View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Ghent University Academic Bibliography

35

Task-Based Language Assessment

Koen Van Gorp
Katholieke Universiteit Leuven, Belgium

Bart Deygers
Katholieke Universiteit Leuven, Belgium

Introduction

Task-based language assessment (TBLA) is an approach to language assessment that focuses on what learners are able *to do* with language as opposed to what they *know* about language. Central to TBLA is the notion of tasks. The performances on tasks provide a teacher or a test user with authentic and contextually relevant information about the (second) language development or language performance of a student. In spite of the pedagogic benefits associated with a task-based approach, it remains a domain in language testing and language teaching that faces many challenges. In particular, questions of reliability, content validity, and authenticity remain to be researched more thoroughly (Wigglesworth, 2008; Norris, 2009; Bachman & Palmer, 2010). This chapter will present some challenges TBLA faces when used formatively (based on the assessment framework of a Dutch language method for primary education) as well as summatively (i.e., for reasons of certification as shown by the Certificate of Dutch as a Foreign Language).

Task-Based Language Assessment and Task-Based Language Teaching

TBLA is a crucial element within task-based language teaching (TBLT) (Van den Branden, 2006b; Norris, 2009). In TBLT tasks are essential pedagogic constructs that "drive" classroom activity (Samuda & Bygate, 2008). In the language-teaching literature, a task is defined in various ways (for a recent overview, see Van den Branden, 2006a; Samuda & Bygate, 2008). Van den Branden (2006a) defines a task as "an activity in which a person engages in order to attain an objective, and which necessitates the use of language" (p. 4). According to Van den Branden, students

The Companion to Language Assessment, First Edition. Edited by Antony John Kunnan.
© 2014 John Wiley & Sons, Inc. Published 2014 by John Wiley & Sons, Inc.
DOI: 10.1002/9781118411360.wbcla058

learn a language when provided with opportunities to use authentic language meaningfully (i.e., meeting the language use needs of learners and society) and engagingly (i.e., in active and interactive language-learning processes). In their definition of a "language use task," Bachman and Palmer (2010) include the notion that a task is always situated in a particular setting. In our view this provides a useful addition to the definition of a task because it adds a socially situated dimension to the definition (McNamara & Roever, 2006; Firth & Wagner, 2007). Expanding Van den Branden's definition of a task leads us to the following definition: A task is a functional activity in a particular setting in which an individual uses language to attain an objective.

In line with the communicative and functional language that tasks elicit, language testers have become interested in task-based assessment following a general move away from purely discrete-point, indirect testing. TBLA is not a new phenomenon, but builds on earlier concerns with communicative language testing (e.g., Morrow, 1979) and language performance assessment (e.g., McNamara, 1996). As such, TBLA subscribes to a framework of language that can be described as "can do/performance"-oriented rather than "ability"-oriented (Bachman, 2011). The basic tenet of TBLA links up with Cureton's advice: "If we want to find out how well a person can perform a task, we can put him to work at that task, and observe how well he does it and the quality and quantity of the product he turns out" (Cureton, 1950, p. 622).

A summative assessment task from *TotemTaal*, a Dutch language syllabus for primary schools (see below), can serve as an example of the performance-oriented approach. In this test students in the final grade of primary school are asked to write a brochure about the rain forest. Throughout the preceding unit, "Jungle fever," the students have been exploring life in the rain forest. At the end of the unit they are asked to write three short informative texts: one describing an animal, one describing a plant, and a last one describing the importance of the rain forest. The students write their texts based on a number of preset questions and on rich visual input. Furthermore, the students are given work sheets that induce them to structure their performance, to work out a catchy title, and to make their texts visually attractive. The teachers give feedback using a rating scale that deals with form and content but also aligns the performance with the expected writing skills as formulated in the syllabus.

The example above illustrates how TBLA allows for a dynamic interaction between cognitive, contextual, and linguistic variables that govern real-life language performance (Skehan, 1998). Furthermore, it stimulates the natural integrated use of language skills (Colpin & Gysen, 2006) and allows for the use of compensatory strategies in situations of real language use (Norris, Brown, Hudson, & Bonk, 2002).

Since TBLA relies heavily on meaningful, real-world language performance, authenticity is a vital task component. Ideally, task-based assessment should directly reflect the tasks and interactions that learners are expected to perform (i.e., interactional authenticity) in real-life situations (i.e., situational authenticity) within a particular domain. However, when writing authentic tasks, identifying the target language use (TLU) may prove problematic. As Bachman (2002) points out, not all tests have a clearly defined TLU. Determining authentic content for a

test based on preset goals may be unproblematic (see *TotemTaal* below), whereas identifying the TLU for "broader" contexts might prove problematic. In those contexts, Bachman stresses the importance of a needs analysis (see "Summative Use of Task-Based Performance Tests" below), while pointing out that authenticity in testing is not without its limits. Indeed, not all real-life tasks can or should be operationalized.

