Mixing many-facet Rasch measurement with qualitative methods to identify rater influences on oral proficiency judgements

Thomas Rogers

Centre for Research in English Language Learning and Assessment,
University of Bedfordshire

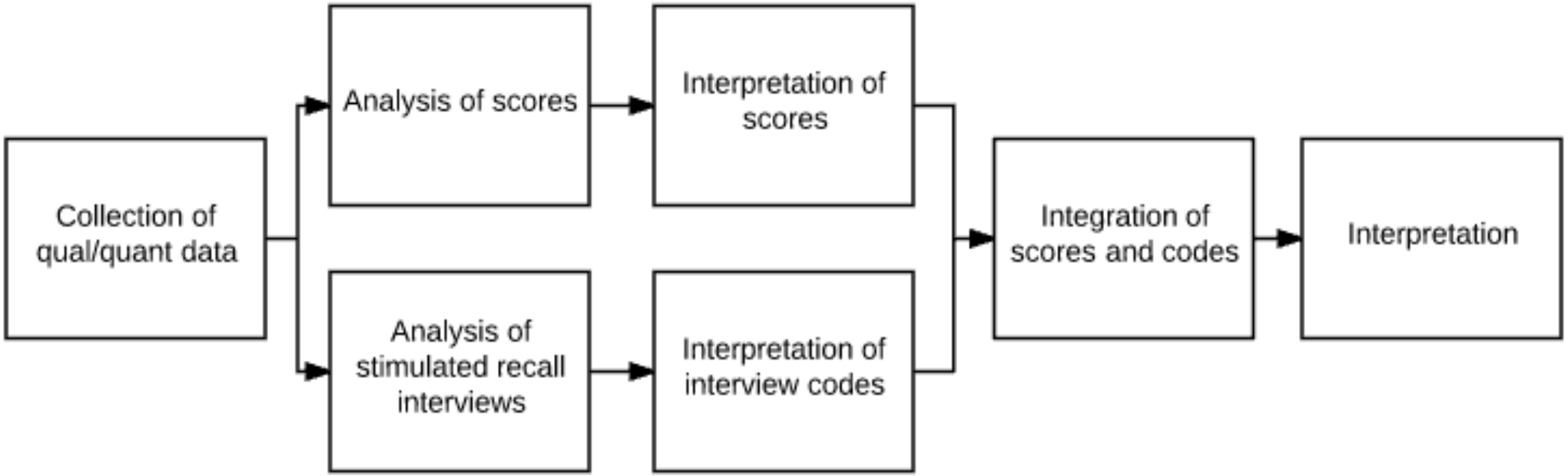
This presentation reports on research using examination data provided by Cambridge English Language Assessment

Purpose

- How raters respond to suprasegmental features of speech
  - Stress – relative salience
  - Rhythm – regularity in delivery
  - Intonation – pitch change over the utterance
- Impact of rating experience

Approach

- Pragmatic mixed-methods – workable solutions, real-world practice, outcome oriented



Procedure

- 12 audio recordings (1 minutes monologue) English language speaking tests
- 30 raters (15 experienced, 15 non-experienced)
- Scores:
  - 9-point intelligibility scale
  - Stimulated recall procedure

