DOCUMENT RESUME

ED 133 770 CS 203 183

AUTHOR Rolfe, Elizabeth
TITLE Assessment of Spoken English.
INSTITUTION New Zealand Council for Educational Research, Wellington.
REPORT NO NZCER-RR-75
PUB DATE 75
NOTE 5p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Communicative Competence (Languages); *Evaluation Methods; Higher Education; *Language Research; Measurement Instruments; Oral Communication; *Oral English; Oral Reading; Rating Scales; *Speech Skills

ABSTRACT

This study examined the usefulness of an evaluation procedure designed to measure performance in spoken English. Rating involved assessment of the prose reading and conversation skills of 57 first-year students at Wellington Teachers' College, New Zealand. Specific topics of investigation included the consistency of "general impression" ratings between evaluators, the extent to which teachers can differentiate between factors on the rating scale, the degree of correlation between assessment of prose reading and conversation, the performance differences between younger and older students, and differences between evaluator ratings in a live interview and in a taped session. Many factors were found to influence the assessment of oral language--the personality of the evaluator, the number of evaluators used, and the administrative practicability of the test instrument itself. Other findings indicated that a high correlation existed between ratings of taped and live situations, that older students performed better than did younger students, that a fair degree of consensus was achieved between evaluators, and that prose reading and conversation were two different skills. (KS)

Item 10

Two of six photographs used in the Conversation Test described in this article.

set75
number one

ooken English

Assessment of Spoken English by Elizabeth Rolfe

Introduction

Emphasis on the spoken word in our English programmes has strengthened considerably over the last decade. The National English Syllabus Committee reflects the change by recommending that more attention be given to the teaching and evaluation of speaking skills. Although much research has been undertaken to clarify problems of reliability in assessing written English, little is known about the technical aspects of assessing spoken English. Some work has been undertaken by Hitchman and Wilkinson in England, and by Pountney in New Zealand, but many problems remain unsolved. Some of these have been investigated by the author in a recent research project, summarized briefly below.

Fifty seven first-year students at the Wellington Teachers' College were assessed on locally prepared tests of Prose Reading and Conversation. For the Prose Reading each student was required to read aloud two passages (chosen from six) — one dialogue and one straight description. For the Conversation Test each student was required to talk briefly about one of a set of six photographs (see examples of test materials). The testing sessions were taped and later marked by four assessors in addition to an on-the-spot assessment by the examiner. The administration of the whole test took about 10 minutes for each student, 6 minutes for selection of materials and 'thinking time' for the student, and 4 minutes for the actual test.

For the first part of the test, the Prose Reading, the student chose two passages, and then had a couple of minutes to glance over them before being asked to read aloud. For the second part, Conversation about a visual stimulus, the student was told that he should attempt to develop a theme independently of the examiner. The examiner was there to ask a few standardized questions at the beginning in order to get conversation started. Subsequently, the examiner was more of a sympathetic listener than an active participant in the conversation.

The rating procedure involved marking on separate factors (such as Interpretation and Delivery for Prose Reading) and then marking for General Impression (see rating scales for Prose Reading and Conversation). Previous work has shown that many assessors prefe general impression marking because it represents a unitary response to a student's spoken English performance. Such assessors would rather judge the whole performance unfragmented, considering the whole to be much more than the sum of the separate parts.

The main aim of the present study was to examine the reliability of evaluation in oral English. Subsidiary problems investigated were: the consistency of general impression marking, the extent to which teachers can differentiate factors on the rating scale, the degree of correlation between marks awarded for Prose Reading and Conversation, the difference between students just out of school and the 'more mature' students, sex differences in test performance, and finally the difference in score distributions of marks from the live situation and from tape recordings.

The Results

1. There was found to be a fair amount of agreement between the marks of individual assessors. The correlations clustered around 0.6 which is similar to those found in the marking of English essay-type answers. The agreement was highest in Prose Reading, particularly in the dialogue passages which required the students to be more expressive.

   The extent of the markers' experience was a factor affecting the consistency of assessments. The two more experienced markers indicated a higher level of agreement with each other than did the two less experienced markers.

2. The consistency of assessments increased with the number of independent markers involved in the evaluation. There was a noticeable increase in the mean correlations (0.62 to 0.69) when a single marker's assessments were compared with the average of **pairs of markers**. There was only a slight increase (0.69 to 0.72) when a third marker's assessments were added to the pool.

