CrossMark

Annual Review of Applied Linguistics, **36** (2016), pp. 230–244.
© Cambridge University Press, 2016
doi: 10.1017/S0267190516000027

Current Uses for Task-Based Language Assessment

JOHN M. NORRIS
Georgetown University
norrisj@georgetown.edu

ABSTRACT

Task-based language assessment (TBLA) has generated interest since the early 1990s, primarily in conjunction with the ongoing development of task-based language teaching (TBLT) and the pursuit of developing appropriate testing models for this approach to pedagogy (Norris, 2002, 2009). However, tasks also offer considerable advantages for language assessment, beyond their obvious relevance within TBLT classes and programs. In fact, major innovations in the general domain of language assessment over the past two decades have occurred in conjunction with the introduction of tasks into assessment design, largely in response to the need for tests that better represent examinees' abilities to use the language (Mislevy, Steinberg, & Almond, 2002), but also because tasks offer a meaningful space for language teachers, testers, learners, and others to examine, understand, and improve language learning endeavors (Van Gorp & Deygers, 2013). This article reviews the considerable range of current uses for TBLA, illustrating different types of assessment with concrete examples and highlighting distinct roles for tasks as a basic unit of analysis in test design, interpretation, and intended consequence. Ultimately, it argues that tasks offer a fundamental, though not exclusive, foundation for useful language assessment, and that task-based assessment, though challenging, is probably worth the effort.

INTRODUCTION

Language assessment, and educational and psychological assessment more generally, has enjoyed a tremendous period of growth and increasing sophistication in the modern era, in particular since the end of World War II, as institutions and societal sectors have come to depend increasingly on systematic, empirical evidence as a basis for making decisions about individuals and groups. Indeed, the 20th century has been marked dramatically by the rise of the science of psychometrics (i.e., the development of highly efficient tests of various human attributes) and an emphasis on large-scale, norm-referenced testing that can reliably rank individuals according to differences in valued abilities, aptitudes, and so on. Education in particular has been shaped and reshaped by the prominent use of such assessments, typically in association with reform movements but also as part and parcel of standard teaching practice.

By contrast, and of particular interest for the current article, a variety of so-called alternative assessments emerged in the final decades of the 20th century, to a large extent in response to such testing traditions, as educators sought other ways to make sense of learners' knowledge and abilities, and as they tried to accomplish

https://doi.org/10.1017/S0267190516000027 Published online by Cambridge University Press

distinct aims from those typically associated with large-scale testing (e.g., Wiggins, 1998). From portfolios to performance assessments, new approaches to testing were explored as a way of reorienting teaching and learning to meaningful goals (e.g., ability to use knowledge) and away from an emphasis on rote memorization and discrete facts, and as assessments were deployed for new purposes such as criterion-referenced achievement testing, classroom-based formative feedback, and others. In language testing, too, an interest in alternative assessments found considerable purchase in the 1990s, as a variety of new assessment methods were developed and put into practice for purposes beyond norm-reference measurement (Brown & Hudson, 1998).

Task-based language assessment (TBLA), like performance assessment in mainstream education (e.g., Kane, Crooks, & Cohen, 1999), appeared during this time as one type of alternative to traditional testing. Brindley (1994) was probably the first to observe and codify the basic dimensions of TBLA, though interest in language performance assessments had certainly been in evidence for some time (see review in McNamara, 1996), as well as practiced through several prominent testing formats like the oral interview technique for assessing speaking proficiency. Task-based assessments, though, raised considerable new interest because of their obvious alignment with the development of task-based language teaching (TBLT; Skehan, 1998) and a concomitant need for assessments that reflected student learning of target task abilities, stressed authenticity of testing methods and rating criteria, and enabled predictions about the extent to which learners could actually use the language rather than merely know something about its rules and vocabulary (Norris, Brown, Hudson, & Yoshioka, 1998).

In the next few years, a flurry of publications appeared that pushed the limits of theorizing about TBLA, the contributions it might make, and the challenges that it presented for language test development and validation (e.g., Bachman, 2002; Brown, Hudson, Norris, & Bonk, 2002; Chalhoub-Deville, 2001; Long & Norris, 2000; Norris, 2002; Norris, Brown, Hudson, & Bonk, 2002; Skehan, 2001; Wigglesworth, 2001). Interestingly, with a few exceptions (e.g., Byrnes, 2002; Norton & Stewart, 1999; Van den Branden, Depauw, & Gysen, 2002), the related discussions and debates were largely hypothetical (and research largely lab-based), as language testers postulated about how TBLA might work and what kinds of problems it posed. Beyond considering the obvious roles for task-based assessment in task-based classrooms, these discussions also played an important part in spurring critical thinking about possible changes needed in other kinds of language assessments, where increasing demands for authenticity, interaction, and language use might well be met by the introduction of more communicative tasks. Practical examples of TBLA in use, though, were hard to come by at the time, and where they did emerge in the literature, they generally reported on small-scale adoption of task-based approaches to assessment within individual programs that were a priori disposed to TBLT ideas.

