

## Implementing and rating a new peer-to-peer assessment of speaking skills in New Zealand

Martin East  
School of Cultures, Languages and Linguistics  
m.east@auckland.ac.nz



1

## What I will cover in this presentation

1. New Zealand's high-stakes assessment system for schools – an introduction
2. An introduction to *interact* – a new way of assessing the spoken proficiency of learners of an additional language (L2)
3. Why did we introduce a new assessment?
4. The reform in practice (a research study)
5. Task examples
6. The rating process
7. Support with rating
8. Final thoughts

2

## New Zealand's high-stakes assessment system

- The National Certificate of Educational Achievement (NCEA)
- Introduced from 2002; revised from 2011
- Operates at three levels:
  1. Level 1, end of Year 11, students 15+
  2. Level 2, end of Year 12, students 16+
  3. Level 3, end of Year 13, students 17+

3

## New Zealand's high-stakes assessment system

- Replaced a traditional norm-referenced, summative examination dominated 'assessment of learning' system
- Criterion-referenced and 'standards-based'
- Internal (teacher-created and teacher-assessed) components are central to the system

4

## New Zealand's high-stakes assessment system

John Hattie (2009) framed the NCEA within a context of a "revolution of assessment":

This revolution relates to Assessment for Learning and it can be witnessed in innovations such as the National Certificate of Educational Achievement (NCEA) and its standards-based approach, the emphasis on reporting more than on scoring, constructive alignment of learning and outcomes, peer collaborative assessment, learning intentions and success criteria, and the realisation of the power of feedback. (p. 259)



5

## New Zealand's high-stakes assessment system

- An 'assessment for learning' stance to high-stakes assessment is reflected in a New Zealand Position Paper on assessment (Ministry of Education, 2011):
  - a move "beyond a narrow summative ("end point" testing) focus to a broader focus on assessment as a means of improving teaching and learning" (p. 4).
  - an assertion that "[t]his approach to assessment has strongly influenced the way in which we have implemented standards-based assessment" (p. 9).

6

### New Zealand's high-stakes assessment system

- Through internal assessments, teachers mark and provide feedback on their students' work in relation to the published expectations of the relevant assessment opportunity.
- This "deliberate focus on the use of professional teacher judgment underpinned by assessment for learning principles rather than a narrow testing regime" is "very different from that in other countries" (Ministry of Education, 2011, p. 9).

7

### New Zealand's high-stakes assessment system

So far so good.

However

- In practice, the high-stakes nature of the system means that issues of accountability, validity and reliability are important.
- As a consequence, external examinations still have a role to play, and teachers' internal assessments are subject to scrutiny and moderation.

8

### New Zealand's high-stakes assessment system

These realities set up genuine tensions within the system.

To illustrate the tensions, I will use as an example a new assessment of foreign language students' spoken communicative proficiency, called *interact*

9

### Assessment Innovation

- Until recently, measuring L2 students' spoken proficiency was done by a **one-time end-of-year interview test** between the teacher and the student (known as *conversation*).
- The new assessment (*interact*) aims to measure spoken proficiency principally by capturing a **range of genuine student-initiated peer-to-peer interactions as they take place in the context of regular classroom work throughout the year.**

10

### Assessment Innovation

- This new high-stakes assessment signals a radical departure from established practices.
- Its introduction has not occurred without considerable controversy and a whole range of different opinions.



11

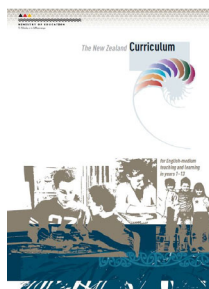
### A story of assessment innovation: What? Why? How?



12

## Background

- The curriculum for New Zealand's schools was recently substantially revised
- The original curriculum, in place since 1993, was replaced in 2007, and a new curriculum fully mandated from 2010



13

## Background

- The revised New Zealand curriculum is intended to be **learner-centred and experiential**, with teachers and learners engaged in a co-construction of knowledge (teacher and learner, and peer-to-peer), rather than top-down and teacher-led, with the teacher representing the authority figure



14

## Background

- Along with curriculum reform, there was a complete overhaul of the NCEA
- New NCEA assessments have now been introduced
- *Interact* is one such new assessment.

