Role Play: Viable, Communicative Language Testing

Andrew Reimann

Utsunomiya University

The following reports on the process of using roleplay as a communicative form of language test to measure and evaluate student's oral language proficiency. If the goal of language teaching is to produce competent users of the target language and the criteria for an accurate test is to infer how those learners will fare in the real world, based on that instruction, then it would follow that a test be used, which mirrors these goals and criterion. This type of communicative test is therefore, geared towards eliciting representative language in real life situations. An effective way to achieve this, within the context of a classroom, is through a series of role play variations. This means of testing is viable because it provides a way to replicate the real life qualities of language and other non-linguistic factors, which are necessary for successful communication.

Comparing the results of a survey designed to elicit information regarding learner's attitudes towards tests in general and communicative testing methods, with student's comments and test scores, this report will endeavor to provide empirical support for the success of role play as a means of testing. Examples of materials, methods and procedures will also be shown and explained to demonstrate how this test can be easily implemented at any level or for any teaching context. In conclusion this presentation will provide a full spectrum of support and evidence showing how role play is both a valid and reliable form of testing, as well as a positive influence on learner's motivation, confidence, autonomy and language proficiency.

下記は、立証と解説を通して実用的で効果的な意味のあるロールプレイを用いて、伝達言語分析についての見識を提供するものである。最良の語学指導とは生徒中心であり、生徒の必要としているものを重視し、対象言語での的確な意思疎通者を生み出すことを目的とするなら、テストの種類はこれらの必要性を学習者がどのように実世界で言語を使用するかを反映するべきである。ロールプレイは正確な手段遂行と明確な興味を生徒に与える状況を作り出す事が可能と思われる。

Introduction

The following reports on a test designed to measure and evaluate student's oral language proficiency, after completing various university conversation courses. The goal of these courses was to use communicative means to expose the students to practical and authentic language, which they could practice and use appropriately, within context. The test was therefore, geared towards eliciting representative language in real life situations. The most effective way to achieve this, within the context of a classroom, was deemed through a series of role play variations. This means of testing was used because it provided a way to replicate the real life qualities of language and other non-linguistic factors, which are necessary for successful communication. Following Alderson's 1981 example of the *Cocktail Party* (p.58-59), if student's goals are to be able to successfully function and communicate with the language in everyday situations, then these types of situations should be reproduced and tested in order to generate the essential test qualities of validity and positive backwash. It may be argued that a role play carried out in the context of a classroom test, does not properly recreate all of the elements involved in communication in the real world, however, as this report will show, sufficient linguistic and non-linguistic factors, although not identical to those in the contexts being tested, are close enough to the real thing to be able to provide valid results and an accurate means for predicting degrees of success in future communication. Testers can isolate, manipulate and quantify any component of communicative competence within or out of context. There is, however, no accurate "flight simulator" for communication or language ability, guesses can be made based on various test scores, nevertheless, in determining language ability the "proof is in the pudding". Until a language learner is

"thrown into the deep end" and experienced the target language first hand, no score or means of measurement can accurately serve as an empirical predictor of success or failure. It is here that role plays can provide the context and authenticity that other tests lack.

Comparing the results of a survey designed to elicit information regarding learner's attitudes towards tests in general and communicative testing methods, with student's comments from interviews and test scores, this report will endeavor to provide empirical support for the success of role play as a means of testing.

Subjects

The subjects of this study were first, second and third year university students consisting of mostly English and Education majors. A total of 109 subjects from four different classes participated in the role play test and all other aspects of data collection. Almost all learners had 6 years of English language instruction and less than 20% had any experience abroad. Classes met twice a week and were limited in size to 28 students. The nature of instruction they had previous to entering university, consisted primarily of grammar translation, rote memorization and theoretical analysis of English. Although learner's goals are typically to improve overall communication skills and ability to function in English, complying with society` requirements, the final goal at the university, is improving level of achievement on the TOEFL test. The resulting preconditioning from previous instruction and negative backwash from contradictory language goals, results in a mixed and confused language learning environment. Student's motivations tend to be high but only when there is a direct and obvious link between a certain language task and

a short-term goal (i.e.) test score or final grade. Few of the students are able to develop a language for life attitude which, regardless of means of instruction, is essential to future success. In light of this, course materials and methods of instruction are aimed towards perking student's interests and fostering positive motivation and life long goals.

