

The Predictive Evaluation of Language Learning Tasks

Zorana Vasiljevic

Faculty of Literature, Bunkyo University

3337 Minami-Ogishima, Koshigaya, Saitama, Japan 343-8511

Tel: 81-48-974-8811 (ext. 2264) E-mail: zorana@koshigaya.bunkyo.ac.jp

Abstract

Teachers are often faced with difficulty in choosing appropriate teaching activities for use in their classroom. In selecting suitable materials for their learners, teachers need to be able to analyze any tasks (i.e., their objectives, procedures and intended outcomes) before they are applied in the classroom. This paper will attempt to outline a systematic procedure for predictive task evaluation. This model should help teachers to identify elements in the task design that are likely to affect the accuracy, fluency and complexity of the students' output *before* the task is implemented in the classroom and thus help them to make decisions regarding task selection and their sequencing.

Keywords: Predictive task evaluation, Task production (fluency, accuracy and complexity)

1. Evaluation in language teaching

Over the last two decades, there has been a growing interest in the purpose and methods of evaluation in language teaching (e.g. Sheldon, 1988; Alderson & Beretta, 1992; Weir & Roberts, 1994; Ellis, 1997,1998). In the literature, the term *evaluation* is used in a number of different ways and based on the writers' purpose, various definitions have been proposed. One of the most 'workable' definitions of evaluation was provided by Richards et al. (1985:98), who described evaluation as 'the systematic gathering of information for purposes of making decisions'. Although this definition may seem broad, it is practical as it can be applied to any component of the language curriculum (needs analysis, objectives, testing, materials, teaching, or the evaluation process itself). Evaluation plays a crucial role in curriculum development as it allows instructors, material designers and administrators to assess the effectiveness and efficiency of a particular language program or any of its components and make informed decisions about how to proceed.

Evaluations can be macro or micro in scale and can be carried out for either accountability or developmental purposes or both of these. In macro evaluation, various administrative and curricular aspects are examined (e.g. materials evaluation, teacher evaluation, learner evaluation), while micro evaluation focuses on the specific aspect of the curriculum or the administration of the program such as evaluation of learning tasks, questioning practices, learners' participation etc. (Ellis, 1998).

The evaluation in language teaching has been primarily concerned with the macro evaluation of programs and projects (Ellis, 1998), and most evaluation studies have been conducted in order to measure the extent to which the objectives of a program have been met, and to identify those aspects that can be improved. As Ellis (1998) observes, this kind of analysis is obviously of interest to teachers as they learn whether or not the goals have been accomplished and whether any changes should be made to the program. However, most teachers are less likely to be concerned with the evaluation of the program as a whole, and more concerned with the extent to which a particular textbook, or a teaching activity is effective in their teaching context.

The evaluation of teaching materials may be done before they are used in the classroom in order to determine whether they suit the needs of the particular group of learners (predictive evaluation), or after the materials have been used in the classroom in order to evaluate their effectiveness and efficiency, and teachers' and learners' attitudes towards them (retrospective evaluation). This paper will introduce a systematic procedure for conducting the predictive evaluation of language teaching tasks

2. Language learning task: definition and componential framework

Put simply, learning tasks are a means for creating the conditions necessary for the acquisition of language. Many definitions of language-learning tasks are found in the literature, but perhaps the most helpful is that provided by Richards, Platt and Weber (1985). They define a language-learning task as:

"...an activity or action which is carried out as a result of processing or understanding language (i.e. as a response). For example, drawing a map while listening to a tape, listening to an instruction and performing a command, may be referred to as tasks. Tasks may or may not involve the production of language. A task usually requires the teacher to specify what will be regarded as successful completion of the task." (Richards et al., 1985: 289)

The definition above clearly highlights the key components of a task: (1) language input, (2) goals (a clearly specified outcome, which determines when the task has been completed) and (3) activities (what learners need to do in order to complete the task successfully).

The interest in task-based language learning has been stimulated by psycholinguistic research, which suggests that learners have their own built-in syllabus, which is often different from the syllabus proposed by instructors (Ellis, 1998). Thus, the sequencing of the linguistic input designed by the instructor may not follow the order of the learner's linguistic intake. Task-based instruction specifies in broad terms what language learners will communicate about and the procedures they will follow, but it gives learners more freedom in terms of the choice of language they use, allowing them to develop their knowledge and skills in accordance with their own interlanguage and order of acquisition.