15

## Background

*Interact* is built on the co-constructive premises that:

- students learn to use the target language most effectively when they are engaged in real language use in the classroom (Willis & Willis, 2007)
- they learn how to communicate through interaction in the target language (Nunan, 2004)
- engagement in meaningful language communication should be an important focal point for assessments (Norris, 2002)

16

## Background



When it comes to using the target language, we want, in a whole range of scenarios, to move students from the language they can initially use only with help and support to the language they can eventually use unaided and without support – **linguistic automaticity**.

17

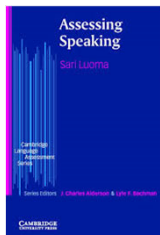
## Theoretical Rationales



18

### Assessing speaking

- According to **Luoma (2004)**, the most common way of organising speaking assessments is “to assess examinees one at a time, often in an interview format” (p. 35).
- Until fairly recently, this well-established format for speaking assessments has not really been brought into question.



19

### Assessing speaking

- But a weakness of single interview tests is that they do not represent normal conversation (van Lier, 1989).
- Normal conversation is spontaneous and can go in a range of directions.

20

### Assessing speaking

- Single interview tests “focus too much on the *individual* rather than the individual in *interaction*” (McNamara, 1996, p. 85, my emphasis).

21

### Assessing speaking

- An alternative is to assess candidates in pairs.
- In the paired format, “the examinees are asked to interact with each other, with the examiner observing rather than taking part in the interaction directly” (Luoma, 2004, p. 36).

22

### Assessing speaking

The paired assessment format can elicit:

- greater balance (more equal interlocution) between the partners (Együd & Glover, 2001; Luoma, 2004).
- a broader spectrum of functional competence (Galaczi, 2010)
- a wider range of interactional patterns (Saville & Hargreaves, 1999; Swain, 2001).

23

### Assessing speaking

Candidates can display:

- collaboration, cooperation and coordination (Jacoby & Ochs, 1995)
- prompting, elaboration, finishing sentences, referring to a partner's ideas and paraphrasing (Brooks, 2009)
- turn taking, initiating topics and engaging in extended discourse with a peer rather than a teacher/examiner (Ducasse & Brown, 2009; May, 2011)

24

## Assessing speaking

We can

- collect a range of examples of speaking that reflect how interactions usually take place (Skehan, 2001).
- measure a more comprehensive **spoken communicative proficiency construct**.
- make better or more useful inferences about language users' proficiency in wider real-life contexts (Galaczi, 2010).

25

## Assessing speaking

- Paired assessments also mirror what is already happening in the regular functioning of a communicatively-oriented classroom in terms of pair/group work (or it will encourage more paired interaction in class) (Galaczi, 2010; Swain, 2001).
- Paired assessments are arguably more representative of 'best practice' in L2 classrooms (Együd & Glover, 2001; Taylor, 2001)

26

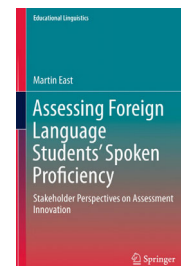
All sounds great in theory  
How does it work in practice?  
And why has the New Zealand  
'experiment' with *interact* provoked  
controversy?



27

## The reform in practice

A two-year project to  
investigate teachers' and  
students' perspectives at  
two significant junctures –  
end of 2012 and end of  
2013  
(funded by the University of Auckland)



28

## A research project for investigating stakeholders' perspectives on reform

- Phase I (2012) included a large-scale nationwide anonymous paper-based survey, targeted at teachers of the five principal international languages taught in New Zealand (Chinese, French, German, Japanese and Spanish)
- I wanted to find out teachers' views, both positive and negative, on *interact* in comparison with *converse*.
- 152 anonymous surveys were completed

29

## A research project for investigating stakeholders' perspectives on reform

- Phase II (2013) included interviews with thirteen teachers who were implementing *interact* at the highest level of examination (CEFR B1 – B2)
- I wanted to find out what these teachers thought of *interact* and how they were putting it into practice.