Research procedure

Context and instruction

Lesson plans and instruction are carried out with the intention of providing students with relevant and authentic language, dictated by their "communicative" needs, while also building their interest, developing learning strategies, motivations and confidence and giving them an opportunity to use the tools they have acquired in practical and realistic activities. Activities include both spontaneous and prepared role plays, dialogue reading in pairs and small groups, information exchange in the form of surveys and questionnaires, context specific problem solving and general task based interaction. A portion of each class and some subsequent assignments are not communicative in nature but are considered necessary in that they provide the learners with tools and skills, helpful for building confidence and increasing accuracy, appropriateness and finally, success in communicative situations. These activities include: keeping a vocabulary notebook in which they enter at least five new words per week, using each word in a sentence as opposed to writing a definition or just giving a translation, completing and filling in speech bubbles within a cartoon or picture of a specific context (see Appendix), listening/cloze exercise in which the students listen to a dialogue while reading a transcription and filling in missing words, a subsequent cloze exercise requires them to fill in missing parts of a conversation on there own

without listening. These various tasks provide the students with important structural tools and building blocks, which should help them navigate through unfamiliar and unpredictable situations. All material is presented first through listening tasks, followed by practice in pairs and finally modeling of a dialogue in front of peers. Review of material is carried out by periodically acting out unprepared dialogues, using only prompts or instructions. These reviews are often carried out in front of peers in order to provide the realistic sense of risk and anxiety found in authentic environments, while also accessing the confidence they have built through practice and the familiarity and safety of the classroom.

Testing

The actual test takes a variety of forms, in order to create a more complete picture of learner's ability and to maximize the amount of feedback and positive backwash returned. The first part of the test involves the creation of a role play and dialogue, which is to be performed in class. This work is done in pairs or small groups and the completed script is checked and scored based on authenticity and appropriateness of language, creativity, degree of realness and naturalness of context created. The students then have a week to prepare and practice their dialogues before performing them. They are encouraged to bring props, costumes or anything that might contribute to the scenario desired. Again these performances are evaluated based on authenticity, appropriateness and creativity, as well as self and peer evaluations. Finally these role play performances are video taped so that the students can watch, praise, criticize or have a good laugh at themselves. This method helps reduce the subjectivity and subsequent bias that is common in this type of testing, while also increasing reliability and empirical qualities, and giving the students more diverse and relevant feedback.

The second part of the test involves the random selection of a role play card (see Appendix) and the impromptu performance of that role play. Students do not have the chance to practice or prepare this role play and must complete the task on the card to the best of their ability, with the skills they have acquired during the class. This simulates the unpredictability and spontaneity of language situations in the real world and is in a way the "acid test" of language proficiency. The added pressure of the test situation and the observation of peers helps create the sense of urgency, anxiety and fear of making errors, that is present in authentic communication. In this way the test not only evaluates linguistic components but those non linguistics as well, such as critical thinking, creativity, brainstorming, communication strategies, degree of extraversion or reticence and risk taking ability, all of which are important factors of communicative competence in any language.

Other instruments

Three instruments designed to gauge student's performance and attitude and to gather data as to the effectiveness of this type of test were administered to all subjects. These methods were chosen on the basis of simplicity, ease of administration and richness and diversity of potential data which would in turn ensure validity of findings. The instruments will be presented and discussed as follows; interview, survey and the actual role play test.

The students were interviewed at three different occasions during the testing process, while creating their role plays, after their performance and after watching the video of their performance. The interviews were generally carried out in groups and at random. Although there is no way to track or compare individual responses to various aspects of the test, and measure subsequent changes, the objective of the interview was to collect general information concerning attitudes, perceptions and progress as well as to provide a third source of data through which to support overall findings. One of the most significant qualities of this type of test is therefore student feedback.