Since the mid-1990s, there has been a growing interest in the effects that the cognitive demands of different tasks may have on students' performance and the restructuring of their interlanguage. One of the most comprehensive frameworks for the analysis of the cognitive characteristics of language-learning tasks was proposed by Robinson (2001, 2003). Robinson argues that successful task performance depends on the interaction of multiple factors that operate in three different dimensions: *task complexity*, *task difficulty* and *task conditions*. Task complexity refers to the cognitive demands (i.e. the attentional, memory, reasoning and other processing demands) that the structure of the task imposes on the language learner. Task difficulty refers to learner factors that may make a task more or less difficult. This includes affective variables such as motivation, anxiety and confidence and ability variables such as aptitude, proficiency and intelligence. Finally, the successful completion of a task also depends on task conditions or the interactive demands of tasks. Interactional factors include participation variables (e.g. one-way or two-way task) and participant variables (e.g. gender, familiarity, power and solidarity). In short, students' task performance is likely to be influenced by the interaction of multiple factors across the three componential dimensions. Teachers and material writers can manipulate these variables either to allow learners access to an existing L2 knowledge base (a focus on fluency) or to promote form control in learners' interlanguage (a focus on accuracy).

3. Evaluation of language learning tasks

The literature on educational evaluation offers numerous checklists and guidelines for the evaluation of language learning tasks. Ellis (1998) proposed a model that identified five basic steps of task evaluation. These were: (1) a description of the task that involves the analysis of contents and task objectives, (2) planning the evaluation, (3) collecting information, (4) analysis of the information collected and (5) conclusions and recommendations. Each of the steps includes several components or dimensions that need to be considered. For example, Step 2 (planning the evaluation) encompasses seven different dimensions of evaluation: approach, purpose, focus, scope, the evaluators, the timing and the type of information, and each of these dimensions has two or more subcategories. Whilst these kinds of guidelines are comprehensive and offer an excellent theoretical base for task examination, they are too detailed to be used in class preparation on a regular basis. It is highly unlikely that a teacher who needs to make a decision as to whether to adopt, adapt or reject a particular teaching activity will have the time or energy to consult a long checklist with 50 or more criteria to consider. Furthermore, the evaluation frameworks commonly found in the literature fail to make a distinction between predictive and retrospective evaluations. Checklists are typically organized in a way that is believed to reflect the teacher's decision-making process: they start with the general evaluation of the overall 'usefulness' of the material, and end with questions that evaluate the task based on the teacher's actual teaching situation. While records on task performance are an important element in the process of material development, they are clearly a part of retrospective evaluation. Although task evaluation should ideally incorporate both predictive and retrospective elements, teachers in practice often have to make decisions about the pedagogical value of a specific task before they meet the learners and thus before they have any information about their intelligence, motivation or attitudes. Furthermore, some learner factors such as motivation, anxiety and confidence, are likely to be less stable and, to a large extent, reflect learners' perceptions of task difficulties. They may be very hard or impossible to diagnose in advance. This means that prior to the implementation of the task in the classroom the teacher should collect baseline information about the cognitive factors in the task design that may affect learners' performance. In other words, complexity differentials should be the major criteria for proactive pedagogic task sequencing (Robinson, 2003).

4. The predictive evaluation of language learning tasks

Any model that aims at providing 'scientific' guidelines for conducting the evaluation of teaching materials is bound to have limitations. As Sheldon (1998:245) observes, "it is clear that coursebook assessment is fundamentally a subjective, rule-of-thumb activity, and that no neat formula, grid or system will ever provide a definite yardstick."

The same can be said for the evaluation of individual teaching tasks: a priori estimation of task difficulty may be

very hard or even impossible to achieve. There is no a magic formula that would guarantee that teachers would always be able to identify all the characteristics of a task that may affect students' performance. Formalizing the evaluation procedure, however, makes such a process more systematic and potentially more objective. This paper proposes a model for predictive task evaluation based on the analysis of task input, outcomes and cognitive elements in task procedures.

4.1 Task input

Input refers to 'the data that form the point of departure for the task' (Nunan, 1989:53). Input for language-learning tasks can be in a verbal or non-verbal form, or a combination of the two.