30

# What is working?

(Teacher survey)

31

## What is working?

In summary:

- authenticity
- preparation for the real-world
- focus on fluency rather than accuracy
- peer-to-peer interactions
- more than one assessment opportunity
- more valid assessment data

32

# What is not working?

(Teacher survey)

33

## What is not working?

In summary:

- workload a major issue
- spontaneous and unrehearsed is unrealistic and unworkable
- pre-scripting / rote-learning still an issue
- partners don't always work together
- can be stressful for students

34

# What could work better?

(Teacher survey)

35

## What could work better?

In summary:

- Reduce the number of assessment points
- Be real about spontaneity
- Allow adequate preparation
- Provide clear guidance on spontaneous and unrehearsed

36

## Two task examples (level 3)

37

## Task examples (level 3)

### Monika's task

New Zealand is a land that produces part of its own energy and has the benefit of being an island far away from pollutants, sparsely populated and windy. Why should you even care about environmental challenges, how do they affect you and your generation? Discuss with a partner aspects of environmental threats and opportunities in the context of New Zealand and globally. You could consider the following: Explain the challenge or opportunity to the environment, why you consider it significant, discussing the impact of inaction, the historic reasons for the situation, negotiating possible solutions.

38

## Task examples (level 3)

Alison presented a quite different task for her Year 13 Japanese classroom: 'Does New Zealand need a cat café?'



39

## The rating process

All information publicly accessible at:  
[www.nzqa.govt.nz](http://www.nzqa.govt.nz)

40

## The assessment criteria

- Contained in assessment 'blueprints' (Bachman and Palmer, 2010) known as **standards**
- Each standard describes what a student **needs to know**, or what they **must be able to achieve**, in order to meet the standard. Having met it, they will gain credits towards the NCEA.

41

## The assessment criteria

- Assessments measure what a student **knows or can do** against the **registered criteria** of a standard in courses they study. If they meet the criteria, they achieve the standard, and gain credits towards a qualification.

42

### The assessment criteria

There are four grades:

- **Achieved (A)** for a **satisfactory** performance
- **Merit (M)** for **very good** performance
- **Excellence (E)** for **outstanding** performance
- **Not achieved (N)** if students do not meet the criteria of the standard

43

### The assessment criteria

- In the case of **not achieved**, schools may allow students to have further assessment opportunities for internally assessed standards later in the year.
- Students can have up to one further assessment per standard per year.

44

What is required at each level?

45

### General requirements (all levels)

#### Language Features

- A repertoire of language features and strategies to maintain the interaction must be demonstrated. Students should not write and learn scripted role plays by heart, or prepare the total interaction beforehand. Indicators of the ability to maintain an interaction must be evident.

46

### General requirements (all levels)

#### Language Features

- Features and strategies include **pausing, negotiating meaning, prompting, seeking clarification** etc. This can only be in evidence when the student is unaware of all questions that are asked, and has not prepared all answers. For this reason, the use of cue cards is also not allowed.

47

### General requirements (all levels)

#### Range

The standard requires a minimum of **two interactions** in **different situations**.

For example, negotiating the best way to spend a Saturday evening with friends will require different language from discussing what students did in the holidays.

48

### General requirements (all levels)

#### Assessing the collection of evidence

- The grade will be awarded for the collection of interactions assessed as a whole, i.e. each interaction will not be assessed individually - the grade will be derived from the overall quality of the work.

49

### Interact at level 1

Interact to communicate **personal information, ideas and opinions** in different situations.

Interactions where students can react in a genuine way, and where they are able to naturally control the direction of the interaction, give students the opportunity to meet the criteria.

About three minutes of evidence is needed

50

### Interact at level 2

Interact to **share information and justify ideas and opinions** in different situations.

At level 2, students need to move beyond simply supplying information and into **justifying expressed ideas and opinions**. This can be done by giving evidence or explanations which support these views and/or the views of others. Over the collected evidence there will be evidence of both sharing and justifying.

About four minutes of evidence is needed

51

### Interact at level 3

Interact clearly to **explore and justify varied ideas and perspectives** in different situations.

The standard at level 3 requires evidence of **spontaneity**, i.e. the ability to maintain and sustain an interaction without previous rehearsal

About five minutes of evidence is needed

52

### Interact at level 3

At level 3, students need to move beyond simply providing information and into exploring and justifying a variety of ideas and perspectives. This will involve **evaluating, explaining and providing supporting evidence**. Students could also support or challenge the ideas and perspectives of others.