TBLA is "an approach that attempts to assess as directly as possible whether test takers are able to perform specific language tasks in particular communicative settings" (Colpin & Gysen, 2006, p. 152). As such, TBLA's construct of interest is task performance itself (Long & Norris, 2000). By emphasizing and assessing task performance, TBLA performs three main functions (Norris, 2009): (1) offering formative or diagnostic feedback to learners and teachers (i.e., assessment *for* learning); (2) enabling summative decisions that are indicative of targeted language-learning outcomes (i.e., assessment *of* learning); and (3) raising the awareness of learners, teachers, and other stakeholders about what language learning is all about by emphasizing valued and authentic language performance and target-task learning throughout the program (i.e., washback). In performing these three functions, TBLA practices provide a crucial link between language objectives and the educational program. The following paragraphs discuss a formative and a summative approach to TBLA. The formative use will be illustrated by a task-based assessment framework for language teaching in primary education, whereas curriculum-independent task-based certification tests will serve to exemplify the summative use of TBLA.

Formative Assessment in a Task-Based Language Syllabus for Primary Education

Most of the discussion about TBLA concerns its summative use. However, as Ellis (2003) points out, teachers will benefit most from a formative TBLA approach. Formative assessment allows teachers to be responsive to learner needs by indicating what students have learned or still need to learn, by providing information about curriculum planning and teaching (e.g., the suitability of classroom activities), and by offering relevant and meaningful learner feedback (Rea-Dickins & Gardner, 2000; Rea-Dickins, 2001). Especially in classroom practice, the distinction between formative and summative assessment is not as straightforward as sometimes portrayed. Formative assessment is not always "tidy, complete and self-consistent, but fragmentary and often contradictory" (Harlen & James, 1997, p. 374). Rea-Dickins and Gardner (2000) refute the idea that cumulative data collection in classroom assessment automatically leads to a reliable and valid representation of learner performance. They also point out that classroom assessment that is generally considered to be low stakes can have serious implications for individuals or groups of learners, and is in that sense high stakes. As a result, issues of reliability and validity should be treated with the same rigor for formative and summative assessment alike.

Notwithstanding its occasional "messiness," formative assessment has the potential to advance students' language learning (Rea-Dickins, 2001). When used

well, it produces coherent evidence about language learners' abilities to perform specific target tasks. To this end, TBLA has to provide "frameworks for tracking and interpreting important aspects of learner development over time" (Norris, 2009, p. 587). Therefore, teachers should be able to do more than acknowledge whether students have performed a specific task successfully. Teachers should be aware of the task specifications, of expected task performance, and of task performance strategies so they can help learners improve their performance. For those reasons, TBLA needs to rely on an assessment framework that generates rich information about in-class learning and teaching processes. Consequently, for teaching purposes and purposes of formative assessment, tasks should be conceptualized as a set of characteristics, rather than holistic entities (Bachman, 2002). These characteristics will be inherent to the task itself, but will also relate to learner characteristics. Task performance yields information about the interaction between learners and tasks, and it is precisely this information teachers need to assess students' progress as well as their ability to perform certain tasks.

Task Specification Framework

One example of how TBLA can provide a close understanding of language learners' development is the assessment framework in *TotemTaal*, a task-based Dutch language syllabus for Flemish primary education in Belgium (Berben et al., 2008b). *TotemTaal* was developed from 2005 to 2008 by a team of task-based syllabus designers at the Centum voor Taal en Onderwijs of the University of Leuven in order to provide teachers with a task-based pedagogy in class. It is a Dutch-language curriculum encompassing listening, speaking, reading, writing, spelling, and language awareness tasks for both first and second language (L1 and L2) learners from grade two of primary education onwards.

As stated above, a good starting point for determining the language goals of any task-based syllabus (and consequently TBLA) is a needs analysis (Long, 2005). For *TotemTaal*, however, the attainment goals were predetermined by the Flemish Department of Education and the curricula by the Flemish educational networks. Based on these legally fixed and unalterable goals, pedagogic tasks were sequenced to ensure varying complexity and content over the different grades of primary school (for a detailed description, see Colpin & Van Gorp, 2007).

To enable task sequencing, monitor task complexity, and track learning opportunities, a task specification framework was developed. This framework defines task characteristics by means of six parameters. Each task challenges students to practice one or more of the four *language skills* while processing or producing a *text type* for a certain *public*, about a specific topic, representing or revealing a *world*, with a certain *function or purpose*. Dealing with the information in the text demands a certain *level of processing*. In addition, the text can be linguistically easy or difficult depending on vocabulary, syntax, structure, code, conventions, and so on. Table 35.1 illustrates these parameters for a reading task.

The task specification framework constitutes the backbone of *TotemTaal*'s task-based approach and of its assessment framework. It guarantees content relevance and representativeness for both pedagogic and assessment tasks. The collection of tasks "Lost in the forest" for grade 5 illustrates *TotemTaal*'s assessment framework.

