# Task-based Performance Assessment of Japanese Second Language Writing

IATEFL 2008
Exeter
Yoshihito SUGITA

## Contents



## Introduction

- The dual-mode system (Skehan, 2001)
- When time is pressing, and contextual support high, memory-based communication is appropriate.
- When there is more time, and precision is important, the rule-based system can be accessed.

## Developing a performance assessment

- Construct-based approach (e.g., Alderson et al., 1995; Bachman & Palmer, 1996; Brown, 1996).
- Procedures for the design, development and use of language tests must incorporate both a specification of the assessment tasks to be included and definitions of the abilities to be assessed.

## The specification of assessment task 1

- You will have 20 minutes to write a 70-80 word letter introducing yourself to your host family. First, think of answers to the following questions.
- How old are you? What do you do? What about your family?
- What are your hobbies and interests?

## The specification of assessment task 1

- Have you been abroad?
- Do you like pets? Regarding food, do you have any special likes or dislikes?
- What do you want to do while you are in England?

## Construct definitions of task 1

| Accuracy | |
|---|---|
| Organizational skills | Linguistic accuracy |
| The writing displays a logical organizational structure which enables the content to be accurately grasped. | Errors of vocabulary, spelling, punctuation or grammar. |

## The specification of assessment task 2

- **You will have 10 minutes to make notes about the following discussion topic, "Why do you study English?" In order to prepare for the discussion, think of answers to the question as many as possible and write them as "To travel abroad."**

## Construct definitions of task 2

| Communicability | |
|---|---|
| Communicative quality | Communicative effect |
| The writing displays an ability to communicate without causing the reader any difficulties. | Quantity of ideas to develop the response and relevance of the content to the proposed task. |

## Procedures for testing

- **Steps to administer the test**

**Identify rubrics of task 1**

↓

**Perform the task within 20 minutes**

↓

**Identify rubrics of task 2**

↓

**Perform the task within 10 minutes**

↓

**Collect the test papers**

## Rating scale development

- ***Who is going to use the rating scale?***

**Assessor-oriented scales are intended to guide the rating process, and focus on comparing the written text with descriptors on the scale (Alderson, 1991).**

## Rating scale development

- ***What aspects of writing are most important, and how will they be divided up?***

**The focus of the assessment is on the acquisition of accuracy (organizational skills, linguistic accuracy) and communicability (communicative quality and effect).**

## Rating scale for task 1 (Accuracy)

| Organizational skills |
| --- |
| The written text<br>-is well organized and well developed (TWE).<br>-shows strong rhetorical control and is well managed (MWA).<br>-has clear organization with a variety of linking devices (FCE). |

## Rating scale for task 1 (Accuracy)

| Linguistic accuracy |
| --- |
| The written text<br>-demonstrates appropriate word choice though it may have occasional errors (TWE).<br>-has few errors of agreement, tense, number, word order/function, articles, pronouns, prepositions, spelling, punctuation, capitalization, paragraphing (ESL). |

## Rating scale for task 2 (Communicability)

| Communicative quality |
| --- |
| The written text<br>-displays consistent facility in use of language (TWE).<br>-contains well-chosen vocabulary to express the ideas and to carry out the intentions (MWA). |

## Rating scale for task 2 (Communicability)

| Communicative effect |
| --- |
| The written text<br>-effectively addresses the writing task (TWE).<br>-has a very positive effect on the target reader with adequately organized relevant ideas (FCE). |

## Rating scale development

- ***How many points, or scoring levels, will be used?***

  **Many large-scale assessment programs such as TOEFL use a six-point scale; however, some questions about scale points can only be determined through empirical means in pre-testing (Weigle, 2002).**

## Procedures for the pre-testing

- **Performances by all 15 Japanese university students were rated by 5 experienced high school teachers of English using rating scales for accuracy and communicability specifically developed for this study.**

## Analysis of the pre-testing

- **The data were analysed using FACETS (Linacre, 2008).**
- **To examine the measurement characteristics of the pre-test, three facets were specified: subject, rater and task. The partial-credit model was chosen because the scoring criteria for the two scales were qualitatively.**

## Results of the pre-testing

- ***Is student ability effectively measured?***

  **Subject ability estimates range from a high of 3 logits to a low of -5 logits, a spread of 8 logits in terms of student ability. The reliability index was .93, which demonstrates it is possible to achieve reliable ability scores.**

## Results of the pre-testing

- ***Are teacher-raters equally severe?***

  **The fixed chi-square for rater severity is 19.0 with df=4 and p= .00, so the raters are not equally severe. Separation of raters is 1.94, with a reliability of .79 which demonstrates the analysis is fairly reliably separating raters into different level of severity.**

## Results of the pre-testing

- ***How much do tasks (i. e., tests) that are designed to be equivalent actually differ in difficulty?***

  **The analysis of the tasks shows that no significant difference occurs between the two tasks (reliability of separation index =.35; fixed chi-square: 1.5, df: 1; significance: p= .21). The tasks do not appear to separate the students to a significant degree, this means that the two tasks can be considered equivalent.**

## Results of the pre-testing

- ***Are scales efficient and consistent with assumptions about distributions of student ability?*** **1.4 logits<step difficulty (SD)<5.0 logits**

| Scale level | Accuracy (SD) | Communicability(SD) |
|---|---|---|
| 1 | 6% | 4% |
| 2 | 21% (-3.99) | 24% (-4.25) |
| 3 | 31% (-1.26) | 20% (-.60) |
| 4 | 23% ( .82) | 19% ( .52) |
| 5 | 11% (2.06) | 26% (1.06) |
| 6 | 7% (2.37) | 7% (3.28) |

## Implications for further study

1. **Teacher-assessments have significant variations; they did not have equal severity.**
2. **The two tasks used in this study were similar in terms of difficulty.**
3. **The rating scales for each task should be modified from a 6 to a 5 point scale.**

## Contact Address


- **Yoshihito Sugita**
  **Yamanashi Prefectural University, JAPAN**
  **Email:** sugita@yamanashi-ken.ac.jp
  **URL:** http://www.yamanashi-ken.ac.jp/