53

### The rating criteria

54

## Rating criteria

### Achievement Standard

Subject Reference	Japanese 1.3		
Title	Interact using spoken Japanese to communicate personal information, ideas and opinions in different situations		
Level	1	Credits	5
Subfield	Languages	Assessment	Internal
Domain	Japanese		
Status	Registered	Status date	9 December 2010
Planned review date	31 December 2020	Date version published	20 November 2014

55

## Rating criteria (level 1)

This achievement standard involves a range of interactions using spoken Japanese to communicate personal information, ideas and opinions in different situations.

### Achievement Criteria

Achievement	Achievement with Merit	Achievement with Excellence
<ul style="list-style-type: none"> <li>Interact using spoken Japanese to communicate personal information, ideas and opinions in different situations.</li> </ul>	<ul style="list-style-type: none"> <li>Interact using convincing spoken Japanese to communicate personal information, ideas and opinions in different situations.</li> </ul>	<ul style="list-style-type: none"> <li>Interact using effective spoken Japanese to communicate personal information, ideas and opinions in different situations.</li> </ul>

56

## Rating criteria (level 2)

This achievement standard involves interacting using spoken Japanese to share information and justify ideas and opinions in different situations.

### Achievement Criteria

Achievement	Achievement with Merit	Achievement with Excellence
<ul style="list-style-type: none"> <li>Interact using spoken Japanese to share information and justify ideas and opinions in different situations.</li> </ul>	<ul style="list-style-type: none"> <li>Interact using convincing spoken Japanese to share information and justify ideas and opinions in different situations.</li> </ul>	<ul style="list-style-type: none"> <li>Interact using effective spoken Japanese to share information and justify ideas and opinions in different situations.</li> </ul>

57

## Rating criteria (level 3)

This achievement standard involves interacting clearly using spoken Japanese to explore and justify varied ideas and perspectives in different situations.

### Achievement Criteria

Achievement	Achievement with Merit	Achievement with Excellence
<ul style="list-style-type: none"> <li>Interact clearly using spoken Japanese to explore and justify varied ideas and perspectives in different situations.</li> </ul>	<ul style="list-style-type: none"> <li>Interact clearly using convincing spoken Japanese to explore and justify varied ideas and perspectives in different situations.</li> </ul>	<ul style="list-style-type: none"> <li>Interact clearly using effective spoken Japanese to explore and justify varied ideas and perspectives in different situations.</li> </ul>

58

## Rating criteria (all levels)

The step up:

- Use language
- Use **convincing** language
- Use **effective** language

59

## Explanatory notes (level 1)

Interactions are characterised by:

- a genuine purpose
  - negotiating meaning
  - initiating and maintaining
  - participating and contributing
  - natural language
  - using different language for different purpose(s)
  - using conventions (eg cultural, courtesies, gestures)
  - contextually appropriate language
  - using simple interactive strategies such as fillers, questioning, thanking, apologising, pausing, prompting, seeking clarification.
- Not all characteristics may be evident in one interaction.

60

## Explanatory notes (level 1)

Interaction may be hindered in some places by inconsistencies in:

- language features
- understanding
- pronunciation
- intonation
- rhythm patterns
- delivery speed or audibility
- stress patterns
- tones.

61

## Explanatory notes (level 1)

*Personal information* includes opinions, ideas and information that relate to the student's life, and may include formal cultural responses.

62

## Explanatory notes (level 1) Convincing and Effective

*Convincing spoken Japanese* refers to interaction showing use of a range of language that is fit for the context and generally successful selection from a repertoire of language features and strategies to support the interaction. Interaction is not significantly hindered by inconsistencies.

*Effective spoken Japanese* refers to interaction showing successful use of a range of language that is consistently fit for the context and skilful selection from a repertoire of language features and strategies to support the interaction. Interaction is not hindered by inconsistencies.

63

## Support with rating

64

## Annotated exemplars

An annotated exemplar is an extract of student evidence, with a commentary, to explain key aspects of the standard.

These exemplars are there to help teachers to make rating judgements, in particular at the grade boundaries.

