

DESIGNING COMPLEX LSP TASKS FOR ASSESSMENT PURPOSES

Christian Krekeler

Konstanz University of Applied Sciences, Germany
Phone: +49 (0) 7531 206-395, E-Mail: krekeler@htwg-konstanz.de

Abstract. *When integrating task-based learning into LSP courses, assessment procedures have to reflect the communicative goals of such tasks. However, performances on authentic and complex tasks are often difficult to score, because they involve not only specific language ability but also field specific knowledge. Moreover, the emphasis of tasks on meaning suggests that the communicative outcome should be regarded as the main criterion of success. This paper focuses on task design and scoring. It is argued that tasks have to be adapted to assessment purposes. However, to avoid a misalignment between teaching/learning and assessment, they should comprise the main features of LSP tasks in the sense of task-based language learning.*

Key words: *LSP assessment, integrated tasks, task-based language assessment, performance assessment*

1. INTRODUCTION AND BACKGROUND

The use of complex tasks for assessment purposes leads teachers to consider affective criteria such as effort rather than specific properties of the performance because students often invest great effort in the task and teachers want to reward them accordingly. This practice is problematic, however, when the main aim is to assess academic achievement. It is also genuinely difficult to score the students' performance on complex tasks. In this paper, I try to disentangle some of the conundrums and also suggest a possible approach to the use of complex tasks for assessment purposes. This article is based on a plenary talk I gave at the 'Languages for Specific Purposes in Higher Education' conference which was held at Cambridge University in 2016. The idea for this talk emerged during the LSP courses I taught in a foundation year programme. In this personal account, I reflect on the challenges of scoring the complex tasks used in this course.

The course 'German for Business' is part of a foundation year programme at our university. In the foundation year programme, international students prepare for academic study at a German university. Students who wish to enrol in a Business degree take part in German classes and study subjects such as mathematics, business studies, accounting, law, information technology and English. In addition, students also attend an LSP course in German for Business. In this course, business-related topics are covered in conjunction

with the language elements required by students to understand the topics and to communicate within the field of business studies.

A task-based approach was adopted for the LSP course. Every unit of the course consisted of three elements: 1) a topic, 2) a language element and 3) a task. Relevant topics from the field of business and economics were firstly identified. Language elements were then chosen which would aid students in understanding and dealing with the topic. Finally, tasks were selected in accordance with the principles of ‘task-based language teaching’ (TBLT). Tasks, topics and language elements were aligned to support the desired learning outcomes.

What are the characteristics of tasks used in the sense of TBLT? Such tasks may engage more than one skill, and are cognitively demanding. They involve a real-world relationship, and students are meant to achieve a communicative goal (Ellis 2009; Skehan 2003; van den Branden 2006). The distinction between ‘focused’ and ‘unfocused’ task is also relevant in this context: Tasks which give students the opportunity to communicate using words and structures they acquired during the course can be described as ‘focused’ rather than ‘unfocused’ (Ellis 2009, 223).

As an example, I refer to a task from the “German for Business” course which was part of a unit on income development. In this unit, different forms of incomes were introduced and the difference between gross and net incomes was discussed. The income development of university graduates was taken as an example. Language elements included reading skills, comparisons, the use of numbers and terminology. The ‘target task’ (D. Willis and J. R. Willis 2011) of the unit on ‘Incomes’ is shown in example 1.

Example 1: Video task

“You are working for the ‘Tagesschau’, the flagship news programme of the German broadcaster ARD. Write a short news item to inform the viewers about the development of incomes. Your item should include a short interview with an expert on incomes. Record the presentation of your news item. You have one week to complete the task with your group.”

I refer to this task as the ‘*video task*’. It was set to fulfil most of the criteria of TBLT: the task was complex as it involved both writing and reading; it was cognitively demanding in that the students had to research statistics on the development of incomes and to interpret them; the communicative goal was to inform (fictitious) viewers of the news programme about the topic. The task was only semi-authentic: Reference was made to an existing news programme, and the development of incomes is a topic which is often reported on. However, the students’ news item would not be broadcast on television or streamed online but simply shown to the group and their teacher. It was expected that students would need to use language elements which had previously been covered in class. As such, the video task can be characterised as a focused task, as students were required to apply the language elements they had learned throughout the unit.

Overall, I felt that the video task was useful for the aims of the LSP course. However, the usefulness of the video task for assessment purposes was limited.

2. CHALLENGES OF ASSESSING COMPLEX LSP TASKS

When complex LSP tasks are used for assessment purposes, a number of difficulties arise. This section contains an overview of such difficulties.

