Classroom Research in Japan: Issues, Studies and Prospects

Peter Robinson
Aoyama Gakuin University

Abstract

How are learner perceptions of task difficulty and the cognitive demands of tasks related? How effective are techniques such as recasting for delivering implicit negative feedback about form during communicative activities? How should syllabus designers operationalise and sequence the units of classroom activity? These issues—all susceptible to empirical research at the interface of language pedagogy and second language acquisition theory—are addressed and provisionally answered by the studies reported in this paper. Prospects for future research arising from the findings, and addressing similar issues, are discussed in the conclusion.

Key Words: Tasks, cognition, feedback, syllabus design

INTRODUCTION

While there has been some basic SLA research in Japan, by far the most prevalent type of research is applied research, particularly SLA research which is pedagogically motivated. While some of this research has been conducted in laboratory settings, most is conducted in classrooms. In this paper I identify three issues that current applied classroom research could, and in many cases has addressed. I describe these issues in the context of reporting the results of three classroom research projects by myself and others. These issues are: i) the relationship between learner perceptions of task difficulty, the cognitive complexity of the task, and learner production on task, ii) the effectiveness of teacher intervention during task performance to shift learner attention from content to form via implicit negative feedback, and iii) the larger issue of syllabus design and the development of holistic communicative ability, specifically the development of pragmatic and conversational skills necessary for oral academic task discussions. In the sections which follow I briefly report the results of studies addressing each of these issues in turn. Each of these issues is clearly related to the others, and an empirical basis for choosing options in the first two issues (what kind of task and what kind of feedback) is essential to motivating a choice of syllabus type, and operationalising it (the third issue). At present each of these issues is far from resolved, and further progress towards resolving them will necessitate findings from classroom research focussed specifically on the Japan context, accompanied by support and motivation from controlled experimental laboratory research taking place within, and outside Japan.

1. TASK DIFFICULTY, TASK COMPLEXITY AND TASK PRODUCTION

1.1 Task Difficulty and Task Complexity

In task-based approaches pedagogic tasks are not sequenced on the basis of their language content but according to variously defined notions of increases in task complexity or difficulty. A crucial research issue is therefore how to establish empirically valid criteria for assessing task complexity and for assessing the effects of task complexity on production and learning (see Robinson, 1995a; Skehan, 1996; Skehan & Foster, 1998; Norris, Brown, Hudson & Yoshioka, 1998). However, in many discussions of task sequencing the terms *complexity* and *difficulty* are used interchangeably and the scope of potential influences on them are argued to be very wide, including cognitive, affective, social, linguistic, response type, text type and many other disparate and often noncomplementary factors. This has led to some confusion which has hindered focused research on task effects on learner performance and learning and obscured the implications of findings for sequencing and syllabus design. With this in mind, and following Spilsbury, Stankov and Roberts (1990) and others engaged in psychological enquiry into task effects, I use the terms complexity and difficulty to describe two distinct influences on the likelihood of achieving success on a task (see Robinson 1997, 1998b). These are summarized in Figure 1 and elaborated on below.

Task complexity	Task conditions	Task difficulty
(cognitive factors)	(interactive demands)	(learner factors)
a) resource directing	e.g., open/closed	a) affective variables
e.g., +/- few elements	one way/two way	e.g., motivation
+/- contextualisation	convergent/divergent	anxiety
+/- no reasoning demands		confidence
b) resource depleting		b) ability variables
e.g., +/- planning		e.g., aptitude, proficiency,
+/- single task		intelligence,
+/- prior knowledge		cognitive style
Sequencing criteria		Methodological criteria
Prospective decisions		On-line decisions
about task units		about pairs and groups

Figure 1. Task complexity, condition and difficulty

Task complexity, I argue, is the result of the attentional, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner, and these are relatively fixed and invariant (Suedfeld & Coren, 1992). That is, for any given learner, an easy task (like adding two single digit numbers, or like giving directions from A to B using a simplified map of a small area with which the speaker is familiar) will always be easier than a hard task (like doing calculus, or like giving directions from A to B and then C using a detailed map of a large area the speaker is unfamiliar with) since the differential in the extent of attentional focus, working memory, reasoning and other cognitive demands is a fixed and *invariant* feature of the tasks. Task complexity, then, will explain *within learner* variance when performing any two tasks. The effects of valid complexity differentials should be revealed by the fact that the cognitively less complex task will always involve a lower error rate, or be completed faster, and so be less cognitive resource demanding (what I propose as the actual cause of these differences) than the more complex task (disregarding the possibility of differential amounts of practice on each, which has the effect of reducing resource demands—a separate factor again, see DeKeyser, 1998).

I distinguish the independently defined *cognitive factors* contributing to complexity, and which are a consequence of the structure of the task which imposes resource *demands*, from *learner factors* which may make a task more or less difficult (as opposed to complex), and which are a consequence of differentials between learners in their available attentional, memory and reasoning *resource pools* i.e., differentials in the resources they draw on in responding to the demands of tasks (see Robinson, 1998b). So in one scenario learners differing in intelligence or aptitude may find a simple maths or language learning task more or less difficult than each other as a result of these inherent ability differentials between them. That is, performing simple addition (versus more complex calculus) may be more or less difficult for two learners because of relative differences in the size and flexibility of the *potential limits* of their attentional, memory and reasoning resource pools as reflected in measures of intelligence, or aptitude for math. But resource pools can also be affected by such *temporarily limiting* factors as motivation. That is, greater motivation for one learner can result in a temporary expansion of resource pools currently available to meet task demands relative to another learner equivalent or even superior in intelligence or aptitude but having less motivation, and so fewer temporarily available resources. Task difficulty, then, I argue, will explain *between learner* variance on performance on any one task, (either easy or complex, cognitively defined) because it is determined by those learner factors contributing to differences between them in the extent of (temporarily or permanently) available cognitive resources.

