Automated Scoring Readiness for Next Generation Assessments

Karen Lochbaum
May 18, 2011
Welcome & Introductions

Moderator: Anne Johnson
Program Manager, Pearson

Presenter: Karen Lochbaum
VP Technology Services
Knowledge Technologies, Pearson
Common Core State Standards
Key Features

- 21st Century Skills
- More authentic tasks and assessment
- More constructed responses => Automated scoring
Benefits of Automated Scoring

- **Immediacy & Efficiency**
  - Evaluate responses in seconds
  - Reduce score turnaround time
  - Give students and teachers instant feedback
  - Reduce costs

- **Accuracy**

- **Consistency, Objectivity**

- **Can detect off-topic, inappropriate and “odd” responses**
Common Core State Standards
Key Features

• **Reading**: Text complexity and the growth of comprehension
• **Writing**: Text types, responding to reading, and research
• **Speaking and Listening**: Flexible communication and collaboration
• **Language**: Conventions, effective use, and vocabulary
• **Mathematics**
The world’s population, or total number of people, lives in uneven clusters on Earth's surface. Some places have many people. Other places are almost empty.
Population is how many people there are in a place. Some places have many people. Other places are almost empty. Demographers are people who study the number of people in specific regions and why they are there. Demographers want to understand why. They want to know why about populations like why some places have lots and lots of people and other places do not have so many. Regions where farming was a big part of the economy have larger populations even though they don't have as many farmers any more. Farms are not as crowded as cities.
You should improve your summary. Use the bars below to see where to focus your effort.

**Section Coverage**
- **Population Distribution**
- **Population Density**
- **Birthrates and Death Rates**
- **The Challenges of Population Growth**

**Length**
- 98 words
- Long
- Good
- Short

**Editing Tools**
- **Copying**
  - Status: Fair
  - Comment: Some copied parts
- **Spelling**
  - Status: Fair
  - Comment: Some misspelled words
- **Repeated**
  - Status: Excellent
  - Comment: No repeated information
- **Unrelated**
  - Status: Poor
  - Comment: Some unrelated content
Text Complexity

- Conceptual level of vocabulary vs. surface level (e.g. word frequency)
- Measure how word meanings are learned and change over time with increasing exposure
Writing

Asks students to answer several questions about a hypothetical, yet realistic, scenario.

You advise Pat Williams, the president of DynaTech, a company that makes precision electronic instruments and navigation equipment. Sally Evans, a member of DynaTech’s sale force, recommends that DynaTech buy a small private plane (a SwiftAir 235) that she and other member of the sales force could use to visit customers. Pat was about to approve the purchase when there was an accident involving a SwiftAir 235.

Document Library
- Newspaper article about the accident
- Federal Accident Report on in-flight breakups in single engine planes
- Internal Correspondence (Pat’s email to you and Sally’s e-mail to Pat)
- Charts relating to SwiftAir’s performance characteristics
- Excerpt from magazine article comparing SwiftAir 235 to similar planes
- Pictures and descriptions of Swiftair Models 180 and 235
Science

Use the technical passage 'Green Ocean Machine' to answer the following.

The passage states that “the new green partner [alga] seems to provide Hatena with most of its energy needs.”

Describe the process that enables organisms to use energy from light to make food. In your description, be sure to include

* the specialized features needed to produce food
* the substances needed to produce food
* the substances produced during this process
Listening & Speaking

**Score Report**

**Versant Junior English Test**

- **Test Taker's Name:** Student 1
- **Test Identification Number:** 12345678
- **Test Completion Date:** May 22, 2009
- **Test Completion Time:** 1:23 PM (PDT)

**Overall Score:** 126

<table>
<thead>
<tr>
<th>Skill Area</th>
<th>Score</th>
<th>70</th>
<th>90</th>
<th>110</th>
<th>130</th>
<th>150</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>126</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td>130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentence Mastery</td>
<td>150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pronunciation</td>
<td>102</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© 2009 Pearson or its affiliates. All rights reserved. Ordinate is a registered trademark and Versant is a trademark of NCS Pearson, Inc. Other names may be the trademarks of their respective owners.