The second instrument, a survey with 18 questions aimed at gathering information on student's attitudes and feelings towards role play testing and activities, subsequently offering another source and perspective from which to establish validity, was given to all subjects at the end of the course (see Appendix). At the time of completing the survey, students had had time to reflect on the test and had been told their scores. This was done to erase any doubt, anxiety or mixed feelings the students might have had about the test, in order to get more accurate feedback.

The final means of data collection consisted of a combined analysis of test results. The scores for the role play test were fairly high and it is unclear how this may have influenced the primarily, positive feedback the students gave about the test. Although many students were well bellow average on some aspects of test criteria, such as gestures, fluency or pronunciation, in order to offer encouragement and positive backwash any student that basically "tried their best" was given credit. Therefore, there were also few extreme scores, emphasis being on providing positive washback with more criticism on the individual components and criteria of the test as opposed to the whole. It was thought more importantly to allow the students to feel positive about their basic communication abilities and constructively criticise them more directly on complimentary aspects of communication, such as mechanics, manner or creativity. The peer evaluations differed slightly in that there was more variance in the range of scores given.

However, there does not seem to be any significant or extreme difference between the two types of evaluation, other than the fact that in some cases the students were harder on themselves. This may reflect a degree of validity or that the students are more aware of their linguistic needs and what is required or expected for successful communication and to realize their goals.

Results and discussion

Interview & survey results

A summary of the most significant findings is as follows: (refer to appendices for more details)

- most important language skills speaking 68% listening 19%
- least important language skills writing 5% and reading 3%
- goals for using English traveling 32%, working 22%, making friends 12%, communication 15% general interest 13%
- average TOEFL scores were between 300-450
- 83%, felt that this score did not reflect their true ability
- 50% reported that they thought their actual level was higher
- 35% prefer speaking tests 15% eliminate testing 12% prefer listening tests
- 10% prefer writing tests 27% no response
- 80% role play test fun but difficult 12% easy 8% no comment
- 89% role play scores accurately reflected their ability
- 56% felt their ability was actually lower
- 72% thought peer feedback comments and evaluations were very useful
- 39% enjoyed video 24% did not enjoy the video 46% video most useful
- 36% video was very interesting, 32% video was embarrassing, 22% video was funny 6% video was a bad experience.

- 63% helpful and positive *"could see their mistakes, were shocked and surprised and changed their image"*.
- 24% felt video was negative *"looked bad and didn't want to see."*
- 55% more confidence to use English
- 76% participate in a role play again

Comments regarding evaluation included; *Positive (75%): "Honest, they (peers) know about me, get much information about my English." And Negative (25%): "not real, can't judge English skill well, friends give good score."* A further investigation into the merits of peer evaluation could be fruitful.

Test results

Score	Low	Mean	High	Mode
Role Play	60%	81%	92%	79%
Peer Evaluation	40%	78.3%	90%	80%
TOEFL	54%	69.2%	87%	68%

Students generally dislike tests as they expose their weaknesses and insecurities and attempt to evaluate and assign a score to something that is very personal, psychological, sensitive and for the most part indefinable; language ability or the inevitable lack thereof. Though the students hold a great deal of anxiety towards the role play test before hand, after the fact their comments from interviews, are mostly positive. Students often expressed that the role play was their most memorable part of the class and that they felt that they really learned a practical skill and accomplished something with language. After watching the video one week later students tend to be

surprised by the level of their ability. Most students had a much lower image of their language proficiency and were pleasantly surprised. Others thought that they had no problem with "l's", "r's" and "th's" and discovered they need to work on these areas. All in all, the video, self and peer evaluations are among the most important aspects of this test, insomuch as they give the students a clear picture of where they are in the ambiguous scale of language ability, which is much more useful than an abstract A, B or C. Further, this type of evaluation allows the students to discover first hand, where their strengths and weaknesses lie and perhaps gives them insight into techniques or strategies, which they can use to improve. Self-discovery in language acquisition is invaluable and much more effective than instruction or teacher's comments, in bringing about desired results. Finally, the amount of positive backwash generated by this test, establishes its merits. The role play test actually measures and evaluates what the language instruction is preparing the students to do in the end. Although their short term goals include passing tests, getting good grades and increasing their TOEFL scores, the students come to realize subconsciously if not overtly, what the purpose of language instruction is and what they can do to improve their proficiency. These aspects of the test are confirmed by student's responses to the survey questions.