Consequently I argue that the learner factors contributing to task difficulty, and so to between learner differences in the size of resource pools available to meet task demands, are of two kinds; a) *affective* variables, such as confidence, motivation, and anxiety, which may change and so affect the size of resource pool availability on a temporary daily, or even minute-by-minute basis; and b) *ability* variables, such as poor eyesight, intelligence, aptitude, and cognitive style which are more permanent and fixed determinants of resource pools, or at least more stable over a course of instruction than changeable (and methodologically susceptible) affective variables. Affective variables cannot be anticipated by the materials and syllabus designer. However, they are extremely important to on-line methodological decisions about pairing and grouping students when implementing the syllabus. Ability variables such as aptitude, intelligence and cognitive style can be diagnosed ahead of syllabus implementation but there is as yet no clear research evidence of stable and predictable interactions between them and second language task complexity. This is a much needed area of research in SLA and an active one in the study of task performance in other educational domains (see Ackerman, 1989; Fleishman, 1978; Snow, 1984, 1989; Stankov, 1988; Stanton & Keats, 1986; and see Hulstijn, 1989; Skehan, 1998; and Sawyer & Ranta, 1998 for SLA related work). A fourth ability variable, proficiency, is often the basis of placement into different levels of a program. Therefore differences in proficiency level between learners will likely not be a factor that is available for on-line methodological manipulation since the result of placement decisions is often to homogenize the proficiency level of groups. However, proficiency level will, of course, influence decisions about task content and the level of language input to a task.

The present study sought to examine the relationship between task difficulty, as assessed by responses to a questionairre assessing affective factors completed immediately after task performance, and the hypothesized difference in complexity of two map tasks. The main research question was: "Is the cognitive complexity of the tasks matched by learner perceptions of differences in their difficulty?" One would expect that it would be. A second research question was: "Is language production qualitatively different on easy and hard tasks?" Elsewhere I have argued (with some support, see Robinson, 1995a) that cognitively complex tasks will push learner output more than easy tasks, and therefore increases in complexity, and in some cases accuracy can be expected relative to easy tasks, but with decrements in fluency on harder tasks (see Robinson 1995a, 1997, 1998b, 1999b). Finally, the issue of sequencing was examined. The research question here was: "Does sequencing tasks from easy to hard versus hard to easy affect learner perceptions of difficulty?" There is some evidence from applied psychological research that perceptions of difficulty are affected by the order in which easy and hard tasks are performed (see Hancock, Williams & Manning, 1995; Robinson, 1998b, 1999a) but there has been no SL research into this issue.

The questionnaire described in more detail below consisted of Likert scale ratings of such affective factors as perceptions of difficulty, motivation, stress, interest, as well as perceptions of ability, fluency and accuracy. Questionnaire responses to the easy and the hard tasks were compared, and the questionnaire data were also measured against actual evidence of task production to see if learner perceptions of difficulty were reflected in differences in production. Since the tasks involved were one-way information gap activities (described in more detail below) the speaker's production was assessed for accuracy and fluency using measures described below, and the hearer's production was examined for interactive units, such as confirmation requests and clarification requests.

1.2 The Students

The students were 44 university undergraduates, aged between 20-24 at Aoyama Gakuin University. All participated voluntarily in this research.

1.3 The Tasks and Research Design

Students were randomly assigned to the role of information giver (speaker) or receiver (hearer) on two map tasks. One map task (the easy one) required the speaker to give directions using a map covering a small area which was familiar to them, Aoyama Gakuin campus. The hard task required them to give directions using a large map of an area they were unfamiliar with, taken from an authentic street map of the Nihombashi area in central Tokyo. The two dimensions of complexity distinguishing the cognitive demands of these tasks , amount of information, and prior knowledge, are related to other dimensions in Figure 2 , and are described in more detail in Robinson, Ting and Urwin, 1995, and in Robinson 1998a, and 1998b. Task sequencing was controlled for. Half the students were randomly assigned to perform the easy task then the hard task, and half to perform the hard task followed by the easy task.

1.4 The Rating Instrument

Immediately following each task both students completed a questionnaire. This is shown in Appendix A. Each item was rated on a nine point Likert scale. Responses to five items assessing difficulty, stress, perceived ability and interest and motivation, were used to assess learner perception of task difficulty. In addition, learner perceptions of their own accuracy and fluency during task performance were rated in order to measure them against actual performance data.

	Simple				Complex
	1	2	3	4	5
Dimensions of complexity					
planning time (before speaking)	+	-	-	-	-
single task (route marked)	+	+	-	-	-
prior knowledge (of a familiar area)	+	+	+	-	-
few elements (a small area)	+	+	+	+	-

(simplified data/map)   (authentic data/map)

(from Robinson, 1998b)

Figure 2. Component Dimensions of Map Task Complexity

1.5 The Measures of Production

Speaker production was assessed for accuracy, fluency and complexity. To measure accuracy a specific measure (target like use of articles—TLU) and a general measure (error free T units—EFT) were used. Complexity was assessed by measures of S nodes per T unit (SperT), and type token ratio (TTR). Fluency was assessed by a measure of words per T unit (WperT).

1.6 Results.

Results of the repeated measure ANOVA show a significant effect of Task Complexity on Ratings of Task Difficulty. The hard task is rated more difficult than the easy task (Figure 3). There is an effect approaching significance for Sequence. The Hard to Easy sequence results in higher difficulty ratings for the hard task. And there is a significant interaction of Role and Difficulty ratings. Hearers rate their ability to be higher, and tasks to be less difficult and stressful than speakers.