For more information, call +1 650-470-3565 or visit us online at www.VersantTest.com
Type or paste your essay into the box below.

User: eport  
Prompt: Community Service

Some people think that all 18 year-olds should perform one year of national or community service (for example, the Peace Corps, emergency medical services, Big Brother and Big Sister programs) before they pursue college or a paying job. What is your position on the issue? Write an essay in which you support your point of view with convincing reasons and examples.

(Expected essay length: 150 - 650 words.)

The more experience a person has to go into a specific field the better off they will be in their field. Therefore in performing national or community service I think it’s a good idea for 18 year-olds to have to complete this service before they pursue college or a paying job. In the issue of national or community service the 18 year old will be better able to perform their job or studies.

Furthermore the service will make the 18 year old more rounded as an individual therefore allowing a better chance of success. To begin, national or community service teaches 18 year olds things that their school might not have gone over because in school they go over certain things that have to be completed in a certain time. These special services allow brief examples that might happen in the real world which prepare these 18 year olds for that job or college. In anything a person does the more experience they have the more they learn which allows the more things they can do. In a job or college this is the key to success.

Next, doing such service will show that the 16 year old is serious about what he or she is trying to get into and that they will be committed to it because they are taking that extra step to be successful and learn more than what they already know. The more rounded an individual is said to be the more successful they are likely to be because they can adapt to survive any type of situation that comes at them. Without seeing what goes on before the 19 year olds pursue college or a paying job is like them walking blind-folded on a thin line and its likely they will fall. Its an elevator to success in their lives, as well as a key to success.

In conclusion national or community services doesn’t hinder a 18 year old.
That is a very good first draft. Use the feedback below to make it even better.

**Scores**

<table>
<thead>
<tr>
<th>Category</th>
<th>Score</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>4 of 6</td>
<td></td>
</tr>
<tr>
<td>Ideas</td>
<td>3 of 6</td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>3 of 6</td>
<td></td>
</tr>
<tr>
<td>Conventions</td>
<td>4 of 6</td>
<td></td>
</tr>
<tr>
<td>Sentence Fluency</td>
<td>4 of 6</td>
<td></td>
</tr>
<tr>
<td>Word Choice</td>
<td>4 of 6</td>
<td></td>
</tr>
<tr>
<td>Voice</td>
<td>3 of 6</td>
<td></td>
</tr>
</tbody>
</table>

**Length**

- 328 words
- **Length**: Long

**Editing Tools**

- **Spelling**: Almost
  - A few misspelled words
- **Grammar**: Fair
  - A few grammar errors detected
- **Repeated**: Excellent
  - No repeated information

**Get Definition**

- repeated
Language

Part D: Sentence Builds

Please rearrange the word groups into a sentence.

Example:

A voice says "finished by April"..."must be..."our projects" and you say "Our projects must be finished by April."
Add the following fractions with unlike denominators. First find equivalent fractions, and enter them. Then, write the sum in simplest form.

\[
\frac{1}{2} + \frac{3}{10} \\
\downarrow \quad \downarrow \\
\frac{5}{20} + \frac{6}{20} = \frac{11}{20}
\]

Analysis:
Student incorrectly converted both fractions. Equivalent fraction requisite skill for this task.

Did the student fill in all blanks? Yes
Is the response correct? No
Did the student convert both fractions to results with the same denominator? Yes
Is the sum of the student’s equivalent fractions correct? Yes
Is the denominator of the solution equal to (or a reduced form of) the direct sum of the converted fractions? Not Checked
Were one of the fractions converted correctly? Not Checked
Did the student convert the first fraction correctly? Not Checked
Did the student convert the second fraction correctly? Not Checked
Did the student avoid the standard mistake: adding numerators and denominators? Not Checked
Did the student choose the same denominator for both addends? Not Checked
Are the denominators of the two fractions the LCM of the two original denominators? Not Checked
Was the answer reduced (if applicable) to lowest terms? Not Checked
Automated Scoring Approach