Table 35.1 Task specification framework for the reading task "Which way out?" (5th grade) in *TotemTaal* (Berben et al., 2008a). Adapted with permission from the authors

Goals							
Parameters▶	Skill	Level of processing	Text-type	Public	World	Function	Attainment goal
Settings▶	Reading	Evaluating	Informative texts	(Un) known peers	Orientation (scientific description)	Inform	3.4

In the unit "Crispy fairy tales," 10–11-year-olds first read part of a poem by Roald Dahl (from his book *Rhyme Stew*) about Hansel and Gretel, who are abandoned by their parents. The students have to find out where and when Hansel and Gretel were abandoned and got lost by reading and interpreting Dahl's rhymes (e.g., "They walk, all four, for hours and hours / They see no robins, pick no flowers. / The wood is dark and cold and bare" [Dahl, 2003, p. 354]). After reading the poem the students look for a way to help Hansel and Gretel find their way out of the forest. In the task "Which way out?" the students are presented with several informative 100 word texts from children's magazines about different means of orientation, such as the compass in the example below (see Table 35.1 for the specifications of this reading task).

> *A compass*
> The needle of a compass always points north. If you know this, you can work out the other wind directions. If you turn until the arrow points to the front of you, then the east lies to your right, the west to your left and the south is behind you. Always keep your compass away from steel and engines, because iron and electricity influence the direction of the compass needle. (Berben et al., 2008c, p. 36)

In pairs, the students determine the best way for Hansel and Gretel to get out of the forest, while taking into account the poem's context: an unknown forest at night. The students reach a decision and formulate arguments as to why their suggested procedure would be successful for Hansel and Gretel (e.g., why wind directions help Hansel and Gretel find their way home). Before discussing their solutions with the whole class, students individually reflect on the task "Which way out?" They decide whether the reading task went well and, using a work sheet, they go through a list of strategies that can help them to reflect on their task performance. After the class discussion the teacher discusses the students' reflections on their reading accomplishments.

Assessment Framework

The assessment framework of *TotemTaal* consists of four components, of which the first three are present in the lesson "Lost in the forest." The implementation of these components in the classroom will be illustrated by the case of Caroline and her teacher. Table 35.2 provides an overview of the four components.

Table 35.2 Assessment framework in *TotemTaal* (Berben et al., 2008b). Adapted with permission from the authors

Function	Incidental formative assessment	Planned formative assessment		Summative assessment
Format	1. Observation and assistance	2. Observation and analysis	3. Reflection	4. Tests
Who	Teacher	Teacher	Student and teacher	Student
Focus	Looking at students' task performance (process)	Looking at students' task performance (process) and outcome (product)	Looking at own task performance and outcome	Looking at students' outcomes of task-based tests (product)
Pedagogic tools	*Guidelines* for teacher assistance of four language skills	*Frameworks and guidelines* for the observation and analysis of four language skills	Teacher *guidelines* for reflective talks *Portfolio* guidelines for students	*Task-based tests* for listening, reading, writing, spelling, and language awareness
Written form	No; "mental notes"	Systematized notes	Work sheets; portfolio	Test score

These four opportunities for gathering information about the students' developing language skills have partly overlapping intended purposes:

1. observation of task performance in order to provide teacher assistance if necessary,
2. observation and detailed analysis of task performance and task outcome of "targeted" individual students,
3. learner reflection and portfolio for self-assessment of task performance and language proficiency level, and
4. task-based tests for summative use.

The first component, observation and teacher assistance, could be viewed as strictly instructional: reading a poem. However, the reading task allows for in-task performance and process assessment. Every pedagogic task offers an opportunity to observe task performance and to assist students in successful task performance. Every task also allows the teacher to take "mental notes" about the students' strengths and weaknesses in a particular task performance. The guidelines in the teacher manual help the teacher to assess task performance and pinpoint the main task components, (i.e., the reading goal and the information-processing demands) and possible pitfalls concerning poem comprehension. The following teacher's note on 11-year-old Caroline deals with her comprehension of the reading goal: "Caroline had difficulty in deducing from the poem that Hansel and Gretel were abandoned in the forest at night. Was this because of the poetic structure of the

text? Check next reading assignment whether she's able to connect several pieces of literal information in the text."

The second observation component is a planned formative assessment (Ellis, 2003). During the task "Which way out?" the teacher focuses primarily on whether students understand the reading goal and on their manner of information processing. The teacher writes down any difficulties a student encounters during task performance as well as the coping strategies he or she may have used. These notes provide an intentional and systematic process and product evaluation of the student's task performance and supplement the teacher's mental notes. If possible and necessary, the teacher informs the student about his or her observations in order to gain further insight into the student's reading. When a teacher steps in to assist a particular student, the teacher's actions get an instructional focus. Specific guidelines for observation and analysis support the teacher in shaping his or her analysis (see Table 35.3).