Results of the repeated measure MANOVA of speaker production on easy and hard tasks shows no significant effect for Task Complexity (Figure 4). Thus more complex tasks do not appear (in this operationalisation) to affect the quality of speaker production. The repeated measure MANOVA shows a significant effect for Task Complexity on measures of hearer interaction (Figure 5). Thus more complex tasks DO result in greater amounts of negotiation work than their simpler counterparts. Correlations of speaker production and affective variables measured by the questionnaire data are significant, though correlations of hearer interaction and

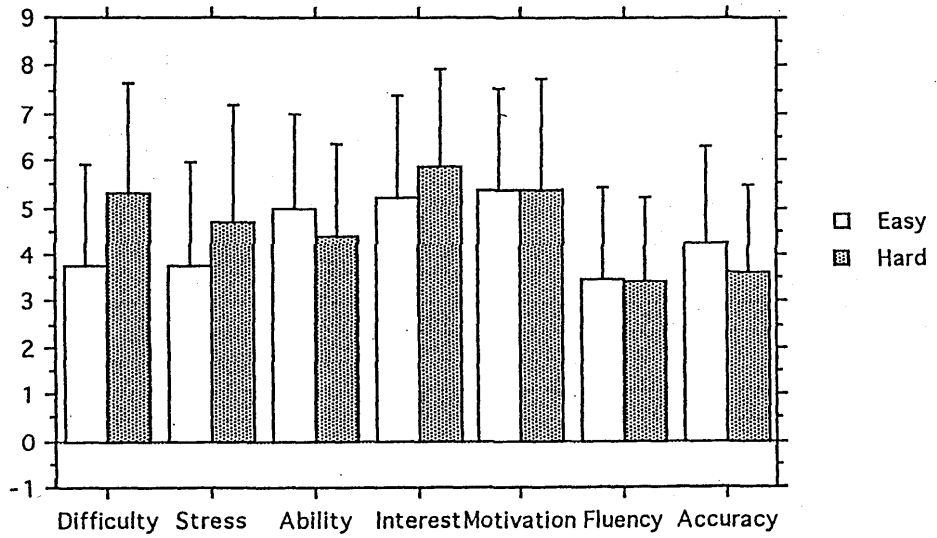affective variables are not significant.

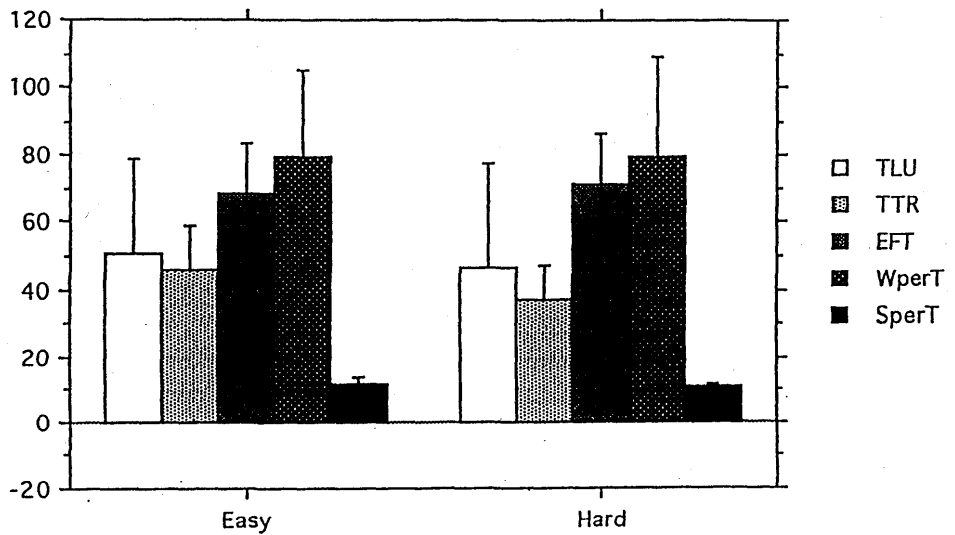**Figure 3.** Difficulty Ratings for Easy and Hard Tasks

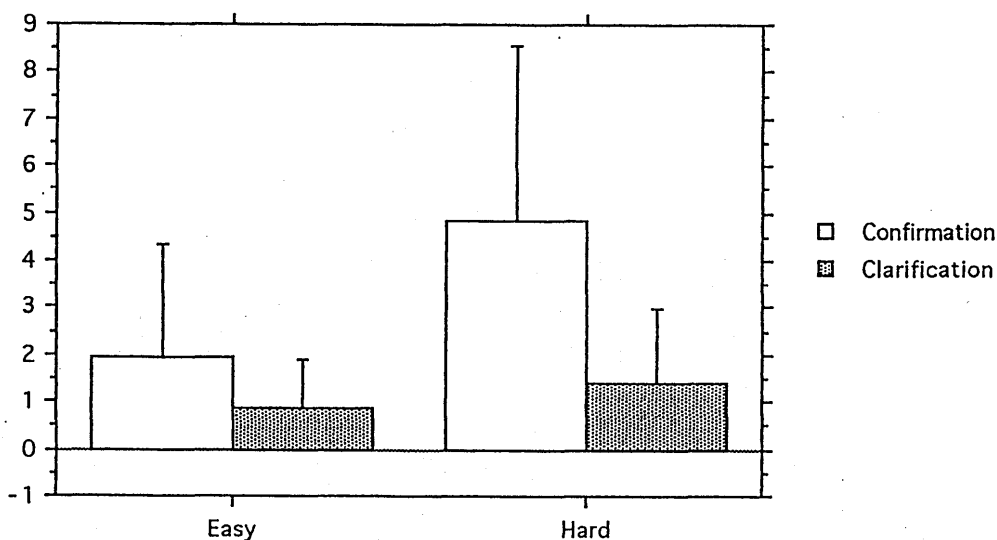**Figure 4.** Speaker Production on Easy and Hard Tasks

Figure 5. Hearer Interaction on Easy and Hard Tasks

1.7 Conclusions

1. The difficulty measures support the proposed complexity differential. More complex tasks are judged to be more difficult, more stressful, and less easy to achieve than simpler counterparts. Motivation is equivalent, though interest is higher on more complex tasks.

2. The measures of speaker production do not support the proposed complexity differential, since there are no differences in speaker production on the two tasks, in either accuracy, fluency or complexity.

3. The measures of hearer interaction do support the proposed complexity differential. There is more interaction and negotiation work on the more complex task.

4. Speaker production (though not hearer interaction) is significantly correlated with their perceptions of the difficulty of the task- specifically WperT which correlate negatively with difficulty, stress, ability and motivation, and SperT which correlate negatively with difficulty and stress on the easy task. These are clearly related findings since a reduction in the number of words per t unit must restrict the number of s nodes. Similarly, there are negative correlations of WperT and SperT with stress on the hard task, as well as a significant positive correlation of TTR with stress. There is also a significant positive correlation of TTR with perceived ability on the hard task.