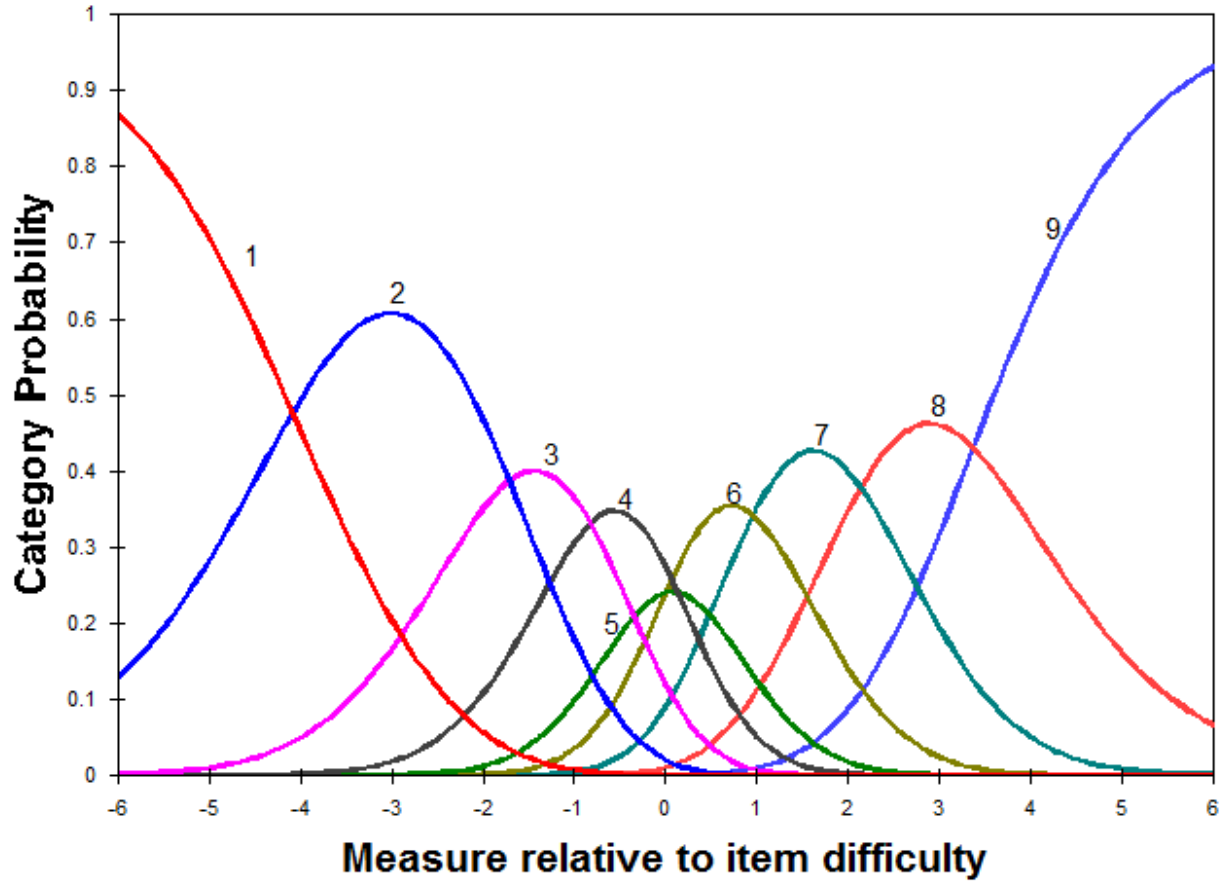
Mixed analysis

- Extreme case analysis
  - explore the qualitative response of quantitatively extreme raters or instances of rating

- Integrated display
  - connect score to rationale visually by populating a matrix

Disorder around the middle of the scale

Category (score)	Total	Average Intelligibility Estimate (logits)	Outfit Mean Square	Rasch Andrich Thresholds
1	1	-1.39	0.9	
2	20	-1.09	0.9	-4.08
3	48	-0.68	0.9	-1.71
4	64	-0.42	0.8	-0.82
5	46	0.03	0.8	0.13
6	57	0.58	1	0.01
7	52	1.41	1.1	0.96
8	42	2.3	1.1	2.14
9	30	3.61	0.8	3.37