3. General Impression marks were shown to be almost as consistent as the composite marks resulting from summation of marks on the separate rating factors. This suggests that the General Impression mark is a satisfactory assessment on its own.

4. The results revealed that there was a considerable amount of overlap between the marks given for Interpretation and Delivery on the Prose Reading Test. Apparently teachers cannot effectively differentiate between these two factors. However, it may be justifiable to retain both as separate factors to be rated, provided markers receive sufficient training in how to discriminate between t

   The overlap between marks awarded Analysis-Content and Language on the Conversation was very high, indicating that the speech qualities that the assessors evaluated in both cases were more or less one and the same thing. This suggests that the two factors should be combined for rating purposes and renamed Content-Language.

   The Delivery factor of Conversation proved to be the most independent, i.e. teachers found it comparatively easy to mark this as a separate aspect of performance.

5. Assessments from the tapes suggested that Prose Reading and Conversation were two different skills. Good performance on one does not necessarily indicate good performance on the other.

Assessment of Spoken English by E. Rolfe, unpublished NZCER Research Report, 1975. Available from NZCER Library.

6. Students who had left school for more than one year performed much better on the Spoken English test than fellow students who were in their first year out of school. The fact that the older age group performed better is not surprising since the Spoken English tests appeared to favour those students with confidence, maturity and a well-developed personality. The words of one assessor were that "the older students had more clarity of thought and speech" while another stated that "their confidence and firmer voices" were their main advantages over the younger students.
7. The current study revealed no significant sex differences in performance on the Spoken English test. Female students performed slightly better in Prose Reading, but not in Conversation.
8. There was no difference between the distribution of marks (i.e. the average mark or the spread of marks) from the live situation and those from tape recordings. However, the comparison was based on the assessments of only one examiner. The correlation between marks from the live situation and those from tape was high. If this finding is confirmed in other studies, it would have important implications for testing practice in this field.

Conclusions

Many factors affect the validity of oral assessment. The quality of the marker is of supreme importance; an assessor should have the kind of personality that can calm an anxious student and encourage a shy student to talk. Experience at Spoken English assessment is an advantage and so too is the opportunity to meet and discuss problems with other assessors. Standardization of test materials and conditions can also help increase the amount of agreement between markers.

Administrative practicability is an important aspect of any testing programme. Spoken English tests are often regarded as impracticable because most are individual tests, thus very time consuming. However the research suggests that it would be possible to have group discussion tests being video-taped or tape recorded and evaluated later. This may not actually cut down the time factor but it may be a satisfactory way of simultaneously involving more students. Also, if the tests are being recorded, the teacher-examiner can give full attention to improving the quality of the test situation without having to be pre-occupied with marking 'on the spot'. The video-tape and tape recorder both appear as means of making Spoken English assessment more practicable. However, the technical equipment used must be of superior quality.

The research described above demonstrated that the pooled assessments of two markers from the tapes of individual students is a most efficient way to gain reliable evaluations of Spoken English. Therefore, teachers should give such assessments on two or more different sub-tests (such as Prose Reading and Conversation); this should preferably be done twice during the year. There would be no need to test all students at the same time; rather the testing could be scattered throughout the school year. The amount of test preparation to be done by students beforehand would be minimal, thus minimising "cramming" and anxiety. All tests could be taped to enable at least a **second** assessment to be made. A sample of the taped tests could then be assessed by expert external assessors for purposes of moderation, that is, determining comparable standards between schools. These external assessors could meet with teachers before the assessments are made, in order to discuss the use of the rating scale(s) and the qualities of Spoken English to be evaluated.

Although more research is needed on some of these problems, enough is now known about the assessment of oral English for such recommendations to be made with some confidence.

Footnote:

Spoken English, for the purposes of the study discussed here, was defined as follows:

i) the ability to read aloud passages of connected English prose and whilst doing so to reveal one's own powers of interpretation and appreciation;
ii) the ability to converse at some depth with an adult on a chosen subject.

The student's power to communicate mood and ideas was relevant to these two dimensions of Spoken English. Also relevant was the student's command of language and his ability to present ideas with a pleasant voice and clear diction.

