The impact on candidates of examiner interventions in oral interview tests

Fumiyo Nakatsuhara & John Field

CRELLA

University of Bedfordshire

CRELLA
University of Bedfordshire

Acknowledgement

- *This presentation draws upon reports on a research project funded by Trinity College London, and carried out under the Trinity Funded Research Programme.*

- *Any opinions, findings, conclusions, or recommendations expressed in this material are those of the presenters and do not necessarily reflect the views of Trinity, its service providers, examination centres and/or distributors.*

CRELLA
University of Bedfordshire

Research Background

Examiner-candidate interaction

- **Examiner variability and its impact on test performance**
  - The scores awarded in interview tests seem to be collaboratively achieved through interactions driven by the interviewer's discourse (e.g. Lazaraton, 2002; Brown, 2003)

- **Issue of training and standardisation of interviewers**

[High stakes exams]

Every possible effort is usually made to minimise interviewer variation (Taylor, 2003)

⬅ *CAVEAT*

The effect of standardisation of interviewer protocols may 'potentially threaten the ... validity of the procedure and... [influence] the interaction and discourse in ways that are detrimental to candidates' (Ross, 1988)

The role of listening in oral interview tests

- **Seedhouse & Egbert (2006)**
  - Interactional problems can be caused by test-takers' *misunderstanding* of what the examiner has said

- **Nakatsuhara (2012)**
  - Candidates' **listening test scores** were compared on a monologue task and on an interview task
    - Significant effect on **Fluency** scores
    - **Communication problems** related to candidates' limited listening proficiency

Oral interview tests are to some extent tapping into the construct of listening-into-speaking i.e. interactive listening skills

Relevance of issues to Trinity GESE (Graded Examinations in Spoken English) exams

- The GESE exams aim **to "replicate real-life exchanges** in which the candidate and the examiner pass on information, share ideas and opinions, and debate topical issues"; **to assess both listening and speaking skills** through communicative interaction (Trinity College London, 2009).

- The tasks across the 12 GESE levels are designed to extend the demands placed upon **the candidate as a listener**.

- The GESE examiners have **more freedom and discretion** than in some other tests in relation to how they intervene in the interactions.

  – They do not follow a strict interlocutor framework, but are instructed to produce **a test plan for 'natural interventions' that meet the language specifications of the grade** **More authentic!**

In order to understand the listening demands placed upon candidates, we need a greater knowledge of the linguistic and pragmatic content of the examiner interventions.

Research Questions

- *RQ1:* What types of examiner intervention are employed in the GESE examinations in terms of their linguistic and discourse features?

- *RQ2:* To what extent do examiner interventions differ in relation to the proficiency level of the candidates?

Research Design

Method

Transcription and analysis of audio-recordings obtained by Trinity for standardisation purposes

- 20 candidates at **Grade A** (AAA: Distinction)
- 20 candidates at **Grade C** (CCC: Pass)

Interviewed by the same 20 examiners

  – Candidates graded A and C examined by the same examiner: as similar as possible in terms of L1, age, perceived communicativeness and gender

  – Demographic info: as representative as possible of the current test-taker population sitting the GESE Grade 7

Level and Tasks

GESE Grade 7: chosen as a representative intermediate level test (CEFR B2) with a large number of applicants

Tasks

Phase		Time
1	Candidate-led discussion of a topic prepared by the candidate	5 mins
2	Interactive task	4 mins
3	Conversation on two subject areas selected by the examiner	5 mins

Data Analysis

Basis: Socio-cognitive framework for validating speaking tests
(Weir, 2005; further elaborated in Taylor, ed. 2011)

- **Stage 1:** Identify turns that relate to examiner interventions and candidates' response to these interventions

- **Stage 2:** Select **contextual parameters** in the spoken input and analyse the examiner interventions for:

  1) **Lexical complexity**
  2) **Syntactic complexity**
  3) **Informational density**
  4) **Number and mean length of interventions**
  5) **Speech rate (only for Phase 2 prompting interventions)**
  6) **Purpose for interventions**

Measure(s) for the selected parameters

1) Lexical complexity: Ratio of the first 2000 and off-list words in the BNC

2) Syntactic complexity: Number of sub-ordinate clauses per AS unit, Number of verb elements per AS unit