4.1.1 Verbal input

The complexity of verbal input depends to a large extent on the authenticity of the material. Authentic texts are likely to contain less frequent vocabulary, more slang and idiomatic expressions, more complex syntactic structures, and, for aural materials, a larger number of incomplete utterances, faster pace and more features of connected speech with less word enunciation and less repetition (Porter & Roberts, 1981). Researchers are still divided with regard to the role that authentic materials should play in the classroom. Supporters of authentic materials (e.g., Brosnan, Brown & Hood, 1984) argue that materials especially written for ELT do not prepare the learners for the aural and written texts they are going to encounter in the real world and that adult learners often do not perceive them as relevant to their needs. Authentic texts, however, expose students to natural language and make the connection between classroom work and real-life tasks more obvious. Some researchers (e.g., Little, Davit & Singleton, 1989) also claim that authentic materials bring learners closer to target culture, thus increasing their motivation.

However, while it is indisputable that the goal of language teaching is to enable learners to engage in real-world texts, authentic materials may not be suitable for learners at all levels of proficiency. Learners' comprehension is known to decrease with an increase in the syntactic and lexical complexity of the input (Nagy, 1988; Nation & Coady, 1988; Qian, 2002). The results of some studies (e.g., Freeman & Holden, 1986; Morrison, 1989) suggest that authentic materials may decrease learners' motivation as they tend to be too difficult. In a study conducted by Peacock (1997), learners' on-task behavior and observed motivation improved when authentic materials were used, but learners also reported that authentic materials were significantly less interesting than the artificial ones. One proposed 'golden-mean' solution to this was the simplification of the genuine-texts. However, this approach produced limited effects. In reading comprehension, familiarity with the content was found to play a much more significant role than the linguistic (syntactic or lexical) simplification of the material (Blau, 1982; Parker & Chaudron, 1987; Yano, Kong & Ross, 1994). Furthermore, the simplification and alteration of the materials (e.g., the limiting of grammatical structures, the control of vocabulary etc.) risk making the input more difficult as some of the meaning clues may be removed from the text (Bronsan et al., 1984).

What are the implications of these findings for the language teacher? One possible criterion for the incorporation of authentic materials within the classroom could be learners' proficiency. Some basic guidelines in regards to this are provided in Figure 1 below.

Insert Figure 1

It seems reasonable to assume that most authentic texts will not be suitable for beginners. Low-intermediate students, however, may benefit from simplified authentic texts. From intermediate level upwards, teachers should try to expose students to real-life materials as much as possible. Genuine texts will provide students with samples of natural language, making the connection between the classroom instruction and real life more obvious. This is likely to increase the motivation of adult learners who often learn language for instrumental reasons and expect teaching materials and activities to reflect real-life experiences. Although teachers may want to introduce students to the target culture as soon as possible, exposure to authentic texts should begin with the materials for which learners are likely to have content schemata. Materials for which learners have no content or subject knowledge may be difficult to understand and consequently have a negative impact on learners' motivation, impeding the language-learning process.

4.1.2 Non-verbal input

In both authentic texts and materials that have been specially written for ELT, non-linguistic clues (e.g., pictures, illustrations, graphs and symbols) are believed to facilitate learners in the comprehension of reading or listening materials. Visual input is also frequently used independently of verbal input in order to contextualize the target language or to stimulate language practice. Assumptions about the beneficial effects of visual aids, however, have rarely been tested and little is known about how learners from different cultures perceive these materials (Hewings, 1991). Pictures and illustrations are often culture-bound, and thus they may reflect values, attitudes and conventions

that learners may not be familiar with. As a result, they may be misinterpreted by students from different cultural backgrounds. For example, in Hewings (1991) study of Vietnamese students in Britain, the plate in the symbol below was interpreted as a table, a door, a swimming pool and a place for dancing.

Insert Figure 2

The example above comes from a textbook, but the same problem may occur with authentic materials. In 2004, I used an article from *The Economist* in my advanced reading class. As a warm-up, I wanted the students to predict the content based on the title (*Your Cheating Phone*), the subtitle (*Do mobile phones make it easier or more difficult to deceive people about your location, activities and intentions?*) and an illustration below.