The guidelines in Table 35.3 are a specification and concretization of a more general skill-specific analytic framework or analysis diagram that informs the teacher on how to analyze reading performances in general. Table 35.4 presents this analytic framework for reading tasks.

The framework for reading tasks in Table 35.4 provides the teacher with information about which aspects are essential to the performance of reading tasks in general and with a systematic way of tracking this information for individual students. The aspects that were identified as relevant, based on recent meta-analyses of effective reading programs (e.g., National Reading Panel, 2000; Slavin, Lake, Chambers,

Table 35.3 Reading task "Which way out?": guidelines for analysis (Berben et al., 2008a). Abbreviated with permission from the authors

See Analysis Diagram for Reading Tasks [Table 35.4]:

- Reading goal: Does the student read the texts? Is the student looking for possible ways to help Hansel and Gretel orient themselves?
- Type of information: Does the student understand that he or she is looking for an appropriate way for Hansel and Gretel to orient themselves? Does the student take into account the circumstances in which Hansel and Gretel are lost?
- *Guidelines for teacher–student talk*:
  - Reading goal: Does the student know that he or she has to read the information about the forest first and next has to look for a way out for Hansel and Gretel?
  - Reading strategies: Is the student reading all the texts with the same eye for detail or not? Is the student interpreting the titles of the different texts?
  - Relationships in the text: Does the student understand the relevant information in the texts? Ask questions: *How can you orient yourself using your watch and the sun?*
  - Visual aspects: Can the student carry out the instructions in the texts on the basis of the drawings?
  - Attitudes: Is the student motivated to find out how Hansel and Gretel can find their way out of the forest?

Table 35.4 Analysis diagram for reading tasks (Berben et al., 2008b). Abbreviated with permission from the authors

Name of the student:		*1st period of observation and analysis*		*2nd period of observation and analysis*		. . .	
		S	*H*	*S*	*H*	*S*	*H*
Reading goal: Is the student's reading goal-oriented?							
Can the student perform the reading task with the text? If so, has he or she understood the reading task and has he or she read in such a way as to reach the reading goal?							
Information processing: Is the student able to find the information he or she is looking for?							
Describe	Can the student find literal or explicitly mentioned information in the text?						
Structure	Can the student connect several pieces of literal information from the text?						
Structure	Can the student find implicit information in the text?						
Evaluate	Can the student compare information from the text with information from a second source or evaluate the information based on his or her own personal frame of reference?						
If not, hold a conversation with the student where you try to find out what went wrong							
Identifying the reading goal	Can the student identify the reading goal?						
Topic	Is the student familiar with the topic?						
Strategies	Does the student go about the reading task in an adequate manner?						
Other:							
Overall: How does the student perform with respect to							
Self-reliance	Does the student attempt to resolve the task on his or her own? Makes he or she use of the tools (strategies) at his or her disposal?						
Attitudes	Willingness to read, reading pleasure, willingness to reflect on own reading behavior						
Reflective ability	Does the student gradually develop the ability to think about his or her own reading skills? Does he or she apply these insights in subsequent reading tasks?						

S = self. H = with help or support from teacher or other student(s). Use + for positive performance, – for negative, ± for things in between.

Cheung, & Davis, 2009), are reading goal, level of information processing, topic, reading strategies, self-reliance, attitudes, and reflection. Other aspects are technical reading skills (e.g., fluency and accuracy), conventions of the text type, relations in the text (e.g., function words expressing grammatical relations), vocabulary, and visual aspects (e.g., illustrations and layout). These aspects, specifically the focus on reading goal, on levels of processing, and on reading strategies, connect all components in *TotemTaal*'s assessment framework. In the teacher manual, the guidelines for teacher assistance, observation, and analysis and those for student reflection specify how these aspects of the reading process can be realized in a specific pedagogic task and what realistic expectations are for students of a certain grade.

In our example, the teacher takes the opportunity the formative assessment task provides to focus on Caroline and to observe her reading behavior in more detail. Observation and intervening (e.g., scaffolding) with Caroline's task execution provide the teacher with new evidence about Caroline's ability to connect several pieces of information from the text. This information will help the teacher when completing Caroline's analysis diagram (see Table 35.4).

