

BestPrac - EARMA virtual meeting

# Writing style and grant application success

Peter van den Besselaar

Vrije Universiteit Amsterdam

TMC Amsterdam

German Center for Higher Education and Science Studies (DZHW) Berlin



# Outline

- What predicts grant success?
  - Linguistic approaches to grant writing
  - Case
  - Results
  - Implications
- 
- Acknowledgement: coauthor Charlie Mom (TMC Amsterdam)

# What predicts grant success?

## Past performance

- Productivity and citation impact (+)
- Earlier grants (+)
- Quality of the proposal (?)
  - Originality, novelty

Correlation between scores	Step 1 CV	Step 2 project
Step 1 project	0.84	0.43
Step 2 CV	0.59	0.69

# What predicts grant success?

## Reputation

- Number international coauthors (+)
- Quality (prestige) of network (+)
  - E.g., Ranking or collaborating organizations
- Journal prestige (+)
  - Journal impact factor

# What predicts grant success? Bias

- Cognitive proximity (yes/?)
- Nepotism (yes)
- Gender bias (yes)

- Panel characteristics correlate with gender bias:
  - Share women
  - International variety
  - Workload
  - Formalization of procedures
  - Gender stereotyping

nature

Explore content ▾ Journal information ▾ Publish with us ▾ Subscribe

nature > news > article

NEWS | 21 May 2021 | Update 07 June 2021

## Prestigious European grants might be biased, study suggests

Institutional affiliations of panellists seem to skew European Research Council decisions – especially in the life sciences.

Diana Kwon

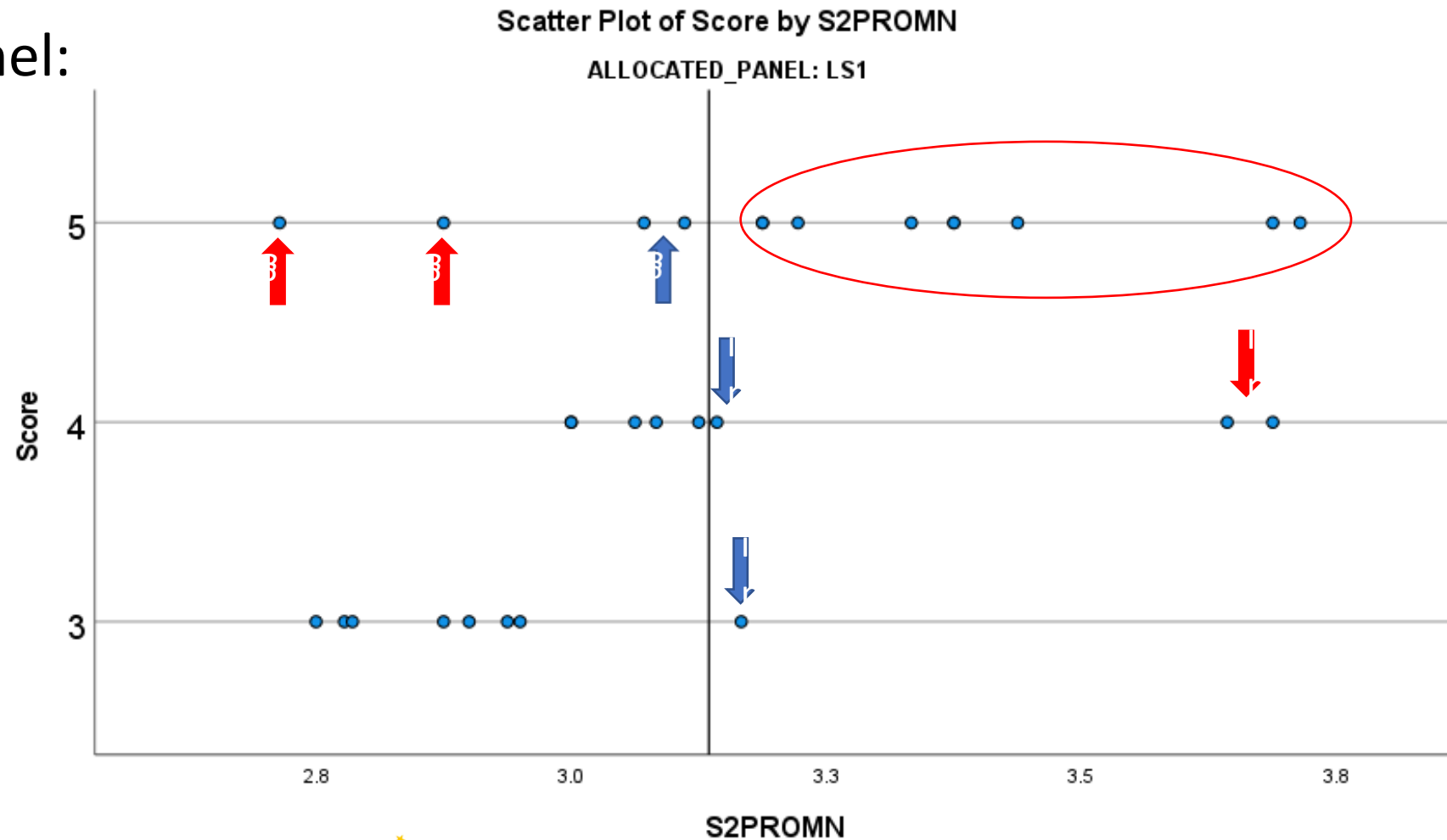
# What predicts grant success?