Insert Figure 3

Although the students were advanced, only two out of twelve learners were able to connect the long nose of the character in picture with the vocabulary in the title and sub-title (cheat, deceive) and the tale of Pinocchio. In this case, although the illustration did not cause miscomprehension, it did not seem to facilitate comprehension either.

These two examples suggest that teachers need to exercise caution in the selection of pictures and illustrations as they may be misinterpreted by the learners and thus fail to provide the intended context for the verbal message. The two criteria that instructors may want to apply in the evaluation of non-verbal input may, therefore, be (1) the extent to which visual input facilitates comprehension of the verbal input (there are many visually pleasing materials that may fail to serve this purpose) and (2) the extent to which learners need some cultural / background knowledge in order to interpret it correctly.

4.2 Task Outcomes

With predictive evaluation, the outcome of a task can be examined at two different levels: (1) the surface level, which describes what it is that learners will have achieved on the completion of the task (e.g., drawing a map, filling in a chart) and (2) the deep level, which describes what learning is expected to take place upon task completion. Ellis (1998) refers to these two criteria as the *student's target* and *teaching objective*.

Surface analysis should present no problems for language teachers. If the instructions are well written, 'the target' should be obvious to both instructors and students. If the teacher cannot easily identify what the target of the task is and if that target is not obvious to the learners, engaging in the task is likely to result in frustration and disappointment rather than language learning. An examination of students' targets should help instructors in anticipating possible problems in understanding task directions and identifying any difficult vocabulary or syntactic structures that learners may encounter during task performance.

Identifying teaching objectives, however, requires more experience and knowledge of the principles of language acquisition. These objectives may be communicative such as exchanging information, and sharing opinions and feelings, or socio-cultural focusing on increasing students' understanding of the target language speech community. They may also be purely linguistic, aimed at drawing learners' attention to some specific feature of the L2 system, or they may be metalinguistic, looking to increase students' awareness of the principles of language learning and helping them manage their own learning process (Clark, 1987).

Deep level analysis (i.e., analysis of the teaching objectives) is a very important, but frequently overlooked stage in predictive task evaluation. Less experienced teachers will often examine the task focusing exclusively on student targets. However, deep level analysis is necessary because teaching objectives have clear implications for the teacher's role during task performance. Although it is very difficult to measure how much linguistic knowledge has been gained from completion of an individual task, a failure to recognize teaching objectives may lead to missed teaching and, consequently, learning opportunities. Teaching objectives have implications for the amount and type of teacher talk as well as the type of feedback learners will get on their performance. This of course, does not mean that all teacher-student interaction is planned. Spontaneous talk is a natural and often very helpful aspect of classroom language-learning. The identification of teaching objectives, however, should help instructors to avoid excessive talk and decide what kind of input, advice or feedback may enhance learners' task performance and at which point in the lesson that information should be provided.

4.3 Procedures

The third component of predictive task analysis is the task procedures. According to Ellis (1998:227) task procedures are 'the activities that the learners are to perform in order to accomplish the task.' They are a crucial element of task design as they have direct implications for task complexity. Cognitively complex tasks have been found to lead to more accurate although less fluent language production (Robinson, 1995; Iwashita, McNamara & Elder, 2001). More accurate performance under the conditions presumed to be more difficult has been attributed to a

lack of contextualized support in more complex tasks. According to Robinson (2001) cognitively simple tasks often allow learners to 'fill in' much of the linguistically uncoded information from the context. More complex, tasks, however, force information givers to direct greater attentional resources to the syntactic preparation of production units. Robinson (2003) argues that the high cognitive demands of the task 'stretch' learners' interlanguage, leading to a more elaborate processing of input, better identification of problematic forms in the output, and as a result greater uptake and longer retention of input.

Two dimensions along which task procedures could be evaluated are: (1) cognitive load and (2) availability of prior knowledge. The cognitive load of the task refers to the variables such as the number of activities that learners need to do in order to complete the task, immediacy of the input and reasoning demands of the task. Tasks which require learners to perform multiple activities (e.g., planning a route and then giving directions from point A to point B on a map to a partner) result in a less fluent but more lexically complex output than single-task condition (Robinson, 1995; 2001). Procedures with greater immediacy ('here and now' as opposed to 'there and then' are likely to result in more fluent less accurate production (Robinson, 1995, 2003; Iwashita et.al., 2001). For example, in Robinson's (1995) study, more accurate and lexically more complex production was observed when learners were describing pictures from their memory ('there and then' condition) than when they were looking at the materials ('here and now' condition). Tasks that require reasoning in addition to simple information transmission make greater cognitive demands, directing attentional resources to the features of language code that can help meet these demands (e.g., the use of logical connectors such as if...then, therefore, because) (Robinson, 1995, 2001, 2003). The effect of task complexity on learners' performance is illustrated in Figure 4 below.