Criteria and standards: Applying criteria and standards is difficult because performances on complex LSP tasks can be very diverse. In the video task, for example, the videos and, more importantly, the speech samples of individual students varied widely: some groups compared the incomes of two countries, whereas other groups focused on the development of incomes in one country or one sector. Video length and the duration of the parts each student had in the videos also varied widely – some students had taken on the role of technician and did not feature in the videos at all. The amount of time and effort invested in the production of the videos also differed from group to group. While some groups had met several times, thoroughly researched the topic and prepared a lengthy video for which they had created a studio atmosphere with professional graphics, others had only invested a minimal amount of time. As a result of these differing circumstances, it is difficult to determine which criteria (i. e. properties of the performance) contribute to the successful achievement of the communicative goal and which standards (i. e. level of attainments) should be applied. In the absence of transparent criteria and standards, judgments often focus on properties of the performance which do not relate to academic achievement, such as effort. There is also the risk of focusing solely on surface errors even though they may contribute very little to the quality of the performance and may not be linked to the learning objectives.

Task difficulty and task complexity: Task difficulty refers to learner perceptions of the task, task complexity to the cognitive complexity required to fulfil the task (Robinson 2001, 2011). It is difficult to determine in advance whether the learners perceive the task to be at the right level of difficulty and whether task complexity is at the right level for the learners (Tavakoli and Parvaneh 2009). The main reason for this is that complex LSP tasks are multi-faceted. The video task, for example, involves a research aspect (i. e. finding a topic and suitable sources), writing (writing the news item), reading, and the recording of the video.

Sources: Complex LSP tasks often involve a research element and the use of sources. The video task was intended to be cognitively demanding, in that students had to find, present and interpret suitable statistics. Students' use of sources varied widely and it was unclear which information and which language elements had been taken from other sources. Without additional information, it was difficult to know whether the use of sources was appropriate and to determine the exact contribution of the students and the extent of 'textual borrowing' involved (Weigle and Parker 2012).

Background knowledge: The role of background knowledge is a fundamental issue in LSP assessment. I follow the view of Douglas who regards background knowledge as an integral part of specific language ability. In LSP contexts, communicative performance is regarded as the result of the interaction between background knowledge with language knowledge (Douglas 2000). Thus, background knowledge will affect communicative performance. For the video task, the issue was that students were left on their own to choose the particular topic of their assessment. Some chose statistics which they did not understand and were unable to interpret in a meaningful way. This had a negative impact on the quality of the language.

Group work: The video task was set as a group activity. While there are many advantages to group work activities, and although the literature offers a range of strategies to reduce problems with scoring performances of groups, assessing tasks of this

nature remains a challenge (Cohen et al. 2014; Davies 2009). Setting the video task as a group work activity was in line with the TBLT approach. Within TBLT, authentic or at least semi-authentic tasks are used; preparing and presenting a news item is, after all, a collaborative task. But the speech samples of each of the group members differed widely: one student only greeted viewers at the beginning and said good-bye at the end. Another student, who served as the group's technician, did not feature in the video at all. The contribution of each member of the group to the making of the video (writing the script) was unclear. Thus, scoring individual performances was difficult.

Before discussing the use of complex LSP tasks for assessment purposes, I will reflect on the role of assessment within the context of LSP courses based on TBLT.

3. ALIGNMENT OF TEACHING/LEARNING AND ASSESSMENT

A fundamental requirement for LSP courses with a TBLT approach is the alignment of teaching/learning and assessment. Such an alignment is missing when assessment formats are different from formats used for teaching. For example, assessments with isolated items may be easier to score, but they do not reflect the approach of the course. Assessments communicate to learners what is valued and what the 'real' learning outcomes of the course are (Brown 2002; Chalhoub-Deville 2008; Norris 2016). For LSP courses with a TBLT approach, complex LSP tasks should therefore be used for the assessment.

The alignment of teaching/learning activities, assessment tasks and learning outcomes is widely discussed in the educational literature and the literature on teaching and learning in higher education (Biggs and Tang 2011; Boud and Falchikov 2007). It is argued that teaching/learning activities and assessment procedures should contribute to the same learning outcomes. For this context, Biggs and Tang coined the term 'constructive alignment' which highlights (1) the (constructivist) learning through activities and (2) the alignment of assessment tasks and teaching/learning with the learning objectives (2011).