Ideally any test of Spoken English should assess the wide range of speech situations that a person encounters in everyday living, such as casual greetings, 'small talk', conversation, group discussion, speech-making and reading aloud — all with varied purpose and audience. To make the exercise practicable, the Spoken English tested in the present study included just two of these.

Selected Bibliography:

Burniston, C., **Creative Oral Assessment — Its Scope and Stimulus** Pergamon Press: London, 1968.

Hitchman, P.J., "The Validity and Reliability of Tests of Spoken English", **British Journal of Educational Research** Vol. 36, 1966, pp. 15-23.

Hitchman, P.J., "The Testing of Spoken English: A Review of Research", **Educational Research**, November 1964, pp. 55-73.

Hitchman, P.J., **Examining Oral English in Schools**, Methuen & Co. Ltd., London, 1966.

Hitchman, P.J., "Examining Spoken English", **Examinations at Secondary Level**, Commonwealth Secretariat, London, 1970, pp. 36-39.

Poutney, C., **Evaluation of Third Form English**, unpublished Dip. Ed. thesis, Auckland University, 1971.

Southern Regional Board, **The Certificate of Secondary Education: Trial Examinations — Oral English**, Examinations Bulletin No. 11, Schools Council, H.M.S.O., London, 1966.

Wilkinson, A., and Stratta, L., "The Evaluation of Spoken Language", **Educational Review**, Vol. 21, No. 3, June 1969, pp. 183-195.

Rating Scale for Prose Reading

[a] Interpretation

9, 10: Delivery indicates a good understanding of the passage — skilful phrasing, fluent rhythm, expressive intonation, flexible use of pace and pause. Mood appreciated and communicated. Easy to listen to.

7, 8:
6, 5, 4:
3, 2:

1, 0: Delivery indicates poor understanding of the passage; phrases too long, too short, jerky or staccato rhythm; overdone intonation, flat, sing-song or otherwise monotonous intonation. Pace too fast, too slow or arhythmic. No appreciation of mood.

[b] Delivery < Voice / Diction (Mechanics)

9, 10: Easily heard. Accurate pronunciation. Variety of intonation. Strong, pleasant voice. Well-pitched. Clear crisp diction. Final consonants adequately defined. Unaffected.

7, 8:
4, 5, 6:
2, 3:

0, 1: Inaudible or too loud. Inaccurate pronunciation (i.e. sounds omitted, substituted or added). Monotonous. Weak, husky, nasal. Pitch too high or too low. Careless, defective diction.

[c] General Impression

9, 10: Good content, language and delivery. Overall very effective communication.

7, 8:
4, 5, 6:
2, 3:

0, 1: Poor on all aspects of this spoken English test. Made no impact on the listener(s).

Rating Scale for Conversation

(Revised as a consequence of the findings of the NZCER study)

[a] Analysis — Content (Ideas)

9, 10: Spontaneous and fluent presentation of ideas. Content of good quality, revealing some depth of thinking. Well-ordered arrangement of ideas. Shows ability to develop a theme. Coherence of ideas. Vocabulary and structure suitable and of adequate range. Ease of presentation. Convincing.

7, 8:
4, 5, 6:
2, 3:

0, 1: Finds it difficult to say anything, or is verbose. Ideas shallow and superficial. Ideas are muddled. Finds it difficult to develop a theme. Fails to keep to the point. Inadequate vocabulary. Uses slang inappropriately. Awkward presentation with too many pauses, false starts and gap-fillers. Lacking force.

[b] Delivery < Voice / Diction (Mechanics)

9, 10: Easily heard. Accurate pronunciation. Variety of intonation. Strong, pleasant voice. Well-pitched. Clear crisp diction. Final consonants adequately defined. Unaffected.

7, 8:
4, 5, 6:
2, 3:

0, 1: Inaudible or too loud. Inaccurate pronunciation (i.e. sounds omitted, substituted or added). Monotonous. Weak, husky, nasal. Pitch too high or too low. Careless, defective diction.

[c] General Impression

9, 10: Good content, language and delivery. Overall very effective communication.

7, 8:
4, 5, 6:
2, 3:

0, 1: Poor on all aspects of this spoken English test. Made no impact on the listener(s).

Note: The rating for General Impression should be done **after** the rating on the other factors. The **total effect** of the prepared talk or conversation is what is called for here.

Rating Scales are adapted from Hitchman, P.J., **Examining Spoken English** (Methven, London, 1970)