65

## Annotated exemplars

Examples for **Japanese as L2**, for interactions at NCEA level 1:

- Low excellence
- High merit
- Low merit
- High achieved
- Low achieved
- High not achieved

<https://www.nzqa.govt.nz/ncea/subjects/japanese/annotated-exemplars/level-1-as90895/>

66

**Low Excellence**

**Commentary**

For Excellence, the student needs to interact using effective spoken Japanese to communicate personal information, ideas and opinions in different situations. This involves successful use of a range of language that is consistently fit for the context, and skilful selection from a repertoire of language features and strategies to support the interaction. Interaction is not hindered by inconsistencies.

The student being exemplified appears on the left in Interaction 1, and is the second student on the left in Interaction 2.

Effective Japanese is shown through a repertoire of language features and strategies to support the interaction. He is able to keep the conversations going at all times.

Skilful selection is shown through an impressive range of fillers which are all used at the appropriate time and place, e.g. ああそうですか、ほんとですか、おもしろそうですね、ところで、あかげるまで、それではいいいえ、そうです。He gives responses and asks questions.

He is able to communicate his own information and ideas and opinions, ワイヘでつりまてからめうあいなストラのたべのをたべることができます、and also enquire about others やすあどこにいきたいですか。He is able to interact with what has been said, for example ロトルアはたかいですからロトルアよりワイヘはいきたいです and せんせいはニューワールドのドルがありますか。

The student shows evidence of being able to talk about past and future events, for example としょかんへいきましたか、and uses both the taidesu form and the masu form of verbs, especially when helping to negotiate whether the group should go to Waiheke or Rotorua, e.g. おんせんはおかぬがかりますからおかねをもっていきまます。

To sit more securely at Excellence, improved consistency in some areas of language use would further ensure that the criteria that communication is not hindered by inconsistencies is met, for example ワイヘで散歩をしながら海の景色をみることで、きますからうみのけしき好きな人ワイヘでいきたいと思います。

**Student response**

Interaction 1 (opens in a new window)

Interaction 2 (opens in a new window)

Interaction 3 (opens in a new window)

Download Interaction 1 (ZIP, 42MB) for offline study

Download Interactions 2 & 3 (ZIP, 38MB) for offline study

67

**Low Merit**

**Commentary**

For Merit, the student needs to interact using convincing spoken Japanese to communicate personal information, ideas and opinions in different situations. This involves use of a range of language that is fit for the context, and generally successful selection from a repertoire of language features and strategies to support the interaction. Interaction is not significantly hindered by inconsistencies.

The student being exemplified is the second speaker in Interactions 1 and 2, and the first speaker in Interaction 3.

Convincing Japanese is shown through a range of strategies, e.g. すきです and 大好きします。The student is also able to make a good response to the comment 七時ははいですね with the reply 七時ははやすぎますか。

Any long pauses in the interactions are maintained by the student's use of えっと, and the most commonly used comment of そうですか, which means that the flow of the interaction is not lost.

The student is able to successfully self-correct to maintain communication: スポーツをすることです。

Interaction is not significantly hindered by inconsistencies.

For a more secure Merit, there needs to be additional evidence of both a range of language that is fit for the contexts, and a repertoire of language features and strategies to support the interaction, for example in the first interaction especially, comments to questions and suggestions are short, and responses could provide additional detail.

**Student response**

Interaction 1 (opens in a new window)

Interaction 2 (opens in a new window)

Interaction 3 (opens in a new window)

Download all interactions (ZIP, 16MB) for offline study

68

## Internal Moderation

- At the end of the day, this is a high-stakes system
- In addition to rating criteria and exemplars, there is an expectation of collegial input to ensure that the assessments are valid, reliable and benchmarked against the national standard

69

## Internal Moderation

- Internal moderation supports the credibility of assessment by ensuring that assessment is **valid**, and grade judgements are **verified**
- The credibility of national assessment relies on quality assurance systems operating in every institution.
- Schools must have assessment policies and procedures to ensure that results reported are **accurate and consistent with the listed standard**.