Whereas the observation and analysis of task performances in the first two components is carried out by the teacher, the third component of the assessment framework encompasses learner reflection and portfolio. Learner reflection allows students to assess their own language experiences and their own language skills. It allows students to gradually build up their ability to self-monitor, to reflect metacognitively, and to regulate their own learning processes. Before discussing solutions with the whole class, students individually reflect on the success of performing the reading task "Which way out?" using a work sheet with a list of strategies (e.g., identifying the reading goal, making use of the illustrations, reading all texts or not, looking at the titles of the texts, underlining relevant information in the texts, and so forth). The strategies on the work sheet relate to the analytic framework for reading tasks. This component becomes instructionally relevant when teacher and student discuss the student's analysis and decide on alternative strategies for tackling future tasks. Through the use of a consistent framework, both teacher and learners are provided with both a frame of reference and a common language to talk about the effectiveness of their reading skills and learning processes. In the case of Caroline, the teacher takes the time to reflect with Caroline on her reading strategies. The teacher focuses on two of the strategies on the work sheet: "I used the illustrations in the text: yes/no" and "I underlined important information in the text: yes/no." Teacher and student discuss whether using these strategies would have helped Caroline in her task performance. Caroline agrees to pay more attention to these strategies and try them in the next reading task.

The fourth and final component in the assessment framework is task-based tests that are performed without teacher or peer support. These tests are directly linked to the attainment goals and provide the teacher with a complementary view of student task performance. These more formal task-based language tests also allow for diagnostic information, since the test items are directly linked to the different information-processing levels in the above-mentioned analytic framework. The test result can then underscore or refine the teacher's analysis of the student's information-processing abilities based on formative assessment, which is likely to improve the reliability of the overall teacher assessment. In this reading test, the

teacher in our example can again determine whether Caroline succeeds in independently combining pieces of information from a text. The teacher can check this directly since the rating scale links particular test items to specific information-processing demands (see also Table 35.4).

In summary, the assessment framework in *TotemTaal* provides teachers with a rich and balanced assessment repertoire that combines "classical" tests with "alternative assessment" (Fox, 2008). The assessment framework in *TotemTaal* is thus not diametrically opposed to that of traditional tests, but embeds curriculum-based tests in a broader assessment and teaching approach. It emphasizes the need for multiple assessment procedures or multiple sources of assessment evidence (Shohamy, 1996) and the collection of multiple performances over time to provide evidence of growth and learning. Multiple sources or procedures enable teachers to make a variety of inferences about the capacities for language use that students have, or about what they can or cannot yet do.

Information about students' task performances is continuously gathered by teachers, both informally and formally. It is also provided by the students' self-assessment. The assessment framework in *TotemTaal* encourages teachers to look at students' performances for both product-evaluation and process-evaluation purposes, allowing for a smooth transition from instruction to assessment. This kind of assessment is an indispensable part of a responsible task-based pedagogy. Most crucially, the formative assessment components are an inextricable part of good teaching (Rea-Dickins, 2001).

Summative Use of Task-Based Performance Tests: The Certificate of Dutch as a Foreign Language

Task-based language testing is widely used summatively. Summative tests determine a student's language skills by using target tasks as an indicator of ability to function in a particular TLU domain. The Canadian Language Benchmarks and the Occupational English Test are well-known examples of task-based language tests, a typology to which the Certificate of Dutch as a Foreign Language (CNaVT) also belongs. Each year, some 3,500 candidates in 40 countries sit one of the six profile-based tests that belong to the CNaVT suite. These tests are either societal (tourism and citizenship), professional (services and administration), or academic (student and tutor) in domain. The current suite of task-based tests is the result of a significant paradigm shift in the test development process of the CNaVT. Having administered competence-based tests of Dutch as a foreign language from the 1970s until the 1990s, the CNaVT moved toward performance-based testing in 1999, which entailed a fundamental reconceptualization of the test construct. For one thing, language level ceased to be the major focal point of the test. Instead, task performance and task outcome (Skehan, 1996) became essential.

Why Assess?

Turning a test inside out brings up a number of fundamental questions, starting with the most basic one of all: Why assess? In the test development process, the

fundamental consideration which precedes all other concerns about test construct and test specifications focuses on the very motive for assessing language at all (Van den Branden, 2006a; Bachman & Palmer, 2010). Since Dutch is quite a small language with limited international resonance, most students of Dutch take it up with a specific goal in mind and require certification to attain that goal, which may be societal, professional, or academic in nature. For other students, a Dutch language test has no other goal than to serve as a motivational yardstick that indicates the extent to which they are able to function in a specific domain.

When reshaping the CNaVT, a needs analysis was conducted in order to identify the test takers' reasons for taking a Dutch L2 test (Van Avermaet & Gysen, 2006). The needs analysis allowed for six profiles to be identified within the three domains. To date, each profile is monitored and updated on a regular basis so as to ensure the representativeness and authenticity of its task types as well as the abilities that are required to perform those tasks (Van Avermaet & Gysen, 2006).

To a large extent, the goals of test takers and those of the CNaVT as a test provider run parallel. Test takers wish to be tested reliably, validly, and authentically so as to be adequately informed about their level of linguistic competence within a given domain. Apart from these shared goals of test takers and test developers, it is of great importance to the CNaVT to generate positive washback. By using functional task-based tests, the CNaVT wishes to inspire teachers to consider using functional tasks in their teaching practice. Both the certification and the washback motive are entwined with the central philosophy that one learns a language not only to use it, but also by using it.