From feedback collected from the post test survey and through student interviews, it would appear that the overall reaction to this type of test was positive. The ethnographic approach to analyzing and collecting data from various sources and perspectives through triangulation, provides an acceptable level of reliability. Comments from students seem to also provide a high level of validity in support for this test. The communicative aspects of the test were further successful

in raising student's awareness of their own language ability, helping to critically analyse their own and their peer's ability giving them meta-linguistic knowledge necessary to exercise some control over their learning, build confidence and perhaps also provide them with a direction towards which to focus and develop learning strategies. Perhaps the most significant results are those that imply a positive change in learner's attitude towards tests, participating in communicative activities and using their language for practical purposes.

Further, comparing role play test results with TOEFL scores supports the validity of the role play test (assuming the TOEFL itself is valid), as the distribution trends seem representative of the class, matching extremes with the bulk of the scores in the median. For purposes of comparison, the TOEFL scores were converted into percentiles from rankings out of ten. Granted, the TOEFL is a different style of test, evaluating completely different and non communicative aspects of language, the representative distribution of the scores may be interpreted as support for the validity of the role play test. Although TOEFL is not a comparable test per se, it remains one of the most widely used international standards for measuring English proficiency and therefore qualifies as a suitable benchmark by which to compare the results described in this study. Of particular interest, is the fact that the TOEFL average is considerably lower than the role play averages. The reasons for this might be due to any number of factors and variables but probably significantly involve the lack of representative and meaningful content and skills and the unclear connection with student's ultimate goals and language needs.

From the survey results, it becomes clear that students responded positively to this type of test and that the results

were more meaningful, representative and valid. This may reflect their need for English, their goals and perhaps also their weaknesses and insecurities, suggesting also that students need a more skill based approach to learning as opposed to their previous language based instruction. From the extreme polarization of some responses, it becomes evident that students are interested in practical language for specific goals and purposes and want to be able to function or communicate at some level through the language. The inconsistency of responses to type of test preferred, considering that they felt strongly about standardized testing, may signify that they are unsure of their speaking skills or that they have no idea of what form a test should take. These skewed results may also be due to the fact that the majority of scores for the role play test were high and that comments from peers were positive. With regard to the evaluation and feedback, most students reported that their favourite part was the video and the actual performance in front of the class. This is interesting in light of the fact that most of them were terrified of this prospect, when the activity was first introduced. Similarly, the majority of students did not enjoy evaluating each other. This might be explained by cultural traits, peer pressure or insecurity and general confidence issues.

Conclusion

Without administering a further written or standardized proficiency test it is impossible to make any generalizable description of student's language proficiency. Nor is it possible to accurately predict how these students will improve in the future or fair, when they encounter a real situation. What this test does accomplish however, is clearly showing the benefits of communicative language testing and going a long way to answer Spolsky's question: "What exactly does it mean to know

how to use a language." Unfortunately communicative language testing is still too subjective and unquantifiable to be used as a standard form of evaluation in most institutions. Similarly the resources and time requirements of this type of test, either strain or greatly exceed most budgets and limitations. Nevertheless, the goal of language education is undoubtedly first to create competent speakers of the desired language and lastly to provide an arbitrary score or description of what experts hypothesize a TOEFL 320 or 550 student should be capable. Educators would most likely unanimously agree, yet the system and the tests used in their classrooms and institutions remain largely unchanged, consisting of predominantly abstract measures of quantifiable structures in isolation, which do not further this goal but actually serve to hinder its achievement. Feasibility and practicality are the arguments used to dismiss large-scale administration of communicative tests. The long-term affect on society of producing low quality, incompetent speakers of language is much more costly.

