

Using FACETS to Investigate the Appropriacy of an IEP Speaking Placement Test

Linxiao Wang & Tingting Kang

English Department

Northern Arizona University

### Abstract

Performance assessment has been largely implemented in speaking placement tests to determine the appropriate level of instruction in intensive English programs (IEP). Such assessment introduces sources of score variance, which may negatively result in inappropriate scores. This study investigated whether the speaking placement test worked as intended. The influences of examinees' speaking proficiency, task difficulty, rater severity, and the rating scales on three speaking task scores awarded to 158 international students were examined using Many-facet Rasch measurement (Linacre, 1989). The results showed that the test placed students into statistically distinct proficiency levels. The three tasks were differentiated in terms of their levels of difficulty. Integrated task appeared to be the most challenging task to the examinees, and narrative task was only a bit more difficult than the independent task. As for the rating scales, a central tendency effect of overusing the middle categories was observed in both narrative and independent tasks. Thirty raters varied in severity with which they scored examinees. We then discussed implications for the use of speaking placement test scores, the design of speaking tasks with increasing difficulty levels, the revision of scoring rubrics, and the future rater training.

## Using FACETS to Investigate the Appropriacy of an IEP Speaking Placement Test

### **Background**

This study examined to what extent the three IEP placement speaking tasks work as intended to distinguish students into six levels. IEP often uses placement tests to distinguish incoming international students into pre-determined levels. Specifically, placement speaking tasks vary with their level of difficulty to discriminate examinees' speaking ability. Additionally, rater's severity level is expected to be consistent to impose some level of objectivity on speaking scores assigned. The categories on each rating scale used are supposed to be appropriately ordered and clearly distinguishable.

### **Research Questions**

This study answered the following research questions: 1) To what extent does the speaking placement test discriminate examinees' speaking proficiency? 2) Are there any examinees that demonstrate unusual profiles across the three speaking tasks? 3) To what extent do the three tasks vary in terms of difficulty? 4) Do raters differ in the severity with which they score examinees? 5) Are there any raters who rate examinee performance inconsistently? 6) Are the in-house or standardized tests' rating scales for each task functioning appropriately?

### **Methods**

#### **Participants**

There were 158 IEP incoming students in Fall 2011 taking the speaking placement test. Thirty raters, ten raters for each task, scored three speaking tasks, namely, narrative, independent, and integrated tasks.

## **Materials**

The Fall 2011 placement speaking tasks were used for the study. The narrative task was to create a story based on a series of pictures. The independent task elicited students' response on a question: *For vacation, would you rather travel to a different country or stay in your home country?* The integrated task asked students to summarize two listenings about popular culture and explain how popular culture influences their life.

Each task was assessed by a rubric of five categories. The rubric for the narrative task was an in-house holistic scale, and the rubrics for the independent and integrated tasks were adopted from TOEFL iBT independent and integrated speaking rubrics.

## **Data Analysis**

This study used the FACETS program (Linacre, 1989) and the linking plan is as follows. The two researchers were Rater 2 and Rater 11. Rater 2 scored a response to Task 1, which was previously scored by Rater 11; Rater 11 scored a response to Task 2, which was previously scored by Rater 2. Then, Rater 2 and 11 scored one response to Task 3 together.

## **Results**

The analysis of the examinees' ability facet showed that this IEP speaking placement test placed students into statistically distinct levels of proficiency. Results also confirmed that the IEP incoming students varied substantially in terms of speaking ability. Also, 7.59% examinees demonstrated unusual profiles across the three speaking tasks. Three tasks were well differentiated in terms of their levels of difficulty. Raters varied with severity with which they scored examinees, but rater severity's impact on candidate performance was not significant. Also, two raters rated examinee performance

inconsistently. In terms of rating categories, certain categories in rating scales functioned inappropriately.

### **Discussion and Implication**

The three tasks varied with difficulty: integrated task appeared to be fairly challenging to the examinees. The narrative task was a bit more difficult than the independent task. We suggest that the three tasks should be revised to reflect the increasing level of difficulty.

A central tendency effect of awarding many middle-category scores was observed in both the narrative and independent tasks. It seemed that the five rating categories for each task were not sufficient for discriminating the IEP program incoming students' speaking abilities. Creating new in-house rubrics or expanding existing rating categories, especially category 2 and 3, would help address the issue. There were only three responses received a score of 4 in the independent and integrated tasks. It might be necessary to revise the descriptor of category 4 to better match with our examinees' ability.

Also, some of the raters used the scoring rubrics inconsistently or with less variance. These concerns should be addressed in the future rater training.

Overall, this research provides empirical evidence for the appropriacy of the IEP speaking placement test.

### **References**

Linacre, J.M. (1989). *Many-faceted Rasch measurement*. Chicago, IL: MESA Press.