3) Informational density: Lexical density (content words / total words)

4) Number and mean length of interventions: Number of interventions, Mean length of interventions, Number of words

5) Speech rate: Articulation rate, Number of pauses and total pause time

6) Purpose for interventions: Language Function list: O'Sullivan et al, 2002
Informational (11 sub-categories),
interactional (15 sub-categories) and
Interactional management (4 sub-categories) purposes

Focus of investigation

a. Types of intervention across 3 phases (RQ1)

b. Variation in interventions between examiners (RQ1)

c. Variation within examiners in relation to proficiency level (RQ2)

Results

a) Types of intervention across 3 phases

Lexical complexity, Informational density

- Almost identical across 3 phases

Syntactic complexity

- More complex in Phase 3 (Conversation), followed by Phase 2 (Interactive) and Phase 1 (Topic) [Sig.]

a) Types of intervention across 3 phases (cont.)

Number and mean length

- Phase 1 (Topic): shorter interventions

- Phase 2 (Interactive): less frequent but longer interventions

- Phase 3 (Conversation): more frequent and longer interventions

Phase	N of intervention	N. Of words / intervention	N of words in total
1 (Topic)	17.5	9.5	155.5
2 (Int.)	16.5	12.3	209.0
3 (Conv.)	19.5	11.3	221.0