References

Alderson, J. C. (1981). Report of the discussion on communicative language testing. In J. C. Alderson and Hughes (eds.) *Issues in Language Testing. ELT Documents*. 11, London: The British Council.

Skehan, P. (1991). Progress in language testing: the 1990s. In J.C. Alderson and B. North (eds.) *Language Testing in the 1990s*. London: Macmillan.

Spolsky, B. (1985). What does it mean to know how to use a language? An essay on the theoretical basis of language testing. *Language Testing, 2*, 2, 1985.

Appendices

English through drama, role play or dialogue

In Groups of 4 you must write and act out a dialogue or a role play of a real life situation.
• This activity will be performed in class.
• All group members must participate equally.
• The dialogue should be at least 5 minutes long.
You should use real life situations that you may encounter at this university, in Japan or abroad and use real language and communication strategies.

Important points are: creativity, realistic language, communication ability and strategies such as eye contact, smile, hand gestures, and physical contact (shaking hands etc…).

You may use examples from the text but try to be creative and original.

Suggested Scenarios: Asking for information or directions, Going Shopping, booking a hotel room or making a reservation, getting lost in a big city or on campus, ordering in a restaurant, in a bank, in a post office. In a hospital, talking to the police, emergency situations, telephone conversation or anything else you find interesting. Please ask me if you have any difficulties deciding a theme.

Dialogues and Role Plays will be acted out on the last day of class before summer vacation.

This activity will be used to evaluate you as a test. Therefore you must participate and try your best!

Member Names	Theme/Role Play	Characters

Role play evaluation and critique form

Name			
Language Authenticity			
Gestures			
Eye Contact			
Volume/ Fluency			
Voice/ Pronunciation			
Creativity			
Humor			
Attitude/ Delivery			
Comments			

Role play self-evaluation and critique form

Tables and charts

Give a score out of 10 for yourself and your group members.

Name			
Language Authenticity			
Gestures			
Eye Contact			
Volume/ Fluency			
Voice/ Pronunciation			
Creativity			
Humor			
Attitude/ Delivery			
Comments			

Language Needs Analysis

- Speaking
- Listening
- Writing
- Reading
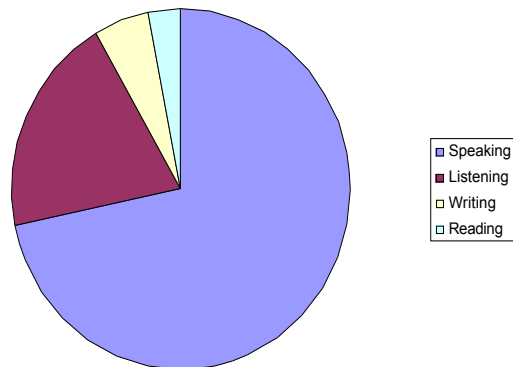
Long Term Needs Analysis

- Travel
- Working
- Communication
- General Interest

Test Preferance

- Speaking Test
- No response
- No Test
- Listening Test
- Writing Test

Test Score Avg.