Insert Figure 4

Accuracy, fluency and lexical complexity of output were also found to be influenced by the extent to which a task allows learners to draw on prior knowledge. More complex tasks where prior knowledge is not available (e.g., explaining a route on the map from point A to point B in an unfamiliar area) were found to result in more lexical variety and more interaction between the learners, but less fluent production (Robinson, 2001).

One of the dimensions of task complexity discussed in Robinson's (2001) framework is planning time. Although teachers may intuitively feel that more time spent planning is likely to result in more accurate and more fluent production, studies that examined how this dimension of task design may affect learners' performance produced mixed results. In some experiments, greater accuracy of production was observed under the planned conditions (e.g., Ting, 1996; Ortega, 1999), whilst in other studies, task pre-planning was not found to have a significant effect on either accuracy or fluency of the output (e.g. Iwashita et al., 2001). More research is needed in order to determine what effect planning time may have on task performance, and until more conclusive evidence is available, it may be difficult for teachers to employ this variable in the predictive task evaluation.

4.4 Summary and Conclusions

A systematic predictive task evaluation should enable teachers to anticipate how specific features of task design may affect learners' performance and thus allow them to make more informed choices about suitability of particular activities for different learning situations. The proposed model suggests evaluation along three dimensions of task design: input, outcome and the procedures. (A template for predictive task evaluation is available in the Appendix.) A verbal task input should be examined in terms of its authenticity, whilst for non-verbal input, possible cultural bias should be taken into consideration. Task outcomes should be examined for their clarity and student targets should be examined both at the surface level and for their expected learning outcomes at a deeper level. An analysis of the underlying teaching objectives should help teachers define their roles and improve classroom interaction. Making evaluation procedures explicit raises teachers' awareness of any factors in the task design that may facilitate or possibly impede task performance, and allows them to make the necessary adjustments in order to optimize classroom practice. This makes predictive task evaluation an important element in teacher development. It forces teachers to go beyond impressionistic assessments by requiring them to determine exactly what the objectives of the task are, and what the learners (and the teacher) will need to do in order to ensure that those objectives are met. Furthermore, conducting a predictive evaluation in a systematic manner should make it easier for instructors to interpret the results of any retrospective evaluation that may follow. The formalization of evaluation procedures directs instructors' attention to the strengths and possible weaknesses in their evaluation process and helps them to identify the ways in which the predictive instruments could be improved for the future use.

An examination of the cognitive complexity of the tasks also allows material writers and instructors to make decisions with regards to the sequencing of the tasks. If the focus is on accuracy, tasks should be sequenced from simple to complex, while if fluency is the priority, the reverse sequence may be more beneficial. Task complexity increases with the number of activities learners are expected to perform in order to complete the task. Procedures

where learners have to rely on their memory, use reasoning or those with little contextual support are more cognitively demanding and are thus likely to lead to less fluent but more accurate output with greater lexical complexity and increased learners' interaction. A gradual increase in the cognitive demands of a task was found to lead to greater functional differentiation of learner language use, more attention to output and a deeper processing of input which is believed to lead to the faster development of interlanguage (Robinson, 2003).

The systematic evaluation of individual teaching tasks also provides a good basis for the evaluation of sets of teaching materials. While collecting and analyzing information for a whole textbook would be a daunting task for most teachers, a series of planned, consistent evaluations of what Gibbons (1980:44) refers to as 'learning units' (sets of tasks felt by the designer to be necessary for the teaching of an item on a syllabus) should give teachers a clear picture about the relevance and appropriateness of materials for the target group of students.

Finally, an analysis of the cognitive characteristics of language learning tasks is also important in testing contexts. If task characteristics affect learners' performance then these characteristics must be taken into consideration at both a test design stage and during the interpretations of students' results.

