

Computer vs. Human Scoring in the Assessment of Hong Kong Students' English Essays

Kinnie CHAN

The Open University of Hong Kong



Background

- **In the early 1960s, Ellis Page developed the first AES, Project Essay Grade (PEG), to make large scale essay scoring practical and effective (Page, 2003; Rudner & Gagne, 2001).**
- **Automated Essay Scoring (AES) systems**
 - **numerous research revealed strong evidence**
 - **demonstrates well correlation with human rater behaviour (Shermis & Burstein, 2003)**



Purpose of the research

- Investigate the effectiveness of an Automated Essay Scoring (AES) system in measuring essay writing ability as scored by human raters



The Automated Essay Scoring (AES) System

- **An online scoring device**
- **It does not need any essays for training purposes**
- **Based on the algorithm of word frequency & sentence length**
- **Provides essay writing ability measure via instant feedback**



Method

- **Winsteps & Facets**
- **137 Grade 6 Hong Kong students from one local primary school**
- **Students responded to 3 prompts**
- **Prompts from NAEP**



Collapsing of categories

The guidelines for combining categories were applied to produce a measurement process under the condition of an equal contribution for each category in the Rasch analysis (Linacre, 1995, 1999 & 2002).



Results

- 32 misfitting students (out of 137 students)
 - Infit MnSq > 1.4

A misfitting student's essay

Please read along as your teacher reads you the following writing assignment:

Imagine this situation:

Your favorite book is missing from your school library. It might be a book that you like to read over and over again. Or it might be a book that your teacher or parent has read to you. Some of your friends also like to read this book. The school librarian is not sure she wants to buy the book again.

Write a letter to convince your school librarian to buy the book again. In your letter, give lots of reasons why the book should be in your school library.

My name is [unclear] I will
have to look for it if we
can't find it we will have to
pay for it if I can't pay for
it I will go write out a
check, but I will try to
I am my book.



Results

- **32 misfitting students (out of 137 students)**
- **3 prompts fit the Rasch model very well**

Results

- 32 misfitting students (out of 137 students)
- 3 prompts fit the Rasch model very well
- 4 raters show sufficient fit
 - Infit MnSq between 0.6 and 1.4

Results

- **32 misfitting students (out of 137 students)**
- **3 prompts fit the Rasch model very well**
- **4 raters show sufficient fit**

Conclusions

- All 4 human raters can score essays consistently.
- The AES can score as consistent as the human raters