- TOEFL Score
- Peer Score
- Role Play Score

6.00   6.50   7.00   7.50   8.00   8.50

Role Play Feedback

- Difficult
- Easy
- No Comment
- Fun

Test Validity

Lower
Higher
Role Play Invalid
Role Play Valid

Lower
Higher
TOEFL Invalid
TOEFL Valid

0%   20%   40%   60%   80%   100%
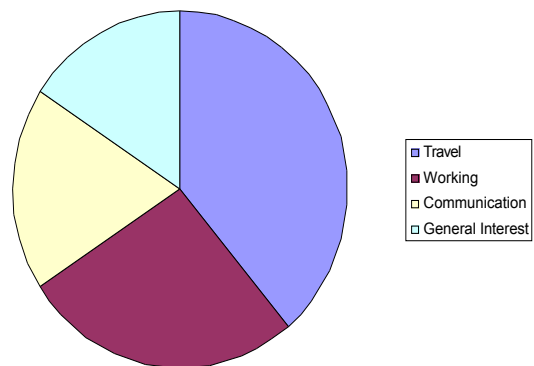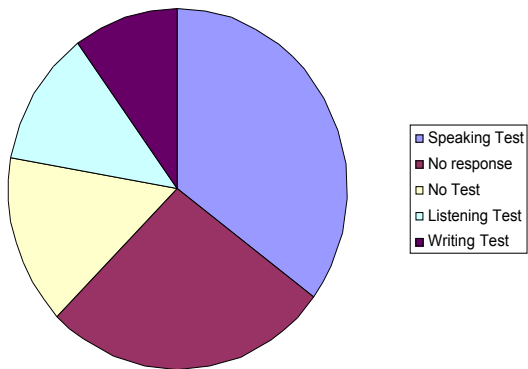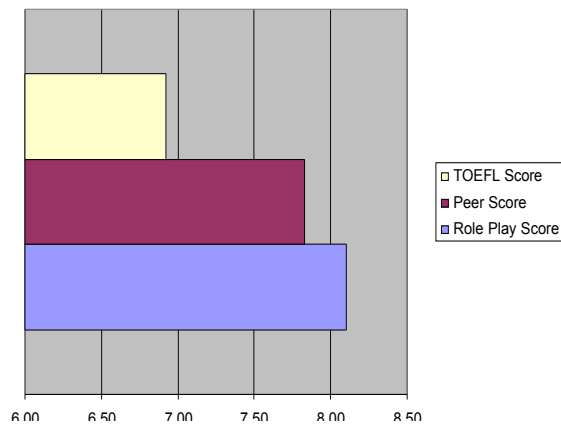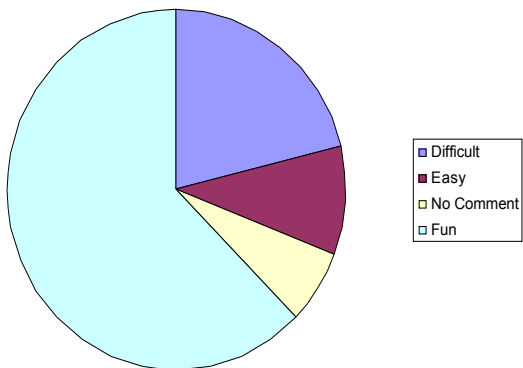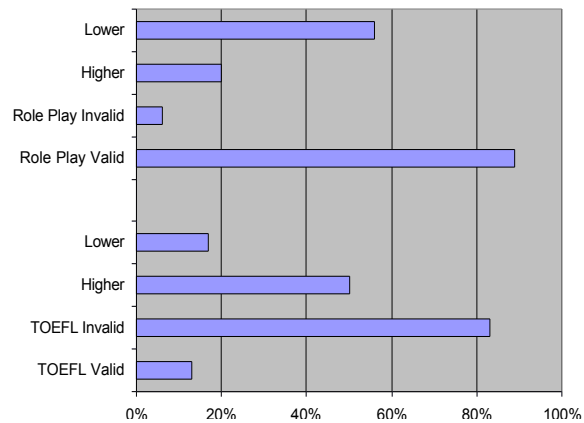
Role play activity survey

1. What is the most important English skill for you?
   Speaking   Listening   Reading   Writing

2. What is the least important English skill for you?
   Speaking   Listening   Reading   Writing

3. What do you want to use English for in the future?

4. What is your TOEFL score?

5. Do you think your actual ability is higher or lower than your score?
   higher   lower

6. Do you think this test reflects your ability?
   Yes   No

7. How would you change the testing system?

8. What did you think of the final test in this class?

9. Do you think your test score for this class reflects your ability?
   Yes   No

10. Do you think your actual ability is higher or lower than your score?
    higher   lower

11. Which part of the Role Play did you like best?

12. Which part of the Role Play did you like least?

13. Which part of the Role Play was most helpful for you?

14. How did you feel watching a video of yourself speaking English?

15. Did this help you?
    Why/why not?

16. Would you like to do a role play again?
    Yes   No

17. Do you have more confidence to use English now?
    Yes   No

18. Do you think evaluation from other students is helpful?
    Why/why not?