5. Task sequencing is nearly significant, suggesting that where easier versions are performed first, the complex task is rated less difficult than when it is performed first- a finding with clear implications for sequencing criteria (simple to complex) to be adopted in syllabus design.

2. TASK FEEDBACK AND GENERALIZABILITY OF FOCUS ON FORM

2.1 Implicit Feedback, Aptitude and Focus on Form

Aptitude, and other individual difference interactions with the effects of different instructional treatments are important for understanding the differences in mental processes that *cause* the effects of instruction for specific learners (see Robinson, 1997a, 1999c). A few studies of the relationship of aptitude to learning have been reported recently, for example Sasaki's work on the interrelation between aptitude, intelligence and proficiency (1996), and experimental laboratory studies of the relationship between aptitude and learning under different conditions of awareness (DeGraaff, 1997; DeKeyser, 1995; Robinson, 1997a). However, none have addressed the effects of individual differences in aptitude on uptake of implicit negative feedback, delivered via recasts. Recasting has been proposed as a candidate for methodologically manipulating a focus on form (Doughty & Varela, 1998; Long, 1997; Long & Robinson, 1998); that is, momentarily drawing learner attention to the form of their output while maintaining an overall focus on the communicative demands of a task (see Aline, 1999 and Muranoi, 1999 for recent studies of alternative techniques for manipulating focus on form in Japanese classrooms, and Doughty and Williams 1998 for a full taxonomy of available techniques and a rationale for their use). Much of the early aptitude research (see Carroll, 1993 for an overview and Skehan 1998 for commentary) examined the effects of aptitude on learning under classroom conditions where audiolingual, or explicit grammar teaching methodology was prevalent. In contrast to these methodologies, task-based learning accompanied by focus on form is likely to draw much more heavily on learner's inductive, incidental learning capacities, and it is important to see whether, and to what extent, aptitude tests measure these successfully, predicting differential success in learning in task-based classrooms.

This second study reports on empirical findings of the effects of implicit negative feedback targeted at specific forms in learner oral output and examines the relationship of learning to aptitude, as measured by Miyuki Sasaki's Japanese aptitude test. The targeted forms were object of comparison relative clauses, e.g., *Brad Pitt is the actor who Sean Connery is older than* —the most marked relative clause type on Keenan and Comrie's Accessibility Hierarchy (1977). In a classroom study, using a different instructional method (rule explanation and practice) Eckman, Bell and Nelson (1988) found that instruction on more marked relative clauses (indirect object relatives) generalized to less marked (subject and object relatives) but the reverse was not the case, instruction on subject relatives did not generalize to more marked forms of relative clause (see Beglar, 1999 for a recent partial replication in Japanese classrooms). Eckman et al. assessed learning using grammaticality judgment and sentence

combining data. The present study additionally assessed listening ability (via an aural grammaticality judgment task) and speaking (elicited imitation) using a pre-post test and delayed posttest design. Our research questions were: "will implicit feedback and task practice focussed on object of comparative relative clauses generalize to other types of relative clause, leading to improvement on all types, on all measures?"; and "will aptitude measures correlate significantly and positively with gain scores?"

2.2 The Students.

Students participating in the study were 58 enrolled in two undergraduate English courses at Waseda University, who formed the experimental treatment groups, and 55 enrolled in another undergraduate English course who formed the control group. All students accepted the task and testing requirements as part of regular coursework though all were told performance on both would not be assessed and did not count towards their final grade for the course.

2.3 The Structures Targeted

These were object of comparative relative clauses, and the types tested in the pre and post-test were, in the grammaticality judgement test , those 55 types used by Hawkins and Chan (1997) in their study of the failed functional features hypothesis. In the other tests (listening, speaking, and sentence combining) two items of each of the 6 relative clause types identified by Keenan and Comrie (1977) were included (and changed each time for content and sequence on the pre, post and delayed post-tests).

2.4 The Tasks and Feedback Procedure

Four tasks were given over the course of the five week treatment. A sample of these is given in Appendix B. Initially, the researchers modelled task performance for the students, then assigned students to random pairs to complete an easy and a harder version of the task (easy tasks had less information to select from than hard tasks). During task performance by the treatment group the researchers (two) circulated, delivering recasts to pairs when aberrant relative clauses were used. After the initial 15 minute settling in period, accommodating late arrivals, the modelling procedure began and took about 15 minutes followed by the pair work, which took about 50 minutes. Following this there was a 15 minute game wind down activity when the researchers asked questions concerning the task which were targeted at the structures, and students as a group answered at will. There were 4 sessions in all.

2.5 The Measures Taken and Experimental Design.

The experimental design was a pre-test (week one), post-test (week five) and delayed

post-test (week ten) for the experimental treatment groups, and the same for a control group who received no task based instruction or feedback. Measures taken were; i) elicited imitation (pre and post-tests for treatment groups only); ii) grammaticality judgment of 55 items (pre, post and delayed post-tests for all groups); iii) listening grammaticality judgment (pre, post and delayed post-tests for all groups); and iv) sentence combining (pre and post-tests for all groups. The measure of aptitude was one developed by Sasaki (1996) based on the MLAT and Pimsleur's PLAB. This was administered after the treatments were delivered, but before the delayed post-test.