70

## Internal Moderation

Schools must undertake quality assurance each year for each standard to be assessed

**School logo / Name**

**INTERNAL MODERATION COVER SHEET**

NZQA Assessment Rules require that schools must establish an internal moderation process that meets NZQA's requirements and which is applied each year to every internally assessed standard being assessed, to ensure that judgements are consistent with the Assessment Standard.

Faculty/Department: \_\_\_\_\_ Teacher in Charge of Assessment: \_\_\_\_\_

Assessors: \_\_\_\_\_

71

## Internal Moderation

**Section B: Critique assessment materials before any assessing**

The critiquing process is to ensure that the assessment materials meet the requirements specified in the standard and provide the opportunity for students to present authentic evidence at all grades.

Before use the assessment materials have been checked against:

The current moderator report and clarification of the standard document	Yes	No
The conditions of assessment	Yes	No
Any external moderation feedback	Yes	No
The standard is unchanged and the task has been previously critiqued.	Yes	No

**If yes, no further critiquing required.**

The assessment is consistent with the explanatory notes/learning context/curriculum level	Yes	No
The assessment allows students to achieve all requirements and grades of the standard	Yes	No
Assessment schedule is consistent with the standard and clarifications documents	Yes	No
Instructions are consistent with the standard explanatory notes/range statements	Yes	No
Judgement/sufficiency statements clearly describe performance levels for each grade, e.g. quality & length	Yes	No
Evidence statements allow for a range of acceptable answers with specific examples for each grade (AME)	Yes	No
Student instructions are clear and language is appropriate	Yes	No
Registered standard number, version, title, level and credits are given.	Yes	No

Critiquers' Name: \_\_\_\_\_ School: \_\_\_\_\_ Date: \_\_\_\_\_

72

## Internal Moderation

<b>Section C: Verify sufficient teacher judgements <i>before reporting results</i></b>		<b>PTO</b>	
The verification process is to ensure that any teacher judgements are consistent with the standard before they are reported to NZQA.			
Verifier Name:	School:	Position:	Date:
Evidence of verification is available overleaf, attached or at location (specify):			
I have sighted evidence of the critiquing and verification processes completed for this standard.			
Signed:	Position:	Date:	
<b>Section D: Retain samples and review assessment materials <i>after results are reported</i></b>			
Results loaded into student management system:		Date:	
Assessment materials and student work are stored ready for external moderation		Yes	No
Location or file path:			
The school's random selection procedure has been used to select work for external moderation (if required).		Yes	No
Assessment materials have been reviewed in response to feedback.		Yes	No
New benchmark samples have been annotated and/or existing examples of grade boundary decisions have been updated.		Yes	No

73

## External Moderation

- National external moderation provides an assurance that assessment decisions, in relation to assessment standards, are **consistent** nationally.
- External moderation of internally assessed standards ensures that assessment judgements (marking of students' work) are at the national standard.

74

## External Moderation

Schools need to submit:

- A copy of the task and any supporting resources
- Selected samples of student work (eight samples)

75

## External Moderation

### The annual national moderator's report

- provides **feedback to assist assessors** with general issues and trends that have been identified during external moderation of the internal Languages standards

76

## External Moderation

This standard is producing some high-quality examples of students using the target language they are learning to express themselves and to interact with other people.

In some moderation there is still an inclination towards role plays, interviews or interactions which have clearly been practised to the specific task. These do not produce natural evidence to support the criteria or intent of the standard.

There is no expectation that students will speak without practice of the language involved, as students will be learning the language in class and practising using it in a variety of ways. However, the specific assessment task should not be practised either with different students before the task, or with the partner before recording. Successful assessment tasks/activities seen in moderation allowed opportunity for students to provide natural evidence of the strategies required by the standard, rather than rehearsed material.

77

## External Moderation

The most successful student evidence typically demonstrates genuine reactions to what has been said, rather than moving to the next prepared question. Students who demonstrate successful outcomes for the Interact standard come into an assessment having thought about the language they will be using, which may include some focus questions they may ask if the opportunity arises, while taking the interaction in natural directions. Providing questions beforehand is unlikely to allow evidence of the interactive intent of the standard to be demonstrated.

Typical characteristics of natural use of language includes features such as seeking clarification, self-correcting, reacting to what is said rather than moving to the next prepared question, making mistakes, pausing for thought, prompting etc.

78