For LSP courses, this alignment means that the task-based approach, and each of its elements, is mirrored in the assessment. This approach to assessment is also termed task-based language assessment (TBLA). TBLA can be defined as a direct performance assessment which uses authentic, complex and cognitively demanding tasks. They often involve integrated tasks (Plakans 2012). Such 'holistic' assessments are scored with regard to the desired communicative outcome of the task (Ellis 2003; Wigglesworth 2008). The validity argument of TBLA has been discussed in the literature. Content validity is usually regarded as the main criterion but Fulcher argues (in the context of EAP) that the focus should remain on construct validity (1999). At the same time, choosing tasks and communicative goals which are representative of the communicative needs of the test takers and the domain, and the alignment of assessment formats to teaching activities and learning outcomes are useful starting points for test design.

4. TASK DESIGN: ADAPTING COMPLEX LSP-TASKS TO ASSESSMENT PURPOSES

The purpose of the assessment was to measure the achievement of the students and to provide a basis for feedback. Consequently, I revised the video task.

Example 2: Radio task

“You are working for ‘Deutschlandfunk’, a public radio station in Germany. Write a short news item to inform the listeners about the development of incomes, consisting of 1) a leader, 2) examples/sources and 3) an interpretation. Make a recording of your news item. Use the information in the table below on the development of incomes for your news item. You have 90 minutes to prepare the text and the recording.

Scoring will be based on the following criteria:

1. *Intelligibility: Was the news item intelligible? Did word stress hinder intelligibility of the news item? (8 points)*
2. *Content and structure: Was the content of the news item convincing? Was the content consistent throughout the three parts of the news item (leader, examples, interpretation)? (6 points)*
3. *Use of figures: Were the figures presented in a listener friendly way? Were the figures used in a meaningful way for the news item? (4 points)*
4. *Comparisons: Did the news item include a comparison? Was the language used for the comparison correct? Was the comparison meaningful for the news item on incomes? (4 points)*
5. *Other aspects: Were there any other aspects of the performance which improved or hindered the news item? (2 points)”*

I refer to the revised task as the ‘radio task’.

Criteria and standards: The radio task included details on the properties of the desired performance and their contribution to the successful achievement of the goal. It therefore supported criteria-based scoring procedures, i. e. scoring based on properties of the students’ performances rather than on a comparison with the performance of other students in the course (Davidson 2012; Sadler 2009). Students were informed that their news item should include specific elements (comparisons, statistics and figures as well as word stress). These language elements had previously been taught in class which indicates the test function (achievement). Students were able to use the information on the criteria which formed the basis for scoring the performances. However, other than an outline of the communicative goal, the radio task did not include information as to the required standard. It would not have been practicable, for example, to include a description of a model performance.

Task difficulty and task complexity: For the revised task, complexity was reduced. The task was altered from a filmed news programme to a voice recording. The students did not have to deal with visuals, facial expressions etc. The use of sources was limited; students were asked to use a single statistic which was provided in the task instructions. Task complexity and consequently the variability of performances were further reduced by changing the task to a supervised and timed format. However, the revised task still mirrored the TBLT approach and was in line with the intended learning outcomes. Despite the provision of the source and the limitation to reading rather than presenting, the radio task was cognitively demanding. It employed more than one skill (writing and reading) and included a complex content element. Thus, the task retained the characteristics of TBLT or TBLA.

Background knowledge and use of sources: For the completion of their task, students were able to use the information provided in the task instructions. The use of sources was

limited to the statistical information supplied in the task. Therefore, the role of background knowledge was limited to an understanding of these specific figures. The provision of a source ensured that students were steered in the right direction and that ‘off-task scripts’ were largely avoided (Weigle 2009, 132). It reduced the complexity of the assessment and enabled a description of the standard of performance desired.

Group work: In the video task, the roles of the students were not fixed and no other strategies were implemented which could have reduced the difficulties of scoring group work. Attention should be drawn to the fact here that the revised radio task was set as an individual task, which therefore circumvented the issue. However, the task was probably less authentic.

In summary, the radio task facilitated scoring procedures while retaining the main features of the TBLT approach. The task was less complex but it was aligned to the TBLT approach of the LSP course.

5. SCORING COMPLEX TASKS

In the final section of this paper, I will briefly consider scoring procedures for complex LSP tasks. In the context of the ‘German for Business’ course, as in many other LSP courses, the main purpose was to measure achievement of the students, not general proficiency, and to give feedback to the students. The assessment task was used only once and for this group only. It was set and subjectively rated by one person, the teacher of the course. There was no performance data or measurement data which could have been used for the development of a scoring procedure because the assessment had not been trialled. As a consequence, scoring criteria were developed *a priori* which has also been described as an ‘armchair approach’ (Fulcher 2017, 383).