Over the intervening decade and a half, much has happened at the interface of task-based ideas and language assessment, both in conjunction with the ongoing development of TBLT as an approach to education and in the broader scheme of

https://doi.org/10.1017/S0267190516000027 Published online by Cambridge University Press

language testing. Though additional discussions about TBLA have emerged in the literature (e.g., Bachman, 2007; Brindley, 2013; Brown, 2004), to date there has been no attempt to consider the scope of actual implementation of related ideas in practice, that is, for operational testing purposes. Therefore, the modest goal of this article is to provide a summary overview of the *practice* of task-based language assessment in the current era, generally seeking to survey the extent to which TBLA ideas have appeared in actual test development and use. Of course, the article makes no claims regarding a causal relationship between earlier discussions of TBLA and current practices, as that type of hypothesis testing is inevitably impossible. Nevertheless, it may be that highlighting the ways in which tasks are being used for practical language assessment purposes can both emphasize the importance of such ideas and identify possible dimensions for future work.

As a point of embarkation, I will offer a general definition of what I think constitutes TBLA, as follows: the elicitation and evaluation of language use (across all modalities) for expressing and interpreting meaning, within a well-defined communicative context (and audience), for a clear purpose, toward a valued goal or outcome.

TASK-BASED LANGUAGE ASSESSMENT IN PRACTICE

If a meaningful measure of the value of an idea is the extent to which it has gained purchase in actual application, then the past decade or so has provided ample evidence in support of the increasing value of task-based approaches to language assessment. Not only have language performance tasks continued to play important roles in classroom-based testing, but their use—often in combination with emerging technological affordances—has expanded within this and other assessment contexts in order to accommodate the real-world needs of test users. Teachers, admissions officers, employers, and others seek transparent indications of the extent to which language learners can communicate in a target language. Task-based assessments seem to fit the bill. In this brief review, I highlight some of the primary uses for TBLA, as it has been put into practice around the world in relation to language education and other domains where language tests play vital decision-supporting roles. The intent here is not to provide a comprehensive account, rather it is to signify the distinct levels of endeavor at which TBLA is being adopted, to summarize key purposes for a task-based assessment approach at each of these levels, and to illustrate the central contribution of tasks in responding to critical assessment demands of various stakeholders across diverse educational or societal sectors.

Tasks as Standards

A first, broad level related to assessment has to do with the role of tasks in the establishment of language policies and learning or ability standards by governmental or professional bodies. Such policies and standards are employed typically as a means for setting out the expectations of language ability (or aspirations for

https://doi.org/10.1017/S0267190516000027 Published online by Cambridge University Press

language learning) that are held for particular individuals, groups, or communities, and they reflect the values espoused by policymakers and, in turn, society. For example, the following task descriptions are a few of the numerous standards stipulated for speakers of English as a second language who wish to become certified to work in the nursing profession in Canada (see CELBAN Centre, http://www.celbancentre.ca/):

Workplace Tasks: Reading

- Scan basic charts, tables, maps, or schedules for information.
- Find routine information on the computer screen, scanner screen, or computerized display screen, if available.
- Read information in the reception or appointment book to find available openings for a new appointment.

These task-based standards, in turn, form the basis for assessments that are used to determine the readiness of nurses (in this case) to use English effectively on the job; thus, for example, the Canadian English Language Benchmark Assessment for Nurses (CELBAN) is a "language assessment that reflects nursing in real life" (CELBAN Centre, http://www.celbancentre.ca/) and which is used as part of the licensure procedure for nurses in Canada. Tasks, in their role as the basic building blocks of such ability standards, provide a crucial means for conveying in no uncertain terms what it is that language is being used to do, and therefore what learners and teachers need to work toward, as well as what testers and tests need to focus on.

It is this practical and transparent aspect of tasks—their ability to encapsulate and signify purposeful language use—that also underlies their adoption as a key dimension of the more generalized Canadian Language Benchmarks (CLB; Centre for Canadian Language Benchmarks, 2012) for all adult English language learners:

The CLB are Task-Based

> The notion of the language task—a communicative "real world" instