TODAY'S GOALS

- What is interactional competence& why is it important?

- How might we assess it?

- How well is our rubric working?

    - Analysing an existing rubric using Many Facet Rasch Measurement

- Recommendations for effective & reliable rubric design

WHAT IS INTERACTIONAL COMPETENCE?

- Interactional competence: the ability to effectively ***co-construct*** an interaction with an interlocutor within a specific context (Kramsch, 1986).

  - *i.e.* involves working together with a partner to achieve a shared goal.

- Central to interaction in a wide variety of real-world contexts.

- Increasingly used in language assessment at university (e.g. Bonk & Ockey, 2003; Leaper and Brawn, 2019; Nitta & Nakatsuhara, 2014) and standardized exams e.g. Cambridge (Galaczi et al., 2011).

- 'Interaction' - one of the 5 spoken production categories in the CEFR scale (Council of Europe, 2001).

- What makes up interactional competence?

  - Interactive listening, topic development – questions, adding information, managing topic shift, dealing with communication breakdowns, inviting contributions, responding appropriately (May, *et al.*)

AN EXAMPLE SCALE (MCDONALD 2018)

Score	Fluency	Accuracy	Complexity	Interaction	Communicative Effectiveness
5	Speaks very fluently with almost no hesitations, false starts, or modifications of attempted utterances.	Errors are barely noticeable.	Confidently attempts a variety of verb forms (e.g., passives, modals, tense, and aspect) even if the use is not always correct. Routinely attempts the use of coordination and subordination to convey ideas that cannot be expressed in a single clause, even if the result is awkward or incorrect.	Confidently initiates and pursues interaction with others. Responds appropriately and naturally to others with adequate detail. May attempt paraphrase or clarification.	Communicates very effectively overall.
4	Speaks fairly fluently with only occasional hesitations, false starts, or modifications of attempted utterances.	Errors are not unusual, but rarely major.	Attempts a variety of verb forms (e.g., passives, modals, tense, and aspect) even if the use is not always correct. Regularly attempts the use of coordination and subordination to convey ideas that cannot be expressed in a single clause, even if the result is awkward or incorrect.	Regularly attempts to initiate and pursue interaction with others. Responds appropriately and naturally to others, though not always with adequate detail.	Communicates relatively effectively overall.
3	Speaks somewhat fluently though hesitations and word-finding delays are not uncommon.	Manages most common forms, with occasional errors, major errors present but not predominant.	Mostly relies on simple verb forms, with some attempt to use a greater variety of forms (e.g., passives, modals, more varied tense and aspect). Makes some attempt to use coordination and subordination to convey ideas that cannot be expressed in a single clause.	Makes some attempt to initiate and pursue interaction with others. Responds appropriately to others, though not always naturally or with much detail.	Communicates somewhat effectively overall.
2	Speaks with a marked degree of hesitation due to word-finding delays or inability to phrase utterances easily.	Demonstrates limited linguistic control: major errors frequent.	Produces numerous sentence fragments in a predictable set of simple clause structures. If coordination and/or subordination are attempted to express more complex clause relations, this is hesitant and done with difficulty.	Makes an effort to initiate and pursue interaction with others, even if unsuccessful at times. Responds to others but often with unnatural pauses and without much detail.	Communicates relatively ineffectively overall.
1	Speech is quite disfluent due to frequent and lengthy hesitations or false starts.	Exhibits a clear lack of linguistic control even of basic forms.	Produces mostly sentence fragments and simple phrases. Little attempt to use any grammatical means to connect ideas across clauses.	Does not initiate or attempt to pursue interaction with others. Provides only minimal responses when prompted.	Communicates very ineffectively overall.

- Analytic rating (Weigle, 2002)
- Alternative approach: Checklists – skills or functions are checked off when used (Wiliam, 2011).

HOW DO WE KNOW IF THE RUBRIC 'WORKS'?

- Rater interpretation, application can vary, raters vary in severity (Eckes, 2015; Myford & Wolfe, 2003; Scullen et al, 2000)

- The rubric scale itself may not be optimally calibrated, e.g. too many scale-points, confusing descriptors, vaguely defined constructs (Janssen et al, 2015; McDonald, 2018).

- Enter Many Facet Rach Measurement (Linacre, 1994)

  - Compares scores to an ideal model

  - Allows comparison of factors, such as demographics, task difficulty, learner performance, rubric functioning or rater severity

THE STUDY

- Aims: To investigate the functioning of an in-tact rubric used to assess the discussion skills of Japanese intermediate learners.

- Speaking tests were formative, assessing skills learned on the course

- 5-minute test, topic known to the students

- 3 video samples of paired speaking tests were rated by 11 raters, (*i.e.* 11 raters scoring 6 performances).

- Analyzed rubrics using Many Facet Rasch Measurement

THE IN-TACT RUBRIC

	EXCELLENT				PASS						
DISC. QUESTIONS	10	9	8	7	6	5	4	3	2	1	0

1. What do you think about?
3. What are the main benefits of___?
2. Should university students___?
4. What are the main problems with _____ ?

Comments:

A score of 6 = basic use of two questions

	EXCELLENT				PASS						
DELIVERY & EFECTIVENESS	20	18	16	14	12	10	8	6	4	2	0

Student uses unit language and pronunciation appropriately (10)

Student has good control over pacing and hesitation (5)

Student actively participates to develop the discussion (5)

Comments:

	EXCELLENT				PASS						
DISCUSSION SKILLS	20	18	16	14	12	10	8	6	4	2	0

1. Introduce topics (let's talk about…)
3. Answer questions/give opinions with extra information
2. Discuss benefits and/or problems
((one of) the main problems with ___ is…)
4. Deals with communication problems
((one of)the main benefits of ___ is…)

Comments

A score of 12 = basic use of 2 skills.

RUBRIC ANALYSIS: LEARNER ABILITY, RATER SEVERITY, CATEGORY DIFFICULTY

```
+---------------------------------------------------------------------------+
|Measr|+students|-raters      |-rubric items           | S.1 | S.2 | S.3    |
+---------------------------------------------------------------------------+
  1  +         +             +                         +  (9)  +(20)  +(19)
       WO

                                                                        17
                                          ---     ---           ---
                                                                        ---
            K                Delivery & Effectiveness
       RT                    Discussion Skills          7     17     16
            B    J
  *  0  * NN     * D  F  G  H  I *                      * --- *       * --- *
            C                                                 16
            E                                          6             15
       MO                                              ---     ---
       KI   A
                                                       5      15     ---
       RM                     Discussion Questions
                                                              ---
                                                       ---           14
                                                              14
  -1 +         +             +                         +  (2)  +(12)  +(12)
+---------------------------------------------------------------------------+
|Measr|+students|-raters      |-rubric items           | S.1 | S.2 | S.3    |
+---------------------------------------------------------------------------+
```
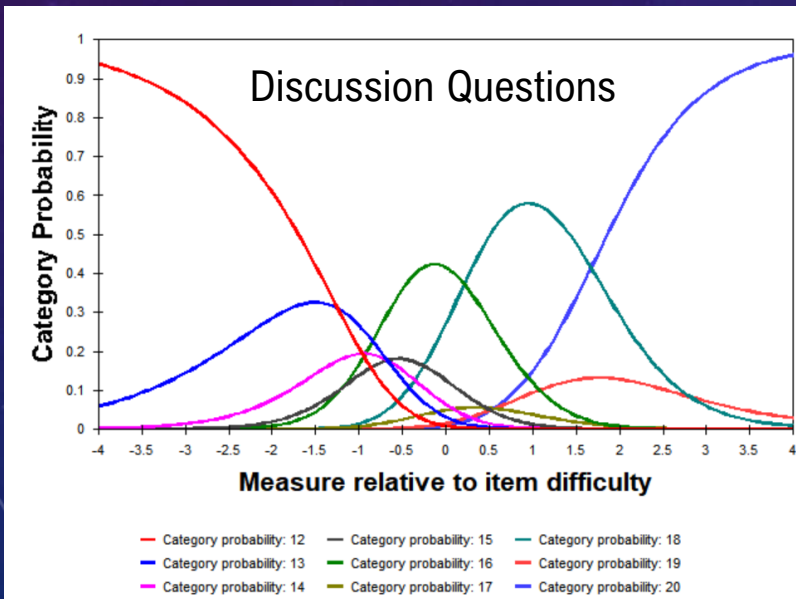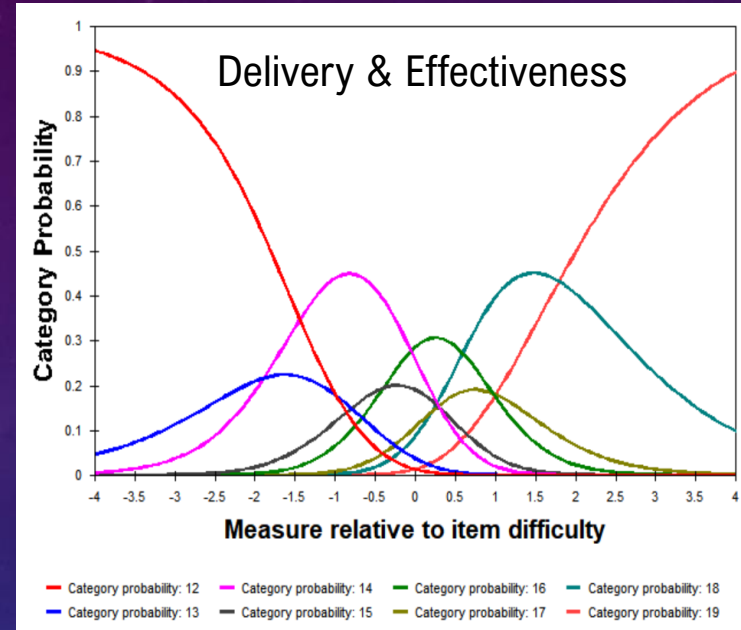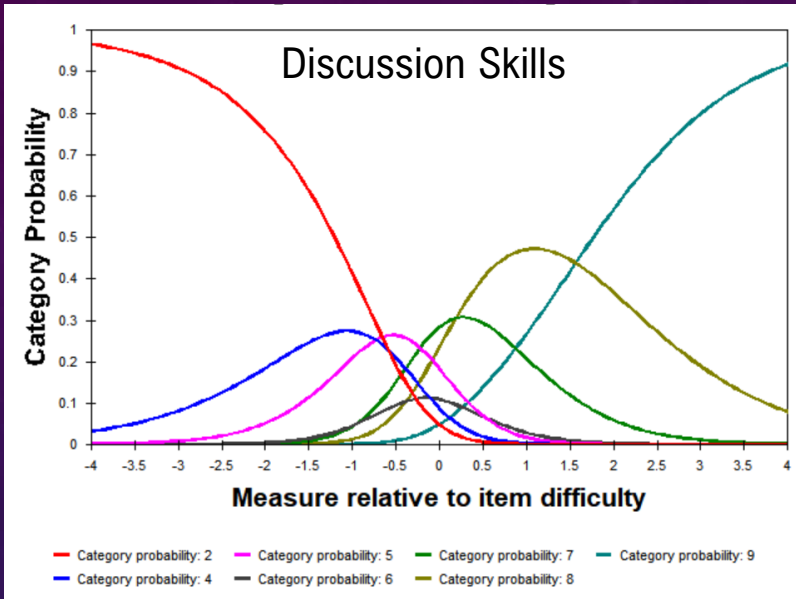
RUBRIC ANALYSIS: CATEGORY RESPONSE CURVES (LINACRE, 2005)
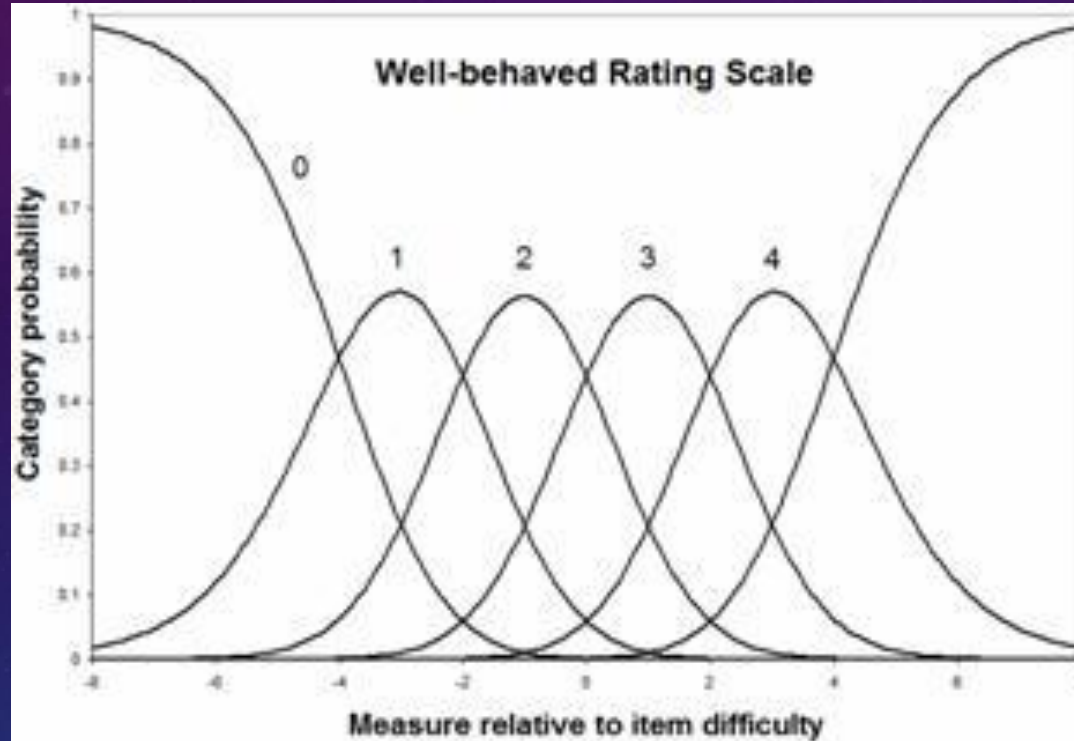
# THE DISCUSSION SKILLS RUBRIC

HOW POINTS WERE AWARDED

Discussion Questions	
Scale point	Number of observations
2	1
3	0
4	2
5	5
6	4
7	16
8	25
9	13

Discussion Skills	
Scale point	Number of observations
12	5
13	8
14	8
15	9
16	20
17	2
18	12
19	1
20	1

Delivery and effectiveness	
Scale point	Number of observations
12	6
13	7
14	23
15	10
16	11
17	4
18	4
19	1

WHAT TO DO ABOUT IT?

Collapse the scale points together: Discussion SKills
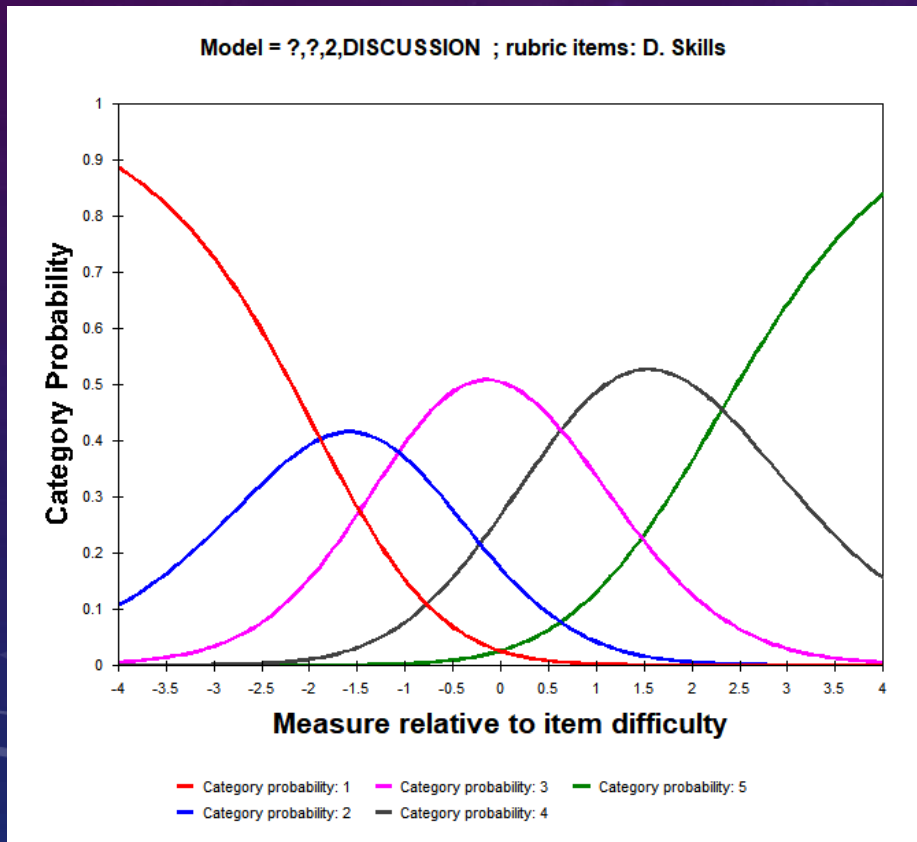
New Scale	
0	1 ~ 11
1	12
2	13 ~ 14
3	15 ~ 16
4	17 ~ 18
5	19 ~ 20

	Course Language	Overall Interaction	Fluency	General Language Use
5	Course language is used effectively to develop and build the discussion. There are no issues with form or pronunciation.	Participants cooperate to build on each other's ideas. There is little or no inter-turn pausing. Interaction is highly effective.	Speaks very fluently with almost no hesitation, repetition or fillers.	Language use is complex and highly accurate. Speech is comfortably comprehensible.
4.5	Between 4 and 5	Between 4 and 5	Between 4 and 5	Between 4 and 5
4	Course language is used and generally develops and builds the discussion. There may be minor issues with form and/or pronunciation.	Participants generally cooperate to build on each other's ideas, although some topics may not be adequately developed. There is some minor inter-turn pausing. Interaction is effective.	Speaks fairly fluently with occasional hesitation, repetition or filler use.	Language use has a degree of sophistication and is generally accurate. Errors do not generally interfere with meaning.
3.5	Between 3 and 4	Between 3 and 4	Between 3 and 4	Between 3 and 4
3	Course language is used, although there may be issues with timing or placement in the discussion. There are some issues with form and/or pronunciation.	Participants attempt to cooperate and build on each other's ideas, although topic development may be limited. There is some inter-turn pausing. Interaction is generally acceptable.	Speaks somewhat fluently. Hesitation repetition or filler use is present, but the speaker can generally maintains the flow of speech.	A mix of both simple and complex language is used. There are some errors. Errors sometimes interfere with meaning, but the speech is generally comprehensible
2	Course language is attempted, but does not develop the discussion well. They may not relate to the previous turn. There are some issues with form and pronunciation.	There is some attempt to cooperate and develop ideas, but contributions are often not well related to the previous turn. There is some noticeable inter-turn pausing. Interaction is faulty.	There is noticeable hesitation. Speaker cannot form utterances smoothly.	Language is mostly simple but complex forms are attempted occasionally. There are frequent errors that sometimes affect meaning.
1	Course language may not be attempted and/or they do not develop the discussion. Placement may be marked/they may not relate to the previous turn. There are some major issues with form and pronunciation.	Participants generally do not cooperate. Speakers appear to produce pre-planned language with little or no topic development or mutual interaction. There is significant mid-turn pausing. Interaction is limited/non-existent	Speech is highly disfluent characterised by short utterances and significant pauses.	The language is very simple and is characterised by frequent errors that often affect the meaning.

REVISED SCALE

- Clearer categories

- Fewer, better defined scale-points

- Less focus on quantification, discrete function use

- more focus on achievement of interactional goals

RUBRIC DESIGN RECOMMENDATIONS

- Fewer scale points rather than more - improves reliability

  - OK, to have a 'lopsided' scale e.g. 2 fail scores but multiple pass scores (2, 5, 6, 7, 8, 9, 10)

- Clear distinction between scale points

- Clearly define categories & constructs

- 1 scale per construct – maximum of 7 categories

- Design the scale to fit the purpose or course goals

- Gather feedback

- Gather data and analyze - even descriptive stats give an insight

REFERENCES

- Bonk, W. J., & Ockey, G. J. (2003). A Many facet Rasch analysis of the second language group oral discussion task. Language Testing, 20(1), 89 – 110 https://doi.org/10.1191/0265532203lt245oa

- Council of Europe. (2001). Common European framework of reference for languages: Learning, teaching, assessment. Available at https://rm.coe.int/16802fc1bf

- Eckes, T. (2015). Introduction to many-facet rasch measurement. Peter Lang.

- Galaczi, E. D., Ffrench, A., Hubbard, C., & Green, A. (2011). Developing assessment scales for large-scale speaking tests: a multiple-method approach. Assessment in Education: Principles, Policy & Practice, 18 (3), 217–237. https://doi.org/10.1080/0969594X.2011.574605

- Green, A. (2014). Exploring language assessment and testing. Routledge.

- Janssen, G., Meier, V., & Trace, J. (2015). Building a better rubric: Mixed methods rubric revision. Assessing Writing, 26, 51–66. https://doi.org/10.1016/j.asw.2015.07.002

- Kramsch, C. (1986). From language proficiency to interactional competence, The Modern Language Journal, 70(4), 366–72. https://doi.org/10.2307/326815

REFERENCES CONTD.

- Leaper, D. A. & Brawn, J. R. (2019). Detecting development of speaking proficiency with a group oral test: A quantitative analysis. Language Testing, 36(2) 181–206. https://doi.org/10.1177/0265532218779626

- Linacre, J. M. (1994). Many-facet Rasch measurement. 2nd ed. MESA Press.

- McDonald, K. (2018). Post hoc evaluation of analytic rating scales for improved functioning in the assessment of interactive L2 speaking ability. Language Testing in Asia, 8 (19). https://doi.org/10.1186/s40468-018-0074-3

- May, L., Nakatsuhara, F., Lam, D., & Galaczi, E. (2020). Developing tools for learning orientated assessment of interactional competence: Bridging theory and practice. Language Testing, *37 (2),* 165 – 188 https://doi.org/10.1177/0265532219879044

- Myford, C. M. & Wolfe, E.W. (2003.). Detecting and measuring rater effects using many-facet rasch measurement: part 1. Journal of Applied Measurement, 4(4), 386–422.

- Nitta, R., & Nakatsuhara, F. (2014). A multifaceted approach to investigating pre-task planning effects on paired oral task performance. Language Testing, 31 (2), 147–175. https://doi.org/10.1177/0265532213514401

- Scullen, S. E., Mount, M. K., & Goff, M. (2000). Understanding the latent structure of job performance ratings. Journal of Applied Psychology, 85(6), 956-970. https://doi.org/10.1037/0021-9010.85.6.956

- Weigle, S. C. (2002) Assessing Writing. Cambridge University Press.

- Wiliam, D. (2011). Embedded formative assessment. Solution Tree Press.