## Personal characteristics

- Age (-)
- Academic age (+)
  
- Self presentation – interview by panel (?)
  - More reliable information?or
  - Stage performance versus academic quality?

# What predicts grant success? Self presentation

- Interaction with the panel:
  - Presenting the proposal
  - Discussion with panel
- How to investigate?
  - Observations
- Stage performance?
  - Training required!!



# Research question: has the writing style an effect on the scores and success?

- Panel observations: difficult to get access
- Alternative: self presentation in written text – the grant application
  
- Clear writing
- Characteristic content elements



# Research on grant proposal texts (1)

## 1. Readability

- Text characteristics
  - Various tests – but why these characteristics?

## 2. Grant proposal as a genre

- Necessary content elements ('Rhetorical moves')
  - Stating the goal
  - Describing the knowledge gap
  - Showing competence of applicant, etc.

## 3. Differences between disciplines

- Variation in linguistic properties
  - Analytic or narrative style
  - Explicit argument needed?

# Research on grant proposal texts (2)

## 4. Context

- Genres embedded in social context
- Have grant proposals and scientific articles the same characteristics?

## 5. Expectations of readers

- What should be in the text?
- **How should it be presented to be convincing?**
  - Focus not on content words but on function words
  - E.g., the style shows competence, not explicitly describing it

.

# Research on grant proposal texts (3)

- Most studies:
  - (Very) small N
  - Often not real life
- Some exceptions
  - Pennebaker (several 10.000 admission essays)
  - Markowitz (19.000 abstracts of grant applications)
- Our dataset
  - Large set of applications (3.000), granted and non-granted
  - Full application texts, including abstract, project description and CV
  - Native and non-native speakers

# Writing style and expectations of reviewers

## A psychological approach

- Good scientists = analytic thinking => analytic writing
  - Pennebaker et al., PlosOne, 2014
- Good scientists should show in text style:
  - Clear writing
  - Complex and analytic thinking
  - Self confidence
  - Confidence in proposal
    - Markowitz, JLSP, 2019
- Text **style** shows cognitive competences and characteristics

# Writing style operationalizations (& hypotheses)

- Clear writing
  - High score on **common words** (dictionary): low on jargon
- Complex and analytic thinking
  - **Long text**, but not too long; **long sentences** – but not too long
  - Higher score on **analytic writing**
- Self confidence
  - High score on **certainty terms**; (low score on **tentative terms**?)
- Confidence in proposal / claims
  - High score on **causal terms**
- Controlling for **English proficiency**
  - Non-native speakers
- Controlling for **performance**
  - Good science =?= good writing
- Controlling for gender
  - Clear gender differences in writing style
  - To be done

# Analytic thinking

(Pennebaker 2011; 2014)