2.6 Results.

Only results from the pre post-test comparisons of the listening test show significant gain for both treatment groups, with a significant difference relative to the pre-test score which is maintained on the post-test by one of the treatment groups (Figure 6). (The elicited imitation data has yet to be analysed). The only correlations with aptitude are for delayed post-test results on the grammaticality judgement test and the memory component of the aptitude test
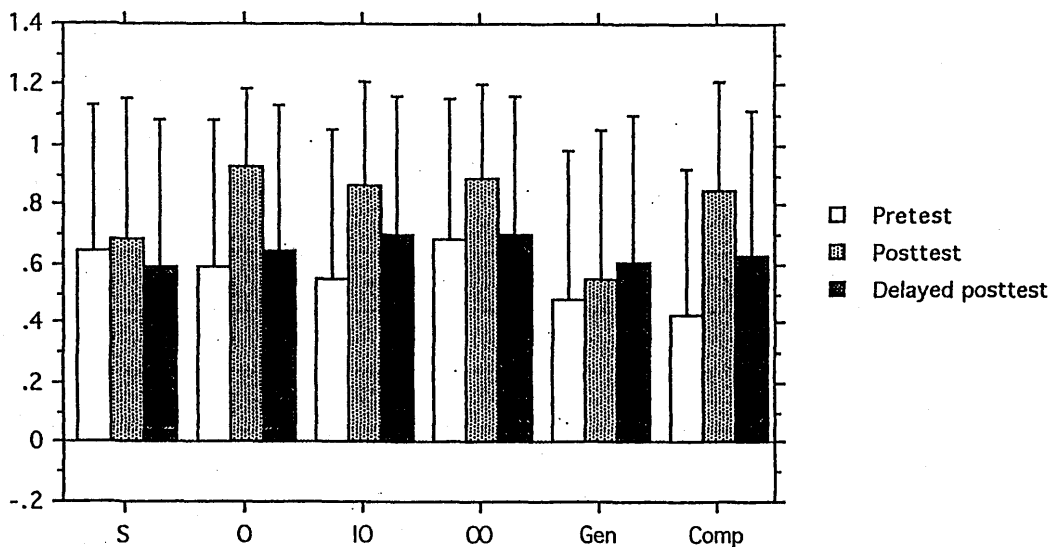

Figure 6. Results of the ANOVA for Pre Post-test Performance on the Listening Grammaticality Judgement Test of Relative Clause Types

2.7 Conclusions

1. Focus on form during task based learning, delivered via recasts has significant effects on listening comprehension/grammaticality judgment alone, and these were maintained by one group in a delayed post-test.

2. Instruction, as measured by performance on the listening test, generalizes from marked to less marked structures, with the exceptions of genitive relative clauses and subject relatives.

3. The memory component of the aptitude test contributes to ability to perform on the delayed grammaticality judgement test, but on no other tests.

3. SYLLABUS DESIGN AND ORAL ACADEMIC TASK ABILITY

3.1 Two Approaches to Syllabus Design

A useful distinction in conceptualising options in syllabus design was made initially by Wilkins (1976; see also Long & Crookes, 1992; Nunan, 1988, Robinson, 1998a; White, 1988) and refers to the learner's role in assimilating the content provided during group instruction and applying it individually to real world language performance and interlanguage development. *Synthetic* syllabuses involve a focus on specific elements of the language system, often serially and in a linear sequence, such as grammatical structures, language functions or reading and speaking microskills. The easiest, most learnable, most frequent, or most communicatively important (sequencing decisions can be based on each of these ultimately non-complementary criteria, and on others) are presented before their harder, later learned, less frequent, and more communicatively redundant counterparts. These syllabuses assume the learner will be able to put together, or synthesize in real world performance, the parts of the language system (structures, functions, skills etc.) that they have been exposed to separately in the classroom.

In contrast, *analytic* syllabuses do not divide up the language to be presented in classrooms but involve holistic use of language to perform communicative activities. One version of an analytic syllabus is adopted in task-based approaches to language teaching (see Skehan, 1998; Long, 1985, in press; Norris, Brown, Hudson & Yoshioka, 1998; Robinson, 1998a, 1998b).The learner's role in these syllabuses is to analyse or attend to aspects of language use and structure as the communicative activities require them to, in line with: a) their developing *interlanguage systems*; b) preferred *learning style* and *aptitude profile*; and c) the extent to which they are *motivated* to develop to an accuracy level which may not be required by the communicative demands of the task. Additionally, interventionist teacher techniques can be used during or following task performance to draw learners attention to aspects of task performance that are non target like, but are judged to be learnable and remediable (see Doughty

& Williams, 1998; Long & Robinson, 1998). For these reasons researchers have argued that analytic approaches to syllabus design, accompanied by focus on form techniques, are more sensitive to SLA processes and learner variables than their synthetic counterparts and do not subvert the overall focus on meaning and communication encouraged during classroom activity .

This third study represents an initial attempt to operationalise a task-based approach to the development of real-world academic oral discussion ability, in which students performed academic oral discussions, then during, or following task participation 'noticed' (Robinson, 1995; Schmidt, 1990) aspects of their performance that could be improved. Two groups operationalised this approach-one in which the post task noticing activities were frequent and structured (Group 2), and one in which the activities were less frequent and less structured (Group 1). This latter group approximated experiential learning through exposure alone while the former group implemented a greater number of interventionist teacher led noticing activities.

We contrasted teaching operationalising this approach with teaching following a more familiar and traditional synthetic EAP syllabus, in which students were first taught academic discussion microskills (agreeing and disagreeing, exemplifying points, turntaking procedures etc), then were encouraged to practice, and then further practice them, largely in isolation from integrative whole task practice (Group 3).

3.2 The Students

The analytic or task-based approach and the synthetic approach to syllabus design were compared over one semester at Aoyama Gakuin University (eight classes delivering instructional treatments, and one class each for pre and post-testing). Three classes of 19, 20 and 21 students, each at an intermediate level of English language ability, participated in the study. The students were English majors in the first term of their freshman year. This was the first of two years in an integrated language skills program that combines 6 hours of weekly instruction in speaking, listening, writing, and reading.

Upon entering the program, the students take a language placement test and are grouped according to three different levels of ability. The curriculum is organized into themes at each of these levels and students undertake a variety of tasks and activities such as writing journals, and essays, reading and reporting on newspaper articles, doing book reports and oral presentations and participating in small group discussions.