Are Task-Based Language Tests Testing Language for Specific Purposes?

The six profiles that are now in use adhere to the central concept of task-based performance testing, which implies employing real-world tasks as assessment material. Since task context will also shape the performance, the context of the task goes beyond what Douglas (2001, p. 172) calls "situational window dressing," and is a vital part of each task. The context in which a task is situated will not only influence the expected register but also, and more fundamentally, co-determine the rating criteria.

Since the CNaVT tasks are authentic and contextualized they are comparable to language for specific purposes (LSP) tests, from which they differ on one crucial point: content knowledge. CNaVT task constructs purposefully eliminate all subject-specific knowledge, which is a defining element of LSP tasks (Douglas, 2000), but also a possible impediment for test fairness if test takers have dissimilar backgrounds (Bachman & Palmer, 2010). In a sense one could define CNaVT tasks as generic LSP tasks: Each task calls upon a contextualized ability which is necessary within but not necessarily exclusive to a specific domain. Presentation tasks, for example, occur in tests of academic Dutch and business Dutch, but the contexts differ. In a recent test of business Dutch, candidates were presented with three possible locations for a large-scale conference. They were asked to pick one venue, based on a number of parameters such as capacity and cost, and argue the case for their choice as the best option for the conference. The criteria used focused on

the adequacy of the presentation in a business context. In the same year, the test of academic Dutch featured a presentation task in which candidates were asked to present a relatively general study concerning Internet use among youngsters. Here too, the rating criteria considered the context in which the task took place. Even though the tasks are similar at the core, the differing content and setting effectively alter the nature of the task. Accuracy, for example, is more important in the academic test, whereas persuasion is a more prominent criterion in the business context.

Striving toward "authentic" tests that reflect the context the test taker will be in (Wu & Stansfield, 2001) goes beyond the tasks and extends to the rating scales. In the example above, the two presentation tasks were rated using dissimilar rating scales because of the different TLU settings. In order to attain this level of authenticity, the CNaVT calls upon the expertise of stakeholders and domain specialists. Including subject specialists in task selection and task development is an important step in developing contextualized task-based tests.

Determining Rating Criteria

Consulting subject specialists employed in the target domain to specify their "indigenous" criteria (Jacoby & McNamara, 1999) constitutes an extra step worth taking in developing authentic, domain-related performance tests (Jacoby & McNamara, 1999; Douglas, 2001). Using domain experts does not imply that testing professionals should take a step back when determining the criteria to be used in the rating scales. Rather, it entails broadening the horizon by using indigenous criteria that "may be used . . . to supplement linguistically oriented criteria" (Douglas, 2001, p. 183). Since the importance and presence of formal linguistic elements and content-related criteria are decided in coordination with subject specialists or domain experts, a given criterion might be considered less important for one task than for the next, depending on the goal. In other words: Consulting domain experts for task selection as well as tapping into their "rich inventory of tacitly known criteria" (Jacoby & McNamara, 1999, p. 224) for establishing rating criteria can increase a test's content validity and its predictive validity.

In an effort to further refine its rating criteria for the academic profiles, the CNaVT conducted a study in 2010 and 2011. The study consulted professionals working within an academic context in which they come into regular contact with students. Ten subject specialists took part in two focus group sessions and a larger group filled out an online questionnaire (n = 231). The respondents of the focus groups were first asked to voice the intuitive criteria they employ when determining the quality of a performance. Later they had the opportunity to refine their intuitive criteria based on test performances. After the first run, which was based on tasks, not on performances, most criteria dealt with content-related matters rather than formal aspects of language. The second session of the focus group, based on task performances, showed a slightly different picture. The relative importance of content diminished whereas the salience of formal aspects of language (e.g., linguistic correctness) increased. In the questionnaire, the respondents wrote down which criteria they employed when deciding on the quality of a performance. Next, they were asked to arrange in order of importance the criteria

that were established in the focus groups. The combined data from the focus groups and the questionnaire allowed for a unique perspective on authentic rating criteria. From the data, the CNaVT was able to conclude that content and structure were generally considered most important. Prototypical linguistic criteria such as spelling and vocabulary were always present, but considered less important than content, structure, and grammatical features. In future, similar data may help "determine the relative importance of different aspects of language ability for a given purpose of assessment in a particular context" (Sawaki, 2007, p. 356), which can in turn influence the weighting of criteria. In line with the logic of considering authentic criteria, the weight that is given to a criterion will differ from task to task and from profile to profile. In the focus groups and the questionnaire results, the criterion vocabulary, for example, was considered quite important for integrated writing tasks, but not important at all for an integrated speaking task in the context of a meeting with student services.