There is no doubt that many other factors may influence task performance. Indeed, there are many variables outside the task design that may affect the outcome of a task in a real classroom. However, as Rea Dickens (1994) points out, evaluation is concerned with immediate practical use rather than ultimate use. While the proposed checklist may not be exhaustive, it gives teachers a workable model to use in lesson preparation and the selection and design of materials. It is hoped that this model will assist teachers in identifying any potential mismatches between their objectives and the actual nature of the materials they are planning to use, and help them to make informed decisions about whether to adopt, adapt, supplement or reject specific teaching tasks as well as how to sequence the selected ones.

References

- Alderson, J.C. and A. Beretta (1992). *Evaluating Second Language Education*. Cambridge: Cambridge University Press.
- Arnold, J. and Scott, R. (1986). *Focus: Coursebook One*. Sevenoaks: Hodder and Stoughton.
- Blau, E. K. (1982). The effect of syntax on readability for ESL students in Puerto Rico. *TESOL Quarterly*, 16: 517-28.
- Brosnan, D., K. Brown, and S. Hood (1984). *Reading in Context*. Adelaide: National Curriculum Resource Center.
- Clark, J. (1987). *Curriculum Renewal in School Foreign Language Learning*. Oxford: Oxford University Press.
- Ellis, R. (1997). The empirical evaluation of language teaching materials. *ELT Journal*, 51(1): 36-42.
- Ellis, R. (1998). The evaluation of communicative tasks. In B. Tomlinson (Ed.): *Materials Development in Language Teaching* (pp. 217-238). Cambridge: Cambridge University Press.
- Freeman, D. and Holden, S. (1986). Authentic listening materials. In S. Holden (Ed.) *Techniques of Teaching* (pp. 67-69). London: Modern English Publications.
- Gibbons, J. P. (1980). A shape for language learning units. *The English Bulletin*, 7 (3): 44-47.
- Hewings, M. (1991). The interpretation of illustrations in ELT materials. *ELT Journal*, 45 (3): 237-244.
- Iwashita, N., T. McNamara and C. Elder (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language Learning* 51 (3): 401-436.
- Little, D., S. Devitt and D. Singleton (1989). *Learning Foreign Languages from Authentic Texts: Theory and Practice*. Dublin: Authentik.
- Morrison, B. (1989). Using news broadcasts for authentic listening comprehension. *ELT Journal* 43 (1): 14-18.
- Nagy, W. (1988). *Teaching Vocabulary to Improve Reading Comprehension*. Urbana, IL: National Council of Teachers of English.
- Nation, I.S. P. and J. Coady, (1988). Vocabulary and reading. In Carter and McCarthy (Eds.): *Vocabulary and Language Teaching* (pp. 97-110). London: Longman.
- Nunan, D. (1989). *Designing Tasks for Communicative Classroom*. Cambridge: Cambridge University Press.
- Ortega, L. (1999). Planning and focus on form in L2 oral performances. *Studies in Second Language Acquisition*, 21:109-148.
- Parker, K., & Chaudron, C. (1987). The effects of linguistic simplification and elaborative modifications on L2 comprehension. *University of Hawaii Working Papers in ESL*, 6:107-133.
- Peacock, M. (1997). The effect of authentic materials on the motivation of EFL learners. *ELT Journal*, 5 (12):



Figure 2. (Arnold & Scott, 1988:82, in Hewings, 1991:241)



Figure 3. (Your cheating phone. *The Economist*, 2004:15)

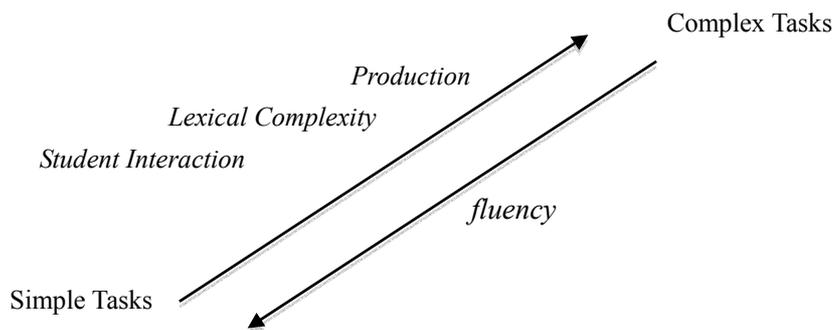


Figure 4. The effects of task complexity on learners' performance.