3.3 The Treatment

In the analytic or task-based approach , students in small groups of 3 or 4 persons worked on a weekly cycle of task (whole group oral discussion), and then post task activities that included self-reflection on their task performance, and/ or group discussion of comments

they made about their own and each others performance using taped audio and video recordings of their group discussions. Group 2 performed more of these than Group 1. In the initial classes a limited number of pre-task orienting activities were used by both task-based groups to orient students to the features of turntaking, gesture, and language use that they could profitably attend to and comment on throughout the rest of the semester in subsequent post task noticing activities. At the beginning of each class groups of 3 or 4 students sat together and watched other students performing discussions, noted the features of those discussions, and rated the group's performance. Selections from recordings of their own discussions were later transcribed by each set of group members and were used while they looked for examples of successful and unsuccessful phrasal or turn-taking language, or discussion performance, etc. They compared their observations with those of their classmates.

In contrast, students in the class following the synthetic, skills based syllabus (Group 3) learned about different kinds of functional language used in discussions such as soliciting opinions, expressing agreement and disagreement. The appropriate expressions were shown to the students and they rehearsed them on a weekly basis in pairs, applying them to follow up activities, with little opportuinity for whole task discussion practice.

Each week, students in all the two task-based classes were randomly assigned to discussion groups of three or four persons. Pedagogy in the skills based class largely involved individual and pair work. To ensure that both groups used topics of similar interest and difficulty, a discussion text *Impact Issues* was used in each class. About 45 minutes was spent on discussion activities during each week of the 8 week treatment. The issues selected for discussion were chosen according to the themes in the Integrated English Program.

3.4 The Rating Instruments

The pre-test and post-test consisted of videotaped group discussions of 5 minutes in length. The individual students in each discussion were scored by 3 experienced native speaker raters (mean length of ESL experience over ten years) who underwent a training session when they practiced use of the rating instrument (Appendix C). The three ratings, from 1 to 5 on a five point scale, for each of four categories (turntaking, eye-contact and gesture, language use, content) were averaged. Interrater reliability was .76.

3.5 Results

Results of the repeated measures ANOVA (Group x Category x Pre-Post-test) of the rating averages show no significant differences for the factor Group, but significant differences for Category and for Pre-Posttest ($p < .01$). As can be seen in Figure 7, all groups improved from pre to post-test, with greater improvement in the areas of content and language than in eye

contact and gesture. A priori planned comparisons revealed a significant difference on the posttest between Task-based Group 1 and the superior Skills-based Group 3. Task-based Group 2 and the Skills-based Group 3 were equivalent.
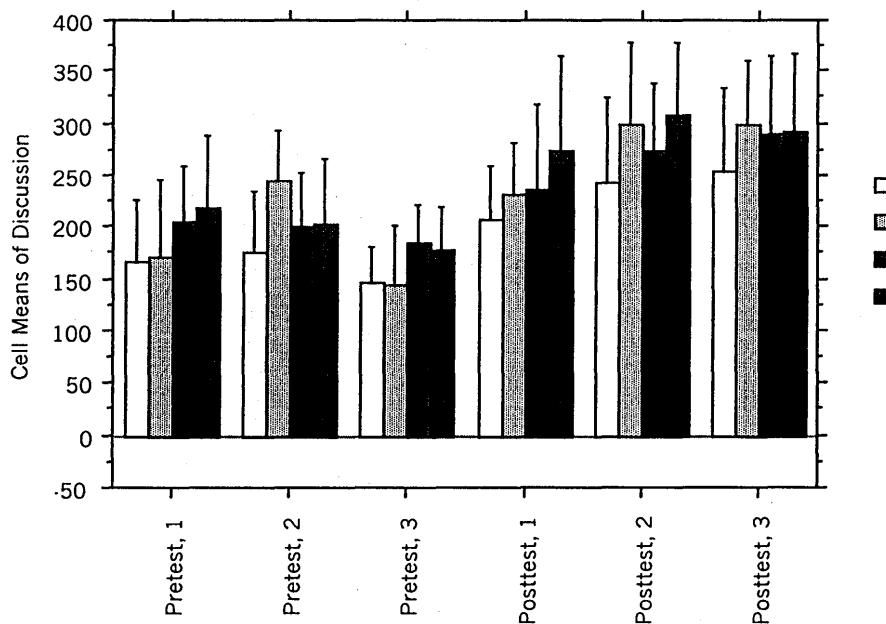

Figure 7. Results of the ANOVA of Pre and Post-test Performance in the Discussion Task Across Groups.

3.6 Conclusions

1. Both structured task-based teaching (Group 2), incorporating focus on form activities, and skills-based teaching (Group 3) are equivalent, with skills-based teaching having advantages over unstructured experiential task-based learning (Group 1). Possibly this is due to transfer of training and expectations from prior language learning experience, since the skills-based approach is most similar to our students' previous English learning experience in Japanese high schools. Longer term studies of the effects of our different kinds of instruction are needed. Nonetheless, the results suggest that structured focus on form, plus extensive task practice is equivalent to carefully targeted and sequenced microskills teaching.

2. Non verbal aspects of discussion abilities—particularly turntaking ability—are the least susceptible to instruction over the short term, in all conditions. It is not clear yet whether

these are best acquired incidentally, over a longer period, compared to verbal aspects which may benefit more from an explicit intentionally directed focus of learner attention, and subsequent rehearsal and memorization. Alternatively, students may have felt more motivated and focussed on verbal aspects at the expense of non verbal aspects (for fuller discussion of these findings see Robinson, Strong, Whittle & Nobe, 1999).

4. SOME AGENDA ITEMS FOR CLASSROOM RESEARCH IN JAPAN

Issues addressed by the studies reported above remain far from resolved, and further research is needed to investigate each of them in Japanese classrooms. These issues include the following:
i) the effects of task complexity on learner perceptions of difficulty, and on measures of learner production,
ii) the effects of sequencing decisions based on differences in complexity on learner perceptions of difficulty,
iii) the effects of different techniques for focus on form in the context of communicative activity, and their relationship to currently available measures of learner variables such as aptitude,
iv) the development of effective techniques for focusing learner attention on pragmatic and discourse level phenomena, important to the development of advanced second language skills such as the ability to take part in academic discussions in an academic context.

Such research would help resolve such questions as: what tasks, and sequencing decisions to select for learners to ensure optimum levels of accuracy, fluency and complexity of production; what learner variables might account for large variance in learning performance; and how to effectively develop advanced level communication skills. An exciting prospect indeed for those wishing to base pedagogic decision making, at least in part, on empirical evidence of learning processes and performance.