- Analytic writing ↔ Discursive writing
  - Focus on objects and comparison ↔ Story telling
- Analytic writing ↔ Articles + Prepositions
- Discursive writing ↔  
Pronouns + Conjunctions + Negations + Adverbs + Auxiliary verbs
- Continuum: Analytic writing = Analytic writing – discursive writing

# Examples

- Analytic terms
  - **Articles**: a, an, the
  - **Prepositions**: all, below, much
- Discursive terms
  - **Pronouns**: I, us, yours, they, it, this, anything
  - **Conjunctions**: and, so, until
  - **Negations**: no, never, not
  - **Adverbs**: even, just, usually
  - **Auxiliary verbs**: are, did, have
- Certainty terms:
  - **Certainty**: absolute, definitive, essential
  - **Tentative**: if, hope, guess, indirect, hypothetical, depending
  - **Causal claims**: infer, intend, solve
- Common words
  - Words in the dictionary **≠ technical terms**

# Case

- ERC starting grant 2014
  - Prestigious grant for early career researchers (up to 7 years after PhD)
  - Almost all fields (N=3030)
- Success rate between 9% and 14%
  - Differs between panels
- Two step procedure
  - In first step, 75% are rejected, in second step 50% wins
  - In the second step is an interview/presentation



# Data

- Grant proposal = PDF file
- Converted into text format
- Cleaning

# Differences between texts

- Splitting file in
  - Abstract
  - CV
  - Project description
  
- Results may be difference for the three parts (Feng & Shi 2004):
  - Different texts = different functions
  - Different texts = different structure
  - Lists in the CV

# Tool: LIWC

- LIWC = Linguistic Inquiry and Word Count
- Processed with LIWC tool => XLS file with for each text (2900 \* 3) =>
  - the share of words belonging to the linguistic categories (“3.5% are pronouns”)
  - for some categories: counts (WC, WpS)
  - for some an index value (composite categories like ‘analytic language’)

	CASEID	S e g.	WC	Analytic	Clout	Authenti c	Tone	WPS	Sixltr	Dic	function	pronoun	ppron
1	433117	1	508	98.62	47.64	32.69	29.05	28.22	29.13	63.78	35.83	3.35	1.57
2	433120	1	317	95.85	52.52	25.42	49.24	26.42	35.33	60.88	30.60	2.52	.00
3	433133	1	382	98.51	58.29	12.63	30.17	29.38	40.05	62.83	35.08	2.36	1.31
4	433136	1	529	96.60	53.78	24.31	80.56	29.39	41.97	64.08	33.27	2.08	.19
5	433156	1	354	97.62	50.00	28.07	81.69	19.67	39.55	56.21	28.81	2.82	.28
6	433157	1	326	94.45	53.68	17.26	42.47	32.60	39.57	65.03	33.74	2.76	.92
7	433177	1	345	97.10	44.24	25.78	17.36	38.33	33.91	64.64	39.42	5.51	1.74
8	433198	1	763	96.60	65.30	12.70	52.83	22.44	34.34	71.43	43.38	4.33	1.97
9	433205	1	371	89.46	55.35	41.29	71.33	30.92	34.23	71.70	42.86	5.93	2.96
10	433212	1	381	96.39	59.33	14.67	34.91	27.21	37.53	58.27	35.17	4.20	1.57
11	433277	1	354	90.01	46.61	16.91	46.60	22.12	42.09	65.54	36.72	3.39	.85
12	433294	1	412	98.83	62.39	14.16	38.73	51.50	35.19	58.50	36.89	3.16	1.46
13	433297	1	387	95.73	87.75	51.26	60.05	24.19	33.85	70.28	41.09	8.27	6.46
14	433309	1	363	99.00	56.57	8.27	25.77	33.00	37.74	69.70	37.19	1.93	.55
15	433326	1	346	94.60	54.60	23.99	64.13	20.35	39.60	69.65	37.86	5.78	.87
16	433409	1	360	97.80	58.78	24.94	17.67	36.00	39.44	61.39	33.06	3.33	1.11
17	433568	1	362	88.25	70.98	30.59	62.44	30.17	40.88	66.02	35.91	5.80	1.93
18	433573	1	440	97.59	70.74	5.85	37.85	23.16	38.64	65.91	37.95	5.45	2.05
19	433643	1	429	97.38	65.42	11.37	86.40	39.00	33.33	67.60	36.13	2.80	1.40
20	433812	1	352	97.29	52.27	17.28	21.41	32.00	34.38	62.78	28.69	3.41	2.27
21	433843	1	322	96.70	58.59	22.46	66.89	26.83	44.10	54.04	28.57	2.48	.62
22	433852	1	469	96.37	67.28	36.80	25.77	27.59	34.75	69.72	38.81	6.18	2.56
23	433943	1	314	99.00	53.82	22.97	67.88	26.17	39.81	68.15	34.39	2.23	.32