The above shows that the CNaVT has adopted not a holistic, but an analytic rating process in which subscores determine the final outcome. Using analytic rating has been shown to be reliable when using trained novice raters (Barkaoui & Knouzi, 2011), as is often the case for the CNaVT. Additionally, analytic rating scales allow for fine-grained distinctions between criteria. In spite of the different criteria that are used for rating, the test results only distinguish between pass and fail, which is in line with Long and Crookes's (1992) statement that TBLA should be organized "by way of task-based criterion referenced tests, whose focus is whether or not students can perform some task to criterion, as established by experts in the field, not their ability to complete discrete-point grammar items" (p. 45).

Washback

As discussed above, the primary purpose of the CNaVT is to provide test takers with authentic and reliable task-based tests that employ valid rating criteria. The second function the CNaVT aims to perform has less to do with testing than with the influence of a test on teaching. The CNaVT's washback philosophy is inspired by the belief that people learn a language in order to use it but also by using it within a meaningful context. Since "well-designed assessment tasks have the potential to provide positive wash-back into the classroom" (Wigglesworth, 2008, p. 114), the CNaVT aims to have an impact on classroom practice by introducing task-based approaches (e.g., by using engaging, real-world tasks as representative practice material) in the slipstream of the formalized tests.

Challenges and Future Directions

Since the early 1990s, TBLT has gained considerable momentum in the field of language education. In its wake, TBLA has developed as a medium summative orientation as well as, although to a lesser extent than, formative feedback. Apart from these two assessment functions, TBLA has proven itself to be a means for raising awareness with all stakeholders about language-learning processes and the ability to perform a variety of valued communication tasks. Although the

benefits of direct and authentic task-based tests are evident, there remain a number of challenges and unanswered research questions. The use of authentic and contextualized tasks in high stakes tests raises questions about generalization that are still waiting to be addressed (Bachman, 2002). Generalization is one of the major issues that links up with the validity of construct interpretations as well as with how task features act as sources of variability in task performances. With respect to formative assessment, it is still unclear whether TBLA can produce a less messy classroom practice than is often observed in research. Developing a coherent assessment framework in a task-based curriculum as *TotemTaal* does is important because it builds up an argument for a reliable and valid TBLA and provides a much-needed interface between theory and practice. However, it is in the implementation of the assessment framework by the teacher in the classroom that the real strengths and weaknesses of TBLA in informing instruction and learning processes are revealed. To ensure the efficiency and effectiveness of TBLA practices, "validity evaluations" (Norris, 2008) have to be set up within educational programs. This research is of great value to prove the utility and worth of TBLA both for teachers in classroom settings and for test institutes developing high stakes certification tests.

SEE ALSO: Chapter 37, Performance Assessment in the Classroom; Chapter 41, Dynamic Assessment in the Classroom; Chapter 42, Diagnostic Feedback in the Classroom; Chapter 72, The Use of Generalizability Theory in Language Assessment

References

Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, *19*(4), 454–76.

Bachman, L. F. (2011, July). *How do different language frameworks impact language assessment practice?* Plenary talk presented at ALTE 4th International Conference, Kraków, Poland.

Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford, England: Oxford University Press.

Barkaoui, K., & Knouzi, I. (2011, July). *Rating scales as frameworks for assessing L2 writing: Examining their impact on rater performance*. Paper presented at ALTE 4th International Conference, Kraków, Poland.

Berben, M., Callebaut, I., Colpin, M., François, S., Geerts, M., Goethals, M., . . . & Van Gorp, K. (Eds.). (2008a). *TotemTaal: Themahandleiding en kopieerbladen 5B*. Mechelen, Belgium: Wolters Plantyn.

Berben, M., Callebaut, I., Colpin, M., François, S., Geerts, M., Goethals, M., . . . & Vanoosthuyze, S. (Eds.). (2008b). *TotemTaal: Inleiding en evaluatie 5*. Mechelen, Belgium: Wolters Plantyn.

Berben, M., Callebaut, I., Colpin, M., François, S., Geerts, M., Goethals, M., . . . & Vanoosthuyze, S. (Eds.). (2008c). *TotemTaal: Werkboek 5B*. Mechelen, Belgium: Wolters Plantyn.

Colpin, M., & Gysen, S. (2006). Developing and introducing task-based language tests. In K. Van den Branden (Ed.), *Task-based language education: From theory to practice* (pp. 151–74). Cambridge, England: Cambridge University Press.

Colpin, M., & Van Gorp, K. (2007). Task-based writing in primary education: The development and evaluation of writing skills through writing tasks, learner and teacher support. In K. Van den Branden, K. Van Gorp, & M. Verhelst (Eds.), *Tasks in action: Task-based language education from a classroom-based perspective* (pp. 194–234). Newcastle, England: Cambridge Scholars Press.

Cureton, E. E. (1950). Validity. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 621–94). Washington, DC: ACE.

Dahl, R. (2003). *The Roald Dahl treasury*. London, England: Penguin Books.

Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge, England: Cambridge University Press.

Douglas, D. (2001). Language for specific purposes assessment criteria: Where do they come from? *Language Testing*, *18*(2), 171–85.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford, England: Oxford University Press.

Firth, A., & Wagner, J. (2007). Second/foreign language learning as a social accomplishment: Elaborations on a reconceptualized SLA. *Modern Language Journal*, *91*(Suppl. s1), 800–18.

Fox, J. (2008). Alternative assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education. Vol. 7: Language testing and assessment* (2nd ed., pp. 97–108). New York, NY: Springer.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education*, *4*(3), 365–79.

Jacoby, S., & McNamara, T. (1999). Locating competence. *English for Specific Purposes*, *18*(3), 213–41.

Long, M. (2005). *Second language needs analysis*. Cambridge, England: Cambridge University Press.

Long, M., & Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly*, *26*(1), 27–56.

Long, M. H., & Norris, J. M. (2000). Task-based language teaching and assessment. In M. Byram (Ed.), *Encyclopedia of language teaching* (pp. 597–603). London, England: Routledge.

McNamara, T. (1996). *Measuring second language performance*. New York, NY: Longman.

McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell.

Morrow, K. (1979). Communicative language testing: Revolution or evolution? In C. J. Brumfit & K. Johnson (Eds.), *The communicative approach to language teaching* (pp. 143–57). Oxford, England: Oxford University Press.

National Reading Panel. (2000). *Reports of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Rockville, MD: NICHD Clearinghouse.

Norris, J. M. (2008). *Validity evaluation in language assessment*. Frankfurt, Germany: Peter Lang.

Norris, J. M. (2009). Task-based teaching and testing. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 578–94). Malden, MA: Wiley-Blackwell.

Norris, J. M., Brown, J. D., Hudson, T. D., & Bonk, W. (2002). Examinee abilities and task difficulty in task-based second language performance assessment. *Language Testing*, *19*(4), 395–418.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, *18*(4), 429–62.

Rea-Dickins, P., & Gardner, S. (2000). Snares and silver bullets: Disentangling the construct of formative assessment. *Language Testing*, *17*(2), 215–43.

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. London, England: Palgrave Macmillan.

Sawaki, Y. (2007). Construct validation of analytic rating scales in a speaking assessment: Reporting a score profile and a composite. *Language Testing*, *24*(3), 355–90.

Shohamy, E. (1996). Language testing: Matching assessment procedures with language knowledge. In M. Birenbaum & F. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 143–59). Boston, MA: Kluwer.

Skehan, P. (1996). A framework for the implementation of task based instruction. *Applied Linguistics*, *17*(1), 38–62.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford, England: Oxford University Press.

Slavin, R., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research*, *79*(4), 1391–466.

Van Avermaet, P., & Gysen, S. (2006). From needs to tasks: Language learning needs in a task-based approach. In K. Van den Branden (Ed.), *Task-based language education: From theory to practice* (pp. 17–46). Cambridge, England: Cambridge University Press.

Van den Branden, K. (2006a). Introduction: Task-based language teaching in a nutshell. In K. Van den Branden (Ed.), *Task-based language education: From theory to practice* (pp. 1–16). Cambridge, England: Cambridge University Press.

Van den Branden, K. (Ed.). (2006b). *Task-based language education: From theory to practice*. Cambridge, England: Cambridge University Press.

Wigglesworth, G. (2008). Task and performance based assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education*. *Vol. 7: Language testing and assessment* (2nd ed., pp. 111–22). New York, NY: Springer.

Wu, W. M., & Stansfield, C. W. (2001). Towards authenticity of task in test development. *Language Testing*, *18*(2), 187–206.

Suggested Readings

Brindley, G. (2009). Task-centered language assessment in language learning: The promise and the challenge. In K. Van den Branden, M. Bygate, & J. Norris (Eds.), *Task-based language teaching: A reader* (pp. 434–54). Amsterdam, Netherlands: John Benjamins.

Brown, J. D., Hudson, T. D., Norris, J. M., & Bonk, W. (2002). *Investigating task-based second language performance assessment*. Honolulu, HI: University of Hawai'i Press.

Byrnes, H. (2002). The role of task and task-based assessment in a content-oriented collegiate FL curriculum. *Language Testing*, *19*(4), 425–33.

Davies, A. (2001). The logic of testing languages for specific purposes. *Language Testing*, *18*(2), 133–47.

Gysen, S., & Van Avermaet, P. (2005). Issues in functional language performance assessment: The case of the Certificate of Dutch as a Foreign Language. *Language Assessment Quarterly*, *2*(1), 51–68.

Norris, J. M. (2002). Interpretations, intended uses and designs in task-based language assessment. *Language Testing*, *19*(4), 337–46.

Skehan, P. (2001). Tasks and language performance assessment. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 167–85). Harlow, England: Longman.