REFERENCES

Ackerman, P. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117,* 288-318.

Aline, D. (1999). Noticing output and its effect on subsequent task production. To appear in P. Robinson, M. Sawyer and S, Ross (Eds.), *Second Language Acquisition Research in Japan.* JALT Applied Materials Vol. 4. Tokyo: JALT Publications.

Beglar, D. (1999). Implicational generalization and the acquisition of relative clauses. To appear in P. Robinson, M. Sawyer and S, Ross (Eds.), *Second Language Acquisition Research in Japan.* JALT Applied Materials Vol. 4. Tokyo: JALT Publications.

Carroll, J.B. (1993). *Cognitive Abilities: A Factor Analytic Approach to Research.* Oxford: Oxford University Press.

Day, R., & Yamanaka, J. (1998). *Impact Issues.* Singapore: Longman.

DeGraff, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition, 19,* 249-277.

DeKeyser, R. (1995). Learning second language rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition, 117, 379-410.*

DeKeyser, R. (1998). Automaticity and automatization. To appear in P. Robinson (Ed.), *Cognition and Second Language Instruction.* New York: Cambridge University Press.

Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 114-138). New York: Cambridge University Press.

Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 197-262). New York: Cambridge University Press.

Eckman, F., Bell, L., & Nelson, D. (1988). On the generalization of relative clause instruction in the acquisition of English as a second language. *Applied Linguistics, 9,* 1-20.

Fleishman, E.A. (1978). Relating individual differences to the dimensions of human tasks. *Ergonomics, 21,* 1007-1019.

Hancock, P., Williams, G., & Manning, C. (1995). Influence of task demand characteristics on workload and performance. *The International Journal of Aviation Psychology. Special Issue on Pilot Workload: Contemporary Issues, 5, 1,* 63-86.

Hawkins, R., & Chan, C. (1997). The partial availability of Universal Grammar in second language acquisition: the 'failed functional features hypothesis'. *Second Language Research, 13, 3,* 187-226.

Hulstijn, J. (1989). A cognitive view on interlanguage variability. In M. Eisenstein (Ed.), *The Dynamic Interlanguage* (pp.17-32). New York: Plenum Press.

Keenan, E., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry, 8,* 63-99.

Long, M. H.(1985). A role for instruction in second language acquisition: Task-based language teaching. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and Assessing Second Language Acquisition* (pp. 77-99). Clevedon, Avon: Multilingual Matters.

Long, M. H. (1997). *Focus on Form in Task-based Language Learning.* Paper presented in the McGraw-Hill Teleconference on Approaches to grammar instruction in communicative teaching, California State University, Long Beach, U.S.A.

Long, M. H. (1999). *Task-Based Language Teaching.* To appear from Oxford: Blackwell.

Long, M.H., & Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly, 26,* 55-98.

Long, M.H., & Robinson, P. (1998). Focus on form: Theory, research, practice. In C.

Doughty & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp.15-41). New York: Cambridge University Press.

Muranoi, H. (1999). Focus on form: Implicit and explicit form-focused instruction incorporated into a communicative task. To appear in P. Robinson, M. Sawyer and S, Ross (Eds.), *Second Language Acquisition Research in Japan.* JALT Applied Materials Vol. 4. Tokyo: JALT Publications.

Norris, J., Brown, J.D., Hudson, T & Yoshioka, J. (1998). *Designing Second Language Performance Assessments.* Second Language Teaching and Curriculum Center Technical Report no. 18: University of Hawaii Press.

Nunan, D. (1988). *Syllabus Design.* Oxford: Oxford University Press.

Robinson, P. (1995a). Task complexity and second language narrative discourse. *Language Learning, 45,* 99-140.

Robinson, P. (1995b). Attention, memory and the 'noticing' hypothesis. *Language Learning, 45,* 283-331.

Robinson, P. (1997a). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning, 47,* 45-99.

Robinson, P. (1997b). Giving pushed output a push: The role of task complexity. *Clarion, EUROSLA Newsletter, vol. 3, no 2,* 37-40.

Robinson, P. (1998a). State of the art: SLA theory and second language syllabus design. *The Language Teacher, 22, 4,* 7-14.

Robinson, P. (1998b). Task complexity, cognitive resources and second language syllabus design. To appear in P. Robinson (Ed.), *Cognition and Second Language Instruction.*

New York: Cambridge University Press.

Robinson, P. (1998c). *Aptitude-treatment interactions in the development of oral academic discussion task performance*. Plenary Address presented at the 5th KOTESOL Conference, Kyunghee University, Seoul, South Korea, October, 1998., and to appear in the Proceedings.

Robinson, P. (1999a). Intelligence, memory and reasoning resources during SLA. To appear in C. Doughty and M. Long (Eds.), *Handbook of Second Language Acquisition Research*. Oxford.: Blackwell.

Robinson, P. (1999b). *Task Complexity and Second Language Syllabus Design*. Ms in preparation for Cambridge University Press.

Robinson, P. (1999c). *Individual Differences and Second Language Instruction*. Ms in preparation for Lawrence Erlbaum.

Robinson, P., Strong, G., & Whittle, J., & S. Nobe. (1999). The development of EAP oral discussion ability. To appear in J.Flowerdew and M. Peacock (Eds.), *Handbook of Research in EAP, Vol. 2: The EAP Curriculum*. New York: Cambridge University Press.

Robinson, P., Ting, S., & Urwin, J. (1995). Investigating second language task complexity. *RELC Journal 26, 2,* 62-79.

Robinson, P., & Yamaguchi, Y. (1999). *Tasks, implicit feedback and aptitude*. Manuscript in preparation.

Sasaki, M. (1996). *Second Language Proficiency, Foreign Language Aptitude and Intelligence*. New York: Peter Lang.