Initially, I had to decide whether ‘generic scales’ (Alderson and Banerjee 2002, 92) could be used to score the radio task or whether specific scales had to be designed. For example, are scale descriptors from the CEFR useful for scoring complex LSP tasks? A typical grade description from the CEFR is: “*Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely [...].*” (CEFR) Could this context neutral descriptor be applied to the radio task in a meaningful way? For the radio task, students may have had to summarise information from a written source in a coherent presentation but they did not have to demonstrate listening skills or spontaneous language use. It would therefore be inappropriate to draw inferences from their performance on the radio task with regard to listening skills or spontaneous language use. This example illustrates that criteria for complex LSP assessments need to be specific.

In the absence of measurement or performance data, the competencies and skills required to successfully complete the LSP task need to be considered. The communicative goal, and therefore the ‘primary trait’ (Weigle 2009, 110) of the assessment task, was the presentation of a news item on incomes. Which skills do students require to achieve this particular communicative goal? With regard to the radio task, I identified reading competence and intelligibility as key components. As a consequence, I selected intelligibility as a central scoring criterion and also included it in the task instructions. The focus on intelligibility during the assessment was also in line with the learning objectives of the unit. In this unit, word-

stress, an important contributing factor for intelligibility (Derwing and Munro 2009), was taught and practised.

Similarly, I had to decide whether to use ‘holistic scales’, which give a single score to the performance as a whole, or ‘analytic scales’, which include scores on a number of criteria (Harsch and Martin 2013; Weigle 2009, 110–21). On the one hand, the use of a holistic scale would be in line with the use of authentic LSP tasks for assessment in that the single score would rightly focus on whether the communicative goal had been achieved or not. For feedback purposes in particular, however, it was more appropriate to consider the features in the performances which contributed to the achievement of the goal. Moreover, while holistic scoring is less time consuming, it is also considered less reliable than analytic scoring (Weigle 2009, 114).

For the radio task, I identified four areas as important features of task performance: 1) intelligibility, 2) content and structure, 3) the use of numbers and 4) the use of comparisons. Finally, an open criterion was included, which reflected that the rating process had been designed without information on performances. The criteria, formed as questions, were also included in the instructions for the students (see task instructions for the radio task in Example 1).

The criteria were used for scoring the performances and they served as a framework for the provision of feedback. Points were awarded for each of the criteria. The various aspects of the performance were weighted differently to account for their presumed level of contribution to the achievement of the communicative goal. While the scoring levels for each criterion were not described in more detail, the pass mark was set at 50 percent (12 out of 24 points). Thus, for each criterion, the performance of a minimally competent student would receive half of the maximum points. Scores were combined for a total score, and students were also informed about the separate scores to support the provision of feedback.

From a measurement perspective, the use of intuitively designed scales is problematic (Fulcher 2017; Fulcher, Davidson, and Kemp 2011; Knoch 2009). However, given that existing proficiency scales are not useful in the context of TBLA and in the absence of measurement or performance data, criteria had to be developed relying on intuition, experience and, as Westhoff argues, also knowledge about SLA (2009). The scoring process of the radio task was transparent. It was also possible to relate the scores to the students’ actual ability and the achievement of concrete learning goals.

6. CONCLUSION

In this personal account of the use of complex LSP tasks for assessment purposes, I argued that assessment should be aligned with learning/teaching and that a task-based approach adopted in an LSP course should be mirrored in the assessment procedures. Obstacles were highlighted that are typically encountered when using such complex LSP tasks for assessment purposes and possible solutions were outlined. Difficulties which arise from using complex LSP tasks for assessment are often a result of the wide variations in task performance. This occurs when LSP tasks are set as a group work activity to be conducted outside the classroom, and when the use of sources is left to the students, who may or may not have sufficient background knowledge to use the information for their task. As a result, task difficulty is unclear as are the criteria for a successful task performance. It was concluded that even complex LSP assessment tasks

need to limit task complexity and indicate criteria for a successful task performance. Context-specific criteria are to be preferred over generic scales, even if they have to be developed intuitively without measurement or performance data. It was suggested that the criteria be included in the task instructions so that learners know what their performance will be assessed on.

REFERENCES

- Alderson, J. Charles, and Jayanti Banerjee. 2002. "Language testing and assessment (Part 2)." *Language Teaching* 35 (2): 79-113.
- Biggs, John Burville, and Catherine So-kum Tang. 2011. *Teaching for Quality Learning at University: What the Student Does*. 4th ed. Maidenhead: McGraw-Hill/Society for Research into Higher Education/Open University Press.
- Boud, David, and Nancy Falchikov. 2007. "Aligning assessment with long-term learning." *Assessment & Evaluation in Higher Education* 31 (4): 399-413.
- Brown, James Dean. 2002. *An investigation of second language task-based performance assessments*. Hawaii: Second Language Teaching & Curriculum Center, University of Hawaii at Mānoa.
- Chalhoub-Deville, Micheline. 2008. "Task-based assessments: Characteristics and validity evidence." In *Researching pedagogic tasks: Second language learning, teaching and testing*, edited by Martin Bygate, Peter Skehan, and Merrill Swain, 210-28. Harlow: Longman.
- Cohen, Elizabeth G., John I. Goodlad, Linda Darling-Hammond, and Rachel A. Lotan. 2014. *Designing Groupwork: Strategies for the Heterogeneous Classroom*. 3rd ed. New York: Teachers College Press.
- Davidson, Fred. 2012. "Test Specifications and Criterion Referenced Assessment." In Davidson and Fulcher 2012, 197-207.
- Davidson, Fred, and Glenn Fulcher, eds. 2012. *The Routledge Handbook of Language Testing*. Routledge handbooks in applied linguistics. London: Routledge.
- Davies, W. Martin. 2009. "Groupwork as a form of assessment: Common problems and recommended solutions." *Higher Education* 58 (4): 563-84.
- Derwing, Tracey M., and Murray J. Munro. 2009. "Putting accent in its place: Rethinking obstacles to communication." *Language Teaching* 42 (4): 476-90.
- Douglas, Dan. 2000. *Assessing Language for Specific Purposes*. Cambridge: Cambridge University Press.
- Ellis, Rod. 2003. *Task-based language learning and teaching*. Oxford: Oxford University Press.
- . 2009. "Task-based language teaching: sorting out the misunderstandings." *International Journal of Applied Linguistics* 19 (3): 221-46.
- Fulcher, Glenn, Fred Davidson, and Jenny Kemp. 2011. "Effective rating scale development for speaking tests: Performance decision trees." *Language Testing* 28 (1): 5-29.
- Fulcher, Glenn. 1999. "Assessment in English for academic purposes: Putting content validity in its place." *Applied Linguistics* 20 (2): 221-36.
- . 2017. "Scoring performance tests." In *The Routledge handbook of language testing*, edited by Glenn Fulcher and Fred Davidson, 178-92. Abingdon, Oxon, New York: Routledge.

- Harsch, Claudia, and Guido Martin. 2013. "Comparing holistic and analytic scoring methods: Issues of validity and reliability." *Assessment in Education: Principles, Policy & Practice* 20 (3): 281–307.
- Knoch, Ute. 2009. "Diagnostic assessment of writing: A comparison of two rating scales." *Language Testing* 26 (2): 275–304.
- Norris, John M. 2016. "Current uses for task-based language assessment." *Annual Review of Applied Linguistics* 36:230–44.
- Plakans, Lia. 2012. "Writing Integrated Items." In Davidson and Fulcher 2012, 249–261.
- Robinson, Peter. 2001. "Task complexity, task difficulty, and task production: Exploring interactions in a componential framework." *Applied Linguistics* 22 (1): 27–57.
- . 2011. "Task-based language learning: A review of issues." *Language Learning* 61 (Supplement s1): 1–36.
- Sadler, D. Royce. 2009. "Grade integrity and the representation of academic achievement." *Studies in Higher Education* 34 (7): 807–26.
- Skehan, Peter. 2003. "Task-based instruction." *Language Teaching* 36 (01): 1–14.
- Tavakoli, Parvaneh. 2009. "Investigating task difficulty: learners' and teachers' perceptions." *International Journal of Applied Linguistics* 19 (1).
- Van den Branden, Kris. 2006. *Task-based language education: From theory to practice*. Cambridge, New York: Cambridge University Press.
- Weigle, Sara Cushing. 2009. *Assessing Writing*. 4th ed. Cambridge: Cambridge University Press.
- Weigle, Sara Cushing, and Keisha Parker. 2012. "Source text borrowing in an integrated reading/writing assessment: Textual appropriation and source use in L2 writing." *Journal of Second Language Writing* 21 (2): 118–33.
- Westhoff, Gerard J. 2009. "A priori assessment of language learning tasks by practitioners." *Language Teaching* 42 (4): 504–18.
- Wigglesworth, Gillian. 2008. "Task and performance based assessment." In *Language testing and assessment*, edited by Elana Shohamy and Nancy H. Hornberger, 111–22. Encyclopedia of language and education 7. Boston, MA: Springer.
- Willis, Dave, and Jane R. Willis. 2011. *Doing Task-Based Teaching*. 5th ed. Oxford: Open University Press.