June 15, 2021

# Model

- Independent variables:
  - Word count (squared)
  - Words per sentence (squared)
  - Analytic language
  - Common language
  - Certainty terms
  - Tentative terms
  - Causal terms
- Control variables
  - English proficiency
  - Performance
  - (Gender)
- Effect variables
  - Score for the project
    - (between 1 and 4)
  - Final score
    - (5 values)
  - Decision
    - (yes/no to step 2)
    - (yes/no grant)

# Methods (3)

- Mixed models regression
  - Predicting the project score
  - Panel as random effect
- Ordinal regression
  - Predicting the final score
  - Panel as categorical variable
- Logistic regression
  - Predicting the success
  - Panel as categorical variable

# Findings

# Project score by linguistic variables

(project text, step 1, mixed models, panel as random effect, standardized at panel level)

variable	Expected effect	Found effect
Word count	+	+
Word count squared	-	-
Words per sentence	+	+
Words per sentence squared	-	-
Analytic thinking	+	-
Common words	+	-
Certainty terms	+	+
Tentative terms	-	0
Causal terms	+	+
English proficiency	+	+
Past performance	+	+



# Similar outcomes for all step 1 outcomes

- **Project score by linguistic variables**
  - project text, step 1, mixed models, panel as random effect
- **Final score by linguistic variables**
  - project text, step 1, ordinal regression, Nagelkerke Pseudo  $R^2 = 0.146$
- **Decision by linguistic variables**
  - project text, decision in step 1, logistic regression, Nagelkerke Pseudo  $R^2 = 0.1$

# Other findings

- The analysis of the **Abstract** text and the **CV** text show a similar pattern, although the effect are (significantly) smaller.
- The effects found for step 1 **disappear in step 2**: obviously then language does not play a role anymore

# Overall outcome

- Word count and words per sentences:
  - positive effect of text complexity
- Certainty and causal terms:
  - positive effect of certainty about self and about the proposal
- Tentative terms:
  - No effect. Why => double role of tentative terms
- Common words:
  - Negative effect: proposal better has more technical terms
- Analytic thinking/writing
  - Negative effect: proposals with better narrative work better
- English proficiency
  - Positive effect
- Past performance
  - Positive effect

# Issue 1: lower level of analytical writing?

- Unexpected result - further analysis needed:
- All score high to very high on analytic writing
  - (Between 95 and 100 on scale 0-100)
  - Combining analytic and narrative writing to one scale is probably wrong
- So in fact: strong analytic writing PLUS good narrative

# Issue 2: Positive effect of past performance

- Good scientist =?= good writing style?
- Adding past performance to the regression does not influence the effect size of the linguistic variables

# Conclusions

- *Better data (rejected applications included) and different texts* lead to **different findings** than in the literature
  - complex (longer text and longer sentences),
  - strong analytic writing combined with
  - a narrative style,
  - and strong technical content works
- Strong confidence (*certainty terms*) and claims (*causal terms*) are effective
- Effects are strongest in Step 1 and in the project text.
- Moderate pseudo-R<sup>2</sup> of 0.146 (final score), and 0.1 (decision), so presentation performance has a substantial effect.

# Implications?

- Can one deliberately use the right linguistic properties to increase the success probability?
- If presentation / writing style is important: role of training and writing support
- Another level of uncertainty in grant allocation

# Thanks for your attention

Paper: [\(21\) \(PDF\) Writing style and research grant success – a linguistic analysis \(researchgate.net\)](#)