj programs :subj triggers  
embedded :within table  
invoked :obj that  
updates :to table

# Lexical Substitution

...invoked

called  
activated  
triggered

by updates

data modification

to another

a different  
some other  
an additional

table

file  
database object

# Identify Synonyms

- **Statistical Thesauri**
  - technical terms: textbook**
  - non-technical terms: on-line Roget**

# Technical Terms

**Statistical Thesaurus built from the textbook:**

program: application .765, code .549, serial .135

update: data modification .576, news .122

table: file .673, database object .528, chair .118

# Strategy

- **Recover** predicate argument structure.
- **Identify** technical terms and non-technical terms.
- **Map** onto the representation of the gold standard.

**Evaluate** c-rater on answers provided by NYU students.

**For more information...**

**[www.ets.org/research/erater.html](http://www.ets.org/research/erater.html)**