- Learn from human scored student responses
- Measure the content and quality of responses by determining
  - The language features that human scorers evaluate when scoring a response
  - How those features are weighed and combined to produce scores
Essay Scoring Process

Essay Score

- Content
- Style
- Mechanics
- Scoring Confidence
  
- Similarity to expert scored essays
- Coherence
- Grammar
- Off-topic Detection
The Intelligent Essay Assessor

Learn to score like human scorers by measuring different aspects of writing

- **Content** -- including subject area knowledge
  - Semantic analysis, measures of similarity to prescored responses, ideas, vocabulary growth, examples, ....

- **Style**
  - Appropriate word choice, word and sentence flow, fluency, coherence, ....
  - Does each sentence logically follow the next?
  - Does each sentence contribute to the essay as a whole?

- **Mechanics**
  - Grammar, word usage, punctuation, spelling, ...
Other Features of IEA

• Uses non coachable measures
  – No counts of total words, syllables, characters, etc.
  – No trigger surface features: “thus”, “therefore”
  – Detects larding of big words

• Knows when it doesn’t know
  – Detects off-topic or highly unusual essays, non-standard language constructions, too long, too short ...
Content Based Scoring

- Use Latent Semantic Analysis (LSA) to capture the “meaning” of language
- LSA knows that
  - Surgery is often performed by a team of doctors.
  - On many occasions, several physicians are involved in an operation.
  mean about the same thing even though they share no words.
- Enables evaluating the content of what is written rather than just matching keywords
Why LSA?

Search for “Cars”...
Why LSA?

- Studies have shown that:
  - People agree on the Keywords for a text only 15% of the time
  - If you have 100 people name a document, you will get 30 different answers
- LSA operates on the level of deep word (latent) meaning

<table>
<thead>
<tr>
<th>Key Word</th>
<th>LSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctor—Doctor</td>
<td>1.0</td>
</tr>
<tr>
<td>Doctor—Physician</td>
<td>0.8</td>
</tr>
<tr>
<td>Doctor—Surgeon</td>
<td>0.7</td>
</tr>
</tbody>
</table>
What does that have to do with automated scoring?

• LSA reads lots of text
• Learns what words mean and how they relate to each other
• Result is a “Semantic Space”
• Every word represented as a vector
• Every paragraph represented as a vector

\[ M(\text{Paragraph}) = M(w1) + M(w2) + \ldots \]
• Every essay represented as a vector
• New essays are placed based on the words they contain
• Every section represented as a vector
• Student summaries are placed based on the words they contain
Spoken Assessments

waveform

spectrum

words

segmentation
Example: Learner

REPEAT: New York City is famous for its ethnic diversity.

Pronunciation: 5.9
Fluency: 3.3
Accuracy: 1 word error (insertion)
Performance Comparison

3.026 seconds
Native speaker

5.502 seconds
Learner

Pronunciation
Fluency
Accuracy
Versant Scoring Logic

17 minutes
Mathematics Representation

c = \sqrt{a^2 + b^2}

Equations are saved using MathML markup, thus preserving the computational meaning of the math even if the presentation is changed.
Keys to Success

Design for automated scoring from the start!
Keys to Success

✔ Item Development
  – Optimize for scoring effectiveness

✔ Item Delivery
  – Math: Input and capture of student response

✔ Field Test and Human Scoring
  – Representative samples
  – Double scoring with resolution
Keys to Success

✔ Psychometrics
  – Automated scoring performance as part of field test item evaluation

✔ Operational Scoring & Monitoring
  – Requirements vary with nature of assessment and acceptable performance criteria
  – Automated scoring in combination with human scoring
Q&A/Discussion

• If you have not done so already, please type any questions or comments you have about the webinar into the Chat Box on your screen.
• You may also email questions directly to anne.johnson@pearson.com after the webinar.

Thank you!

Please join us May 25th when we discuss “Through Course Common Core Assessments: A Proposed Design for English Language Arts” pearsonassessments.com/nextgenwebinars