✓ **Congruent with the test specifications**
-Phase 1: Examiner interventions mainly serve to facilitate the candidate-led discussion of a topic prepared by the candidate
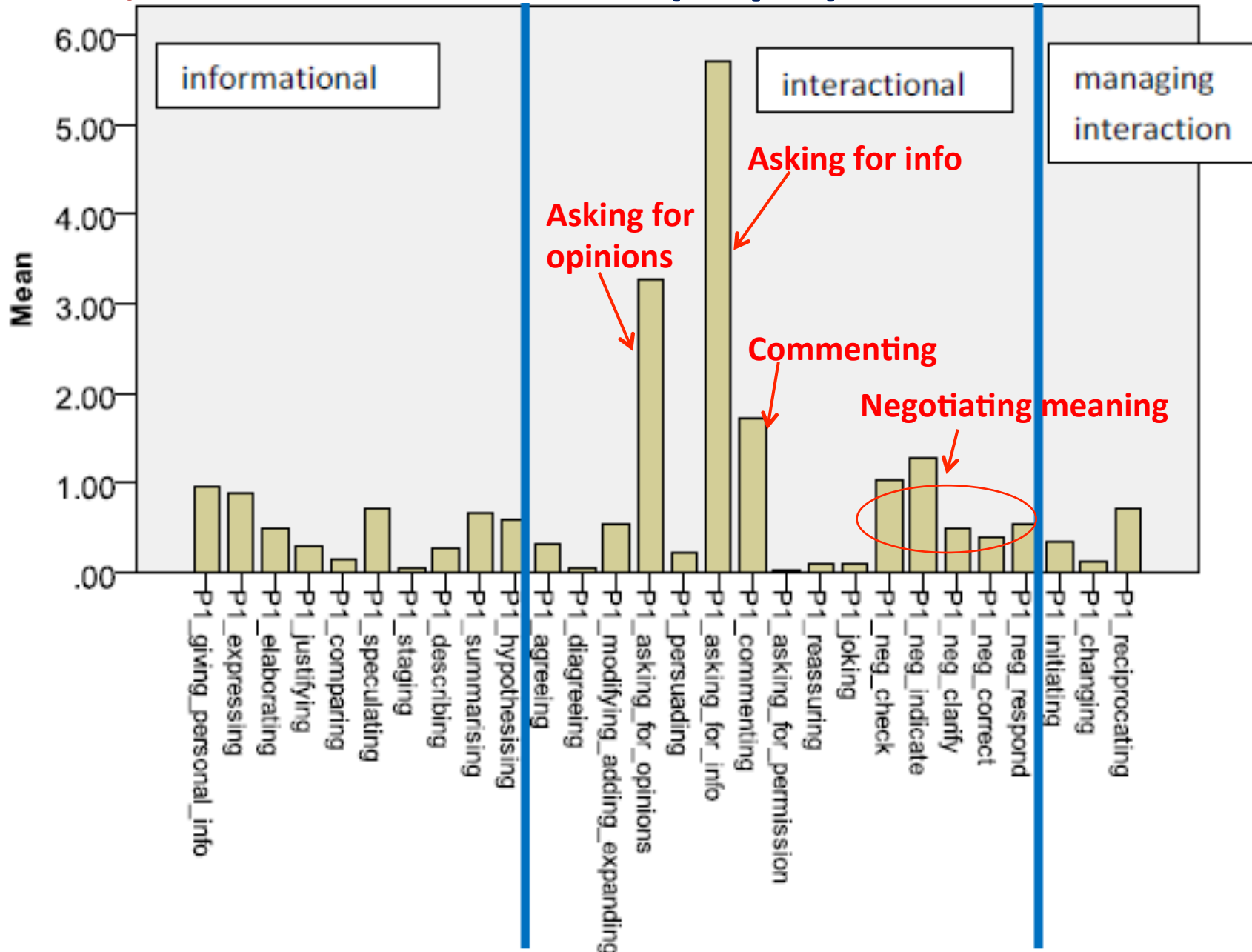**-Phase 2:** It is essentially the candidate's