

Universiteit Utrecht

## Measuring writing proficiency

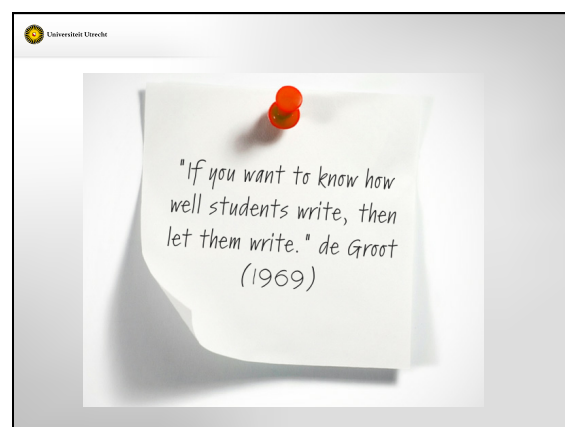
### Effect of genre on the generalizability of writing scores

Conference 'Writing research across borders'

Renske Bouwer<sup>a</sup>, Anton Béguin<sup>b</sup>, Ted Sanders<sup>a</sup> & Huub van den Bergh<sup>a</sup>

<sup>a</sup> Utrecht University  
<sup>b</sup> Institute for Educational Measurement (Cito), Arnhem

March 1, 2015



Universiteit Utrecht

### Measurement problems

- Effects of the rater:
  - Raters differ in leniency/severity
- Effects of the task:
  - Tasks differ in level of difficulty

→ Problem:  
It's hard to determine student's writing ability based on quality ratings of only one written text.

March 1, 2015 3

Universiteit Utrecht

### Generalizability theory

Text quality score = Writing skill + Measurement error

↓

|        |       |             |
|--------|-------|-------------|
| Raters | Tasks | Unexplained |
|--------|-------|-------------|

For reliable assessments: keep the measurement error as small as possible

4

Universiteit Utrecht

### How many tasks and raters?

**Writing assessment: at least 5 tasks & 3 raters**  
Primary, secondary & higher education

**Type of tasks assigned to students differ across studies**

- Persuasive texts (van Weijen, 2009)
- Persuasive and functional texts (Schoonen, 2005)
- Persuasive texts and filling in forms (van den Bergh, 1989)

→ Is generalization of similar texts warranted across genres?

5

Universiteit Utrecht

### Research question

How can we make valid and reliable decisions about writing proficiency of students at the end of primary education (11/12 years old)?

- How many tasks and raters are necessary for the assessment?
- What is the influence of genre?

6

**Method**

**Participants**  
67 students (grade 6) from 3 different primary schools

**Material**  
4 genres: formal letters, narratives, persuasive and personal texts  
3 texts on different topics for each genre

**Rating procedure**  
3 raters for each text (raters are student teachers)  
Global ratings using a benchmark

7

**Results I**

| Source   | Variance in % |
|--|---------------|
| Students   |               |
| Genre  |               |
| Students * Genre   |               |
| Task within genre  |               |
| Students * Task  |               |
| Raters   |               |
| Interaction students*task*genre*rater & unexplained variance |               |

8

**Results I**

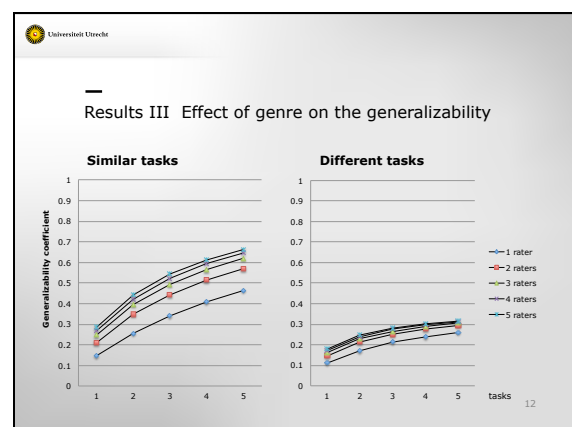
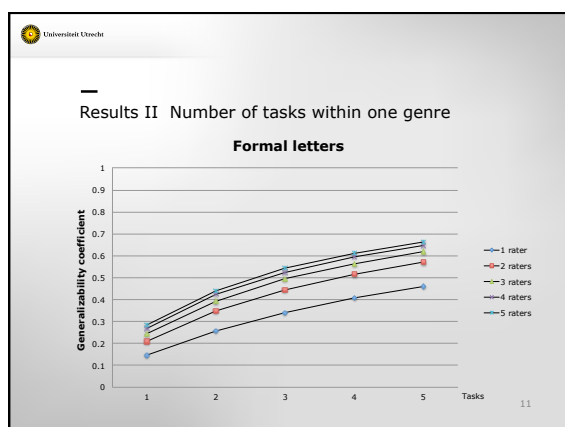
| Source   | Variance in % |
|--|---------------|
| Students   | 9,98          |
| Genre  |               |
| Students * Genre   |               |
| Task within genre  |               |
| Students * Task  |               |
| Raters   |               |
| Interaction students*task*genre*rater & unexplained variance |               |

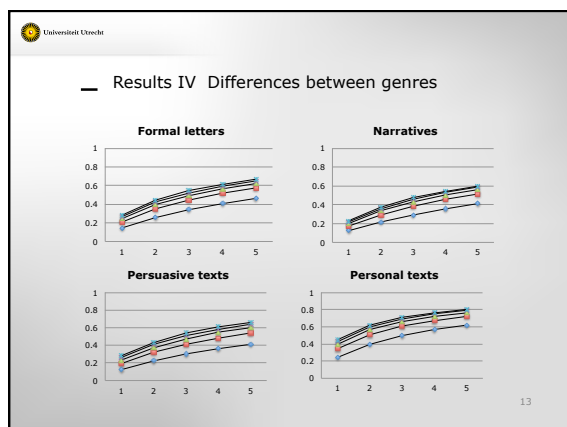
9

**Results I**

| Source   | Variance in % |
|--|---------------|
| Students   | 9,98          |
| Genre  | 11,42         |
| Students * Genre   | 4,01          |
| Task within genre  | 1,71          |
| Students * Task  | 19,13         |
| Raters   | 18,05         |
| Interaction students*task*genre*rater & unexplained variance | 35,71         |

10





— Conclusion: how to measure writing proficiency?

It is possible to determine how well young students write, but:

- ① Let them write multiple texts, rated by multiple raters.
- ② If tasks in the writing assessment are very similar, decisions about writing ability should be limited to writing in that specific genre.

— Questions or suggestions?

**schrijfproject**  
Voor beestachtig goede teksten!

For more information:  
[www.schrijfproject.nl](http://www.schrijfproject.nl)  
 Twitter @schrijfproject  
 E-mail i.r.bouwer@uu.nl

| Formal letter   | Argumentative text                            | Adventure story               | Personal narrative                         |
|---|---|-------------------------------|--|
| Collection of toys at a supermarket                       | Pros and cons of a smoking ban                | Adventure on a sports field   | Personal experience about being frightened |
| Collection of stamps for musical tickets                  | Pros and cons of a candy prohibition          | Adventure about a forest-fire | Personal experience about being caught     |
| Collection of points on wraps of chocolate for a music CD | Pros and cons of telling tales about somebody | Adventure about poison        | Personal experience about being home alone |