

Automated Scoring of Writing Assessments

IPMA Assessment Council

June 6, 2000

Automated Scoring

- Automated scoring is one application of automated text processing technology
- This technology is currently emerging from the R&D phase
- OPM is investigating practical applications of this technology by government agencies

Potential Benefits

- Completely automate scoring of low-stakes writing samples
- Reduce the number of graders needed to evaluate each writing sample
- Screen out lower quality writing samples
- Change the nature of the writing sample evaluation task

Practical Considerations

- Choose the right technology
- Convince others that the technology is real
- Avoid the transcription trap
- Remain concerned about rater validity
- Look inside of the black box

The Right Technology

- Project Essay Grade (PEG)
 - Ellis Page, Duke University
 - Indicators of Writing Quality
- Latent Semantic Analysis (LSA)
 - Thomas Landauer, Colorado
 - Content Coverage
- E-rater
 - Jill Burstein, Educational Testing Service
 - Argument Structure

The Technology is Real

- Six levels of technology

1. Office automation
2. Professional tools
3. Programming tools
4. Emerging technology
5. R&D technology
6. Star Trek!

The Technology is Real

- Text classification, not artificial intelligence
- Correlations of .85+ with trained raters
- Being used in GMAT and other high stakes assessments

Need for Caution...

- Limits to the kinds of writing samples that can be evaluated
- More superficial than evaluation by human judges
- More vulnerable to certain faking strategies
- Can be resource-intensive
- Legally defensible?

Does my project fit?

- Data in electronic text format
- 200+ words in each sample
- Need to evaluate 300+ writing samples
- Address a single topic with common content coverage
- Set of 100+ graded by human experts
- Low stakes or include human graders

Additional Applications

- Identifying themes in critical incident descriptions for job analysis
- Summarizing responses to open-ended training evaluation questions
- Balancing content between two or more forms of a traditional job knowledge test
- Identifying features of multiple choice questions that contribute to item difficulty

Themes

- Technology drives us forward, ready or not
- Accountability will remain with content experts and measurement professionals
- Partial automation rather than full automation for the near future
- Large rather than small samples of text