responsibility to initiate and maintain the discourse, and examiners respond to the candidate's questions
-Phase 3: Examiners are required to take a lead in discussing two topics
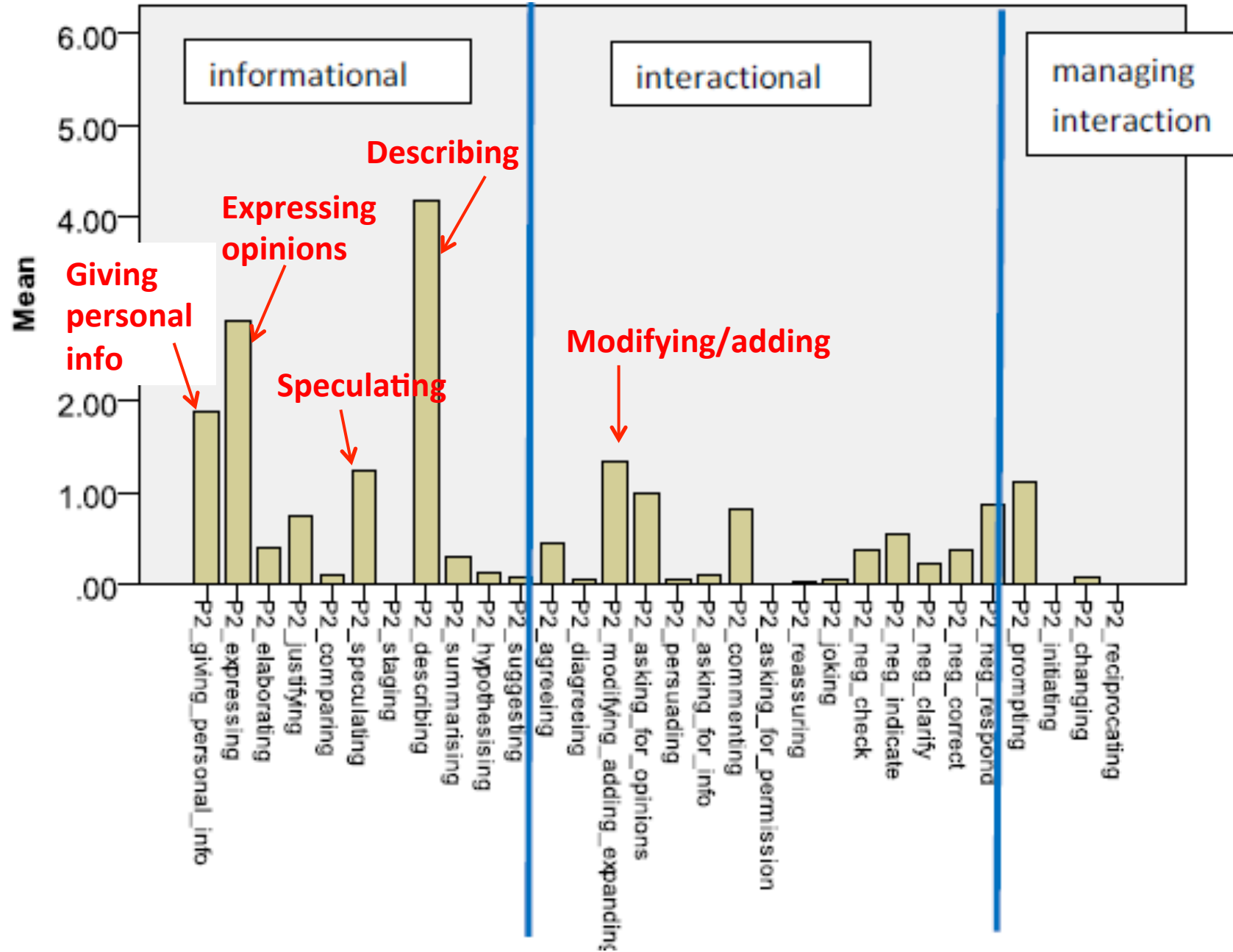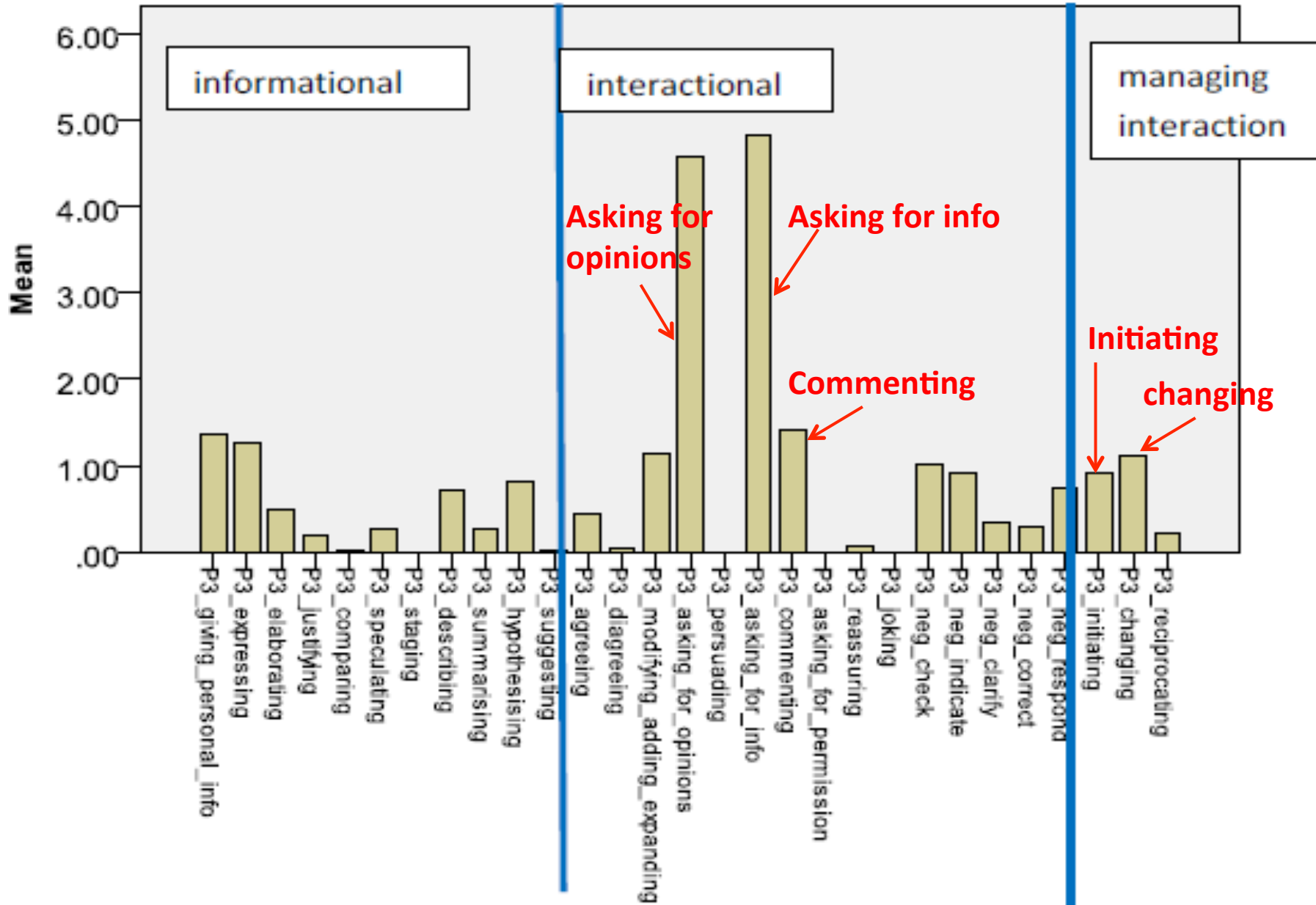
Purpose