Rater misfit

Audio	Fair Average Score	Leniency estimate (logits)	Infit Mean Square	SR-RR Correlation
NR_15	6.32	0.34	2.16	0.74
NR_04	6.55	0.53	1.93	0.72
NR_02	6.16	0.2	1.79	0.71
NR_06	6.07	0.14	1.57	0.82
NR_11	5.02	-0.55	1.43	0.89
NR_05	5.14	-0.48	1.38	0.84
NR_13	5.48	-0.26	1.31	0.8
ER_13	5.57	-0.21	1.3	0.85
ER_06	6.61	0.59	1.27	0.64
ER_09	4.91	-0.62	1.22	0.77
ER_08	5.57	-0.21	1.15	0.82
NR_12	6.16	0.2	1.11	0.64
ER_14	5.36	-0.34	1.03	0.76
ER_01	6.99	0.97	0.99	0.7
NR_10	5.14	-0.48	0.89	0.8
NR_07	6.07	0.14	0.8	0.84
ER_15	6.13	0.18	0.77	0.85
ER_05	6.04	0.12	0.75	0.79
ER_07	5.95	0.05	0.61	0.8
NR_03	4.39	-0.94	0.58	0.94
ER_11	6.69	0.66	0.54	0.87
NR_08	4.65	-0.78	0.53	0.92
ER_12	7.14	1.13	0.51	0.81
ER_03	5.86	-0.01	0.51	0.9
ER_04	4.2	-1.06	0.47	0.95
NR_01	6.4	0.4	0.45	0.93
NR_14	5.98	0.07	0.43	0.89
ER_02	6.99	0.97	0.39	0.83
ER_10	5.67	-0.14	0.37	0.93
NR_09	4.9	-0.62	0.25	0.95

Extreme cases –misfitting raters

One rater infit > 2, three raters infit > 1.5

- all non-experienced
- apart from one rater they were as sensitive to suprasegmental as other non-experienced raters
- idiomatic rating behaviour in terms of:
  - Accent familiarity
  - Rating approach
  - Focus on non-pronunciation features
  - Emotional engagement with the speaker
  - Identifying unique features

Idiomatic approaches of misfitting raters

	NR_02	NR_04	NR_06	NR_15
Familiarity		I thought again the familiarity of the accent on this one made it easier I think it's French and I come across French people reasonably frequently and I did French at school . I think that made it easier . there's definitely quite a few mistakes in there in how she pronounces words and where she puts stress in words . erm but I think because from the beginning you get into that accent and it's familiar and I've heard people talk about it before		she reminded me of somebody in my office who speaks in a very similar way
Rating approach	she is quite difficult to understand . erm . because somehow she's blurring the line between ice and eyes . erm not quite . I'm not sure how to describe what's happening but . erm . it's enough of a glossing over of the sounds			I don't know what the marking scheme is for this sort of test
Unique features				depth' and 'warmth' of a speaker's voice
Non-pronunciation features	he pluralises a couple of things where he shouldn't . or his grammar isn't very good in other places . I understand what he's saying because his speech flows /decently ((uncertainly)) well I mean it's not great (.) but I think that helps	she's got more vocabulary to draw on than to form her sentences and she's able to not just describe but I think at the end she's also then starts putting in an opinion about it being better to play on a team than to play individually so . she's got enough language to talk around a topic more than just describe what she can see	yea (.) I think it's like if it was written down I'd be wanting . I'd be with my red pen . putting the a's and the the's ((plurals of 'a' and 'the')) and the things and I think that's one of the issues again with this one and because there's words missing and the thinking process is going on while you're talking it's not flowing	lack of basics of the English language I think . verbs adjectives nouns . there was just a core base of the English language that was missing from her vocabulary . it just wasn't there . which prevented her from really providing a reasonable description
Emotional engagement	this one drove me nuts . she's she's adding extra syllables . err . which is aggravating . every . every time she's making a declaration she's going /up . it's in the /city . it's in a /building ((imitating test taker's intonation)) . and it's very difficult to follow even though . when she's pronouncing these words I can understand the individual words but overall it's hard to understand her because of the way she's speaking […] well it's distracting	"I found her really easy to understand and it's not necessarily because she's better than some of the others but there's something about her voice that's quite lilting and makes it joined up because it feels flowing . I feel like I can imagine this person . I feel like she gesticulates quite a lot while she's talking . erm and . yea there's something about the pace and just . there's a natural flow to her voice		

Extreme cases – raters assigning fives

- Score of 5 assigned 46 times
  - Broad spread of 5s assigned (mean number of times a rater assigned a 5 = 1.5, min: 0, max: 3), so they are not all coming from a few errant rater
  - Qualitative data suggests raters opted for five when:
    - raters grappling with the broad range of ability within one performance, some contradiction
    - slightly more irritation

Contradiction

NR_05: the pronunciation of the words and the order of the words all seems clear and correct . it’s just the hesitancy and the correcting herself all the time makes it difficult to understand when she speakers for a long time . but she did get a lot better towards the end . so maybe not actually

[…]

NR_05: she didn’t pause . she didn’t stop and correct herself

Integrated display

- place candidate performance against score
- test takers grouped by fair average scores
- patterns among each group identified
- illustrative quotes from the transcripts placed on the display

...what to do about misfitting raters?

Display Groups

Score categories		
Category	Fair average	Audio clips
3	>0.29 <0.3	Audio_01 Audio_09 Audio_06
4	>0.39 <0.4	Audio_12 Audio_08
5	>0.49 <0.6	Audio_02 Audio_05 Audio_11
6	>0.59 <0.7	Audio_04
7	>0.69 <0.8	Audio_03 Audio_14
8	>0.79 <0.9	Audio_10

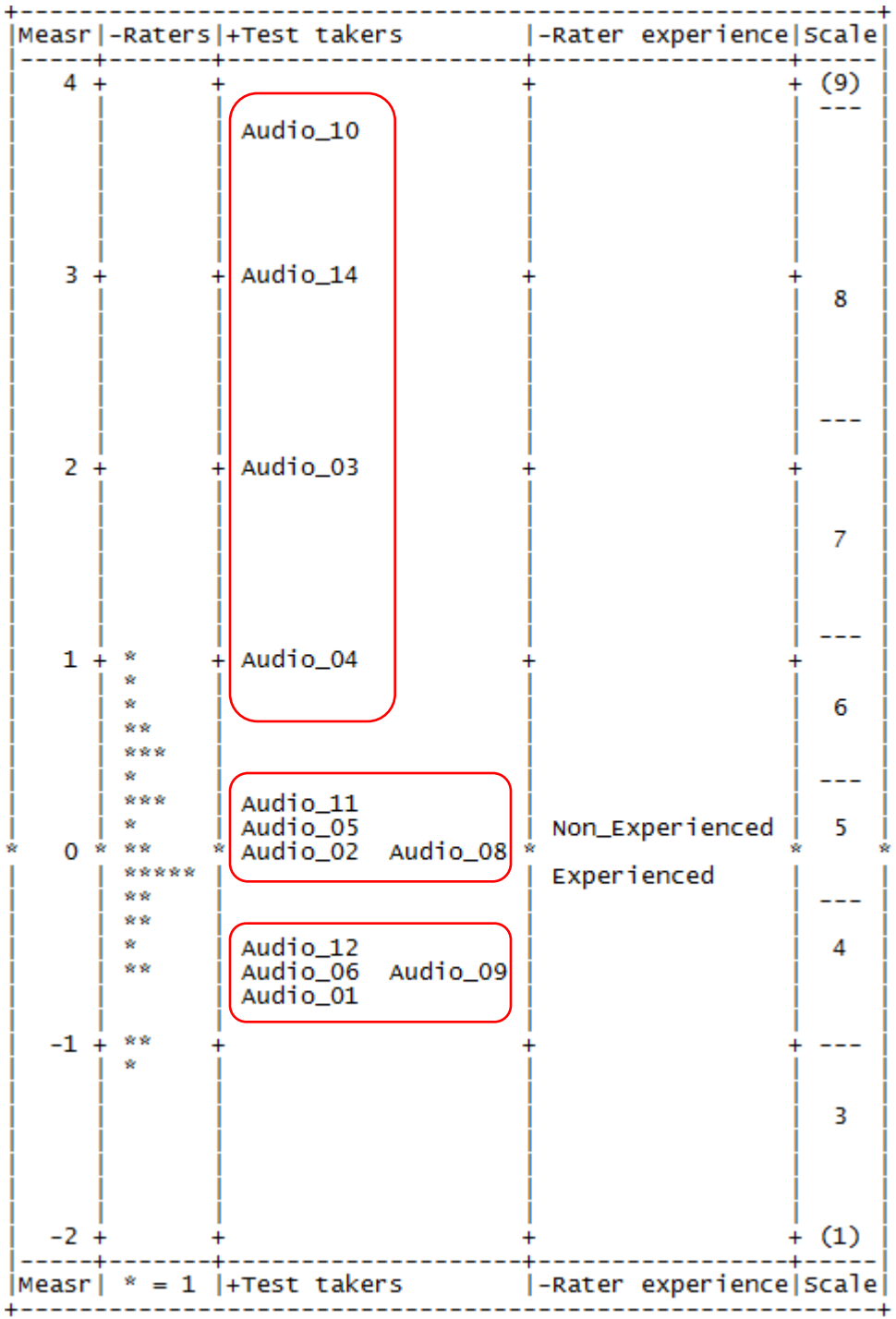
Test taker fair average scores	
Test taker	Fair Average
Audio_01	3.65
Audio_09	3.87
Audio_06	3.98
Audio_12	4.06
Audio_08	4.73
Audio_02	5.05
Audio_05	5.09
Audio_11	5.17
Audio_04	6.35
Audio_03	7.28
Audio_14	7.96
Audio_10	8.43

Category	Stress	Rhythm	Intonation
3		"there's almost a kind of a like a sort of staccato little breaks in that speed erm . it's sort of quite jittery" "he seemed to stagger it with many pauses [...] several pauses interrupted the natural flow of his speech" "she was really struggling with a lot of hesitation and so on . maybe just nervousness but I also feel that erm she didn't have the capability"	"it was quite a flat intonation really . there was no rise or fall in intonation" "it was quite monotonous . there wasn't much going on in terms of the sentence so I was listening to the words individually rather than the whole sentence" "it was that flat intonation that really made it hard [...] that sort of firing at you . yea but not giving you any clues as to the attitude or meaning really"
5	"there's different with vowels and the internal stress of the words makes some of them more difficult to understand" "the intonation of the word for example prefer instead of prefer" "I suppose it breaks up the rhythm a bit if you say a lot of /æ lɑt ɑv/ rather than a lot of /ə lɑt əv/ whatever" "occasional inappropriate stress"	"she's got a habit of adding a syllable on to each word er which is quite common erm so once I got used to that it was fine it was a bit choppy . but still it wasn't monotonous so I found listening to her was quite pleasant it had a nice rhythm going on a nice cadence ." "she's able to pronounce these words (.) I mean (.) family rain child ((listing)) she can pronounce those very clearly (.) but because of the way she's speaking it's so (.) fragmented and (.) disrupted by er the pauses that she's taking (.) she's clearly thinking very carefully about how she's pronouncing the words (.) but by doing that (.) by breaking up the speech and her speech is going up constantly it becomes very difficult to understand "quite a bit of hesitancy and fragmentation as well wasn't there which again would make it hard for people to follow"	"unnatural rising intonation [...] rather listy rather than connected speech" "I found and I think other people would find the rather sing song intonation very unnatural . but [...] once you get a bit used to it you can . okay she's got sing song intonation I'll get on with the rest and try and understand it" "there's kind of a . like a bounce . I don't know quite how to describe it . but it kind of goes up and down more than English does I think . English is flatter"
8	"there's very little incorrect sentence and word stress " "the word stress was correct" "she's got stress with meaning helping us understand there . she's very much in control of what she's saying and therefore that takes the listener along with her"	"she pauses in just the right places so she's really got the rhythm of the language down to a tee" "hardly hesitating . maybe there was just a little"	"the inflection is almost like an English person " "very nice stress . very nice intonation " "the intonation kind of guides you through where you are in the sentence ." "it's more of a natural cadence (.) like the other one I mentioned before (.) so it's not all going /up (demonstrating) she's got some declarative (.) down intonations there "

Six-point display summary

Category	Summary
3	Very strong agreement that rhythm is very broken, intonation is very monotonous.
4	Intonation still monotonous but some sense that the speaker is beginning to be able to apply intonation. Rhythm still broken but weaker agreement among raters on this.
5	Effective rhythm beginning to emerge but in a limited manner. Irregular or unexpected tones influencing perception of rhythm.
6	Some persistent hesitation influencing control of rhythm.
7	Occasional reference to errant intonation although large numbers of raters are now beginning to describe the discourse as natural or ‘native-like’.
8	Intonation and rhythm broadly appropriate with effectively placed stress. Some hesitation persists.

Variable map



Category	Stress	Rhythm	Intonation
Category 1 **Audio_01** **Audio_06** **Audio_09** **Audio_12**	"he's not stressing the words the sounds the individual sounds" "some strange stress patterns" "it just doesn't seem natural . the individual words are stressed incorrectly" "some really odd stresses and pauses in the way she was speaking" "by being monotone there's no emphasis going on" "the stress wasn't quite how I would expect it to be"	" so it's quite stilted" "there's no sort of English rhythm to it is there " "yea there was some attempt to emphasise the key words [...] it was kind of regular there was an attempt to stress the main words [...] fairly fragmented that would put a bit of strain" "it's sort of a jagged flow with uneven rhythms and I suppose those patterns are transferred from his native language" "it was very staccato and there were a lot of umms and ahhs in it which really interfered" "she was really struggling with a lot of hesitation and so on . maybe just nervousness but I also feel that erm she didn't have the capability . and a lot of transfer from her mother tongue" "it's very hesitant it's very slow delivery" "she did have the sort of rather jerky rhythm"	"it's fairly flat sort of staccato" "it was that flat intonation that really made it hard " "there was kind of an attempt to use intonation wasn't there " "very strong transfer of sounds and intonation patterns to the words and utterances" "he is quite monotonous"; "I found his speech very monotonous" " I find that the intonation is a bit off key" "it was quite a flat intonation really . there was no rise or fall in intonation" "so this person speaks in a very kind of monotone . just flat [...] not stilted per say but just very monotone . that was quite unnatural sounding"
Category 3 **Audio_03** **Audio_04** **Audio_10** **Audio_14**	"she's got stress with meaning helping us understand there . she's very much in control of what she's saying and therefore that takes the listener along with her " "in some places the sentence stress helped us . but he didn't have a very very wide range" "the stress is mostly accurately used but she may have for example . 'as far as' . rather than 'as far as' ((demonstrating lack of vowel reduction)) . she was giving it a [L1] stress pattern" "sentence and word stresses is generally appropriate"	"there was a little bit of hesitancy but she always managed to fill it in with an appropriate connective or little discourse marker of some sort to keep it all together" "she was pacing herself very leisurely [...] hardly hesitating . maybe there was just a little" "he's breaking it up in quite a natural way isn't she . erm . it's not broken at individual word level or she's not struggling for words" " she pauses in just the right places so she's really got the rhythm of the language down to a tee"	" I think she mirrors very well English language intonation so that helped just . didn't have to pay any great attention . she annunciates very well it's very clear . the cadence" "a kind of unnatural intonation pattern for native speaker listeners because the [L1 interference]" "there wasn't a lot of range of intonation . but there was enough . we had some nice intonation with meaning" "she had quite good weak forms and appropriate linkage though it was the rising intonation that was a bit irksome" "native speaker like intonation"

Integrated display summaries

Category	Summary
3	Very strong agreement that rhythm is very broken, intonation is very monotonous.
4	Intonation still monotonous but some sense that the speaker is beginning to be able to apply intonation. Rhythm still broken but weaker agreement among raters on this.
5	Effective rhythm beginning to emerge but in a limited manner. Irregular or unexpected tones influencing perception of rhythm.
6	Some persistent hesitation influencing control of rhythm.
7	Occasional reference to errant intonation although large numbers of raters are now beginning to describe the discourse as natural or 'native-like'.
8	Intonation and rhythm broadly appropriate with effectively placed stress. Some hesitation persists.

Category	Comments
1	Limited rhythm caused by hesitancy, short intonational phrases makes speech monotonous
2	Adequate stress beginning to emerge, hesitancy still causing issues with rhythm. Intonation appropriate at times.
3	Stress is under control. Hesitancy still persists which limits perception of effective rhythm. Much more control over intonation but different raters still perceive it differently.

Limitations

- not face-to-face rating as standard practice
- more female than male raters (63%)
- stimulated recall procedure relies on long-term memory

Questions