Sawyer, M., & Ranta, L. (1998). Aptitude, individual differences and instructional design. To appear in P. Robinson (Ed.), *Cognition and Second Language Instruction*. New York: Cambridge University Press.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics 11,* 129-158.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Skehan, P., & Foster, P. (1998). Cognition and tasks. To appear in P. Robinson (Ed.), *Cognition and Second Language Instruction*. New York: Cambridge University Press.

Snow, R.E. (1989). Aptitude-treatment interactions as a framework for research in individual differences. In P. Ackerman, R. Sternberg & R. Glaser (Eds.), *Learning and Individual Differences,* (pp. 11-34). New York: W.H.Freeman.

Snow, R.E. (1994). Abilities in academic tasks. In R. Sternberg & R. Wagner (Eds.), *Mind in Context: Interactionist Perspectives on Human Intelligence,* (pp. 3-37). New York: Cambridge University Press.

Spilsbury, G., Stankov, L. & Roberts, R. (1990). The effects of a task's difficulty on its correlation with intelligence. *Personality and Individual Differences, 11,* 1069-1077.

Stankov, L. (1987). Competing tasks and attentional resources: Exploring the limits of the primary-secondary paradigm. *Australian Journal of Psychology, 39,* 123-137.

Stanton , W.R. & Keats, J.A. (1986). Intelligence and ordered task complexity. *Australian Journal of Psychology, 38,* 125-131.

Suedfeld, P., & Coren, S. (1992). Cognitive correlates of conceptual complexity. *Personality and Individual Differences, 13,* 1193-1199.

White, R. (1988). *The ELT Curriculum.* Oxford: Blackwell.

Wilkins, D. (1976). *Notional Syllabuses.* Oxford: Oxford University Press.

Name_____

Circle one number on each of the lines below to show how you felt while you were doing this task.

I thought this task was easy.								I thought this task was hard.
1	2	3	4	5	6	7	8	9

I felt relaxed doing this task.								I felt frustrated doing this task.
1	2	3	4	5	6	7	8	9

I didnt do well on this task.								I did well on this task.
1	2	3	4	5	6	7	8	9

This task was not interesting.								This task was interesting.
1	2	3	4	5	6	7	8	9

I dont want to do more tasks like this.								I want to do more tasks like this.
1	2	3	4	5	6	7	8	9

I spoke slowly while I did this task.								I spoke quickly while I did this task.
1	2	3	4	5	6	7	8	9

I made many mistakes while I was speaking.								I made no mistakes while I was speaking.
1	2	3	4	5	6	7	8	9

Appendix A. The Rating Scale Used to Assess Perceptions of Task Difficulty

113

Appendix B. An Example of The Relative Clause Tasks

Two people are starring in a movie called "The Dangerous Game".
One is Kevin Costner. Ask questions like the ones below to find out which actor your partner has chosen to appear with Kevin Costner. Mark the picture with a cross(×), when you have identified it.
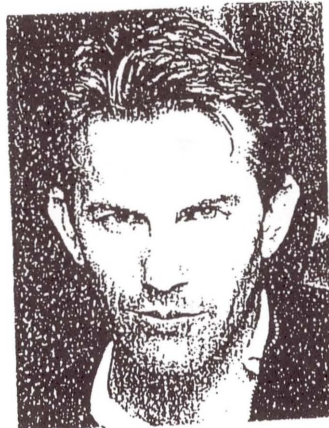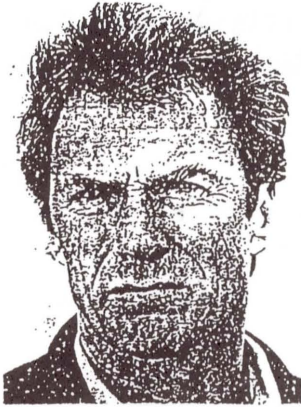
Examples:
  Is he the actor who Kevin Costner is _older_ than?
  Is he the actor who Kevin Costner is <u>more handsome</u> than?
  Is he the actor who Kevin Costner is <u>more popular with young girls</u> than?
  Is he the actor who Kevin Costner has <u>shorter hair</u> than?
Useful expressions:
  younger, taller, shorter, fatter, thinner, stronger, richer, sexier,
  has bigger eyes, runs faster, etc.

Turn taking.

* (1) follows a predictable circular pattern, preceded by lengthy pauses.

* (2) follows a less rigid format, often preceded by lengthy pauses.

* (3) fairly spontaneous and unplanned, hesitations and pauses still occur.

* (4) fairly spontaneous, with few pauses.

* (5) no obvious pattern, and no pausing.

A_____B_____C_____D_____

Eye contact and gesture.

* (1) minimal to no eye contact-no gestures.

* (2) limited eye contact - often directed at one person when speaking-may look down or away if not speaking- gestures are rare.

* (3) eye contact maintained, but not used for turn taking, or emphasizing points- some rhetorical and spontaneous gestures.

* (4) good even distribution of eye contact- follows eye contact signals to participate- gestures accompany agreeing/emphasizing etc.

* (5) even, confident distribution of eye contact-uses appropriate gestures- when listening uses gestures and other cues to take the floor.

A_____B_____C_____D_____

Phrasal language.

* (1) speakers simply state opinions-no phrases for agreement/ disagreement, or emphasis-no clarification requests.

* (2) no variety in the phrases used to agree/disagree and emphasize-- clarification requests are rare.

* (3) varied use of fixed phrases- occasional clarification requests and confirmation checks.

* (4) a greater variety of phrases and speech acts- confirmation checks and clarification requests are common.

* (5) a rich, natural variety of non formulaic phrases- uses comprehension checks and clarification requests.

A_____B_____C_____D_____

Discussion content.

* (1) uninteresting, unengaging content-no supporting details or examples- main points hard to identify.

* (2) main points identifiable-content predictable-few supporting details and examples.

* (3) main points supported by details and examples- imaginative and interesting- listeners occasionally smile and laugh.

* (4) interesting and thoughtful-main ideas and examples are clearly distinguished- often surprises, amuses or otherwise stimulates listeners.

* (5) interesting, engaging content-clearly delivered- a high level of personal response and rationalization - listeners show surprise, amusement and high interest.

A_____B_____C_____D_____