Phase 1 (Topic)

informational

interactional

managing interaction

Asking for opinions

Asking for info

Commenting

Negotiating meaning

Purpose

Phase 2 (Interactive)

Purpose — Phase 3 (Conversation)

Purpose

✓ The data confirms that the test includes a wide range of types of intervention purpose

→ **a variety of pragmatic functions that the listener has to interpret.**

b) Variation between examiners

Lexical complexity, Informational density, Speech rate

- Little variation

Syntactic complexity, Number and mean length, Purpose

- Some variation

[Purpose]

Some interventions appeared to be somewhat more complex to interpret, due to ways in which some language functions were realised (Green, 2012 *'Language Functions Revisited'*).

e.g. Hypothesising (a lack of context prior to hypothesising)

E: if if you had children and they didn't want to go to school what would you say to them?

c) Variation within examiners in relation to proficiency level

Syntactic complexity, Informational density

- No difference

Lexical complexity, Number and mean length, Speech rate

- Interventions tended to be a bit more lexically complex, more frequent and longer, with fewer pauses for Grade A students than for Grade C students [but NOT sig.].

Purpose

- **Grade A students** with more interventions for:
  - Expressing opinions;
  - Speculating;
  - Describing;
  - Agreeing;
  - Commenting;
  - Negotiating meaning (indicating understanding)

Examiner's greater participation in the interaction

- **Grade C students** with more interventions for:
  - Asking for information;
  - Negotiating meaning (correcting an utterance made by the candidate);
  - Negotiating meaning (responding to requests for clarification)

Keeping the conversation going

Conclusions

Main Finding 1: Phases

The experience and expertise of the GESE examiners assisted in differentiating interventions **across the 3 phases of the test** in terms of:

- syntactic complexity
- number and mean length
- purpose

in ways that are congruent with the GESE task specifications

⟹ This validates the Trinity argument that the 3 phases of the test involve different roles for the examiner, and engage the candidate listener to different degrees.

Main Finding 2: Examiner variation

The data showed some variation between examiners in relation to:

- syntactic complexity

- number and mean length

- purpose

But some characteristics of the interventions were consistent across administrations:

- **lexis**

- **informational density**

Main Finding 3: Sensitivity to level

Some examiners showed sensitivity to candidate level by adjusting their interventions in terms of:

- number and mean length

- purpose for intervention

- speech rate of Phase 2 prompts

❖ This suggests a recognition of the different needs of candidates at Levels A and C during the interaction.

❖ It also indicates an awareness of differences in candidates' listening levels, and willingness to adjust the listening demands of interventions to the perceived level of the candidate.

The issue of training and standardisation of interviewers
Lazaraton (2002: 151-152)

'Variability in behaviour is frequent ... Using an interlocutor frame, monitoring interlocutor behaviour, and training examiners thoroughly are all ways to reduce, or at least control, this variability.

It is unlikely, however, that it would be possible, or even desirable, to eradicate the behaviour entirely, since 'the examiner factor' is the most important characteristic that distinguishes face-to-face speaking tests from their tape-mediated counterparts.

Yet, we should be concerned if that factor decreases test reliability, even if it appears to increase the face validity of the assessment procedure.'

Trinity's approaches to addressing this issue
• **Monitoring:** Making very constructive use of audio recordings of live tests for the purpose of monitoring and standardisation of the examiners
• **Research:** Commissioning research to find out how we can grade more finely the listening demands imposed upon candidates by examiner interventions without losing the 'human' factor in the interaction!

Thank you! ☺

Fumiyo.Nakatsuhara@beds.ac.uk / fumettina@gmail.com
John.Field@beds.ac.uk

The 9th Annual EALTA Conference
31 May - 3 June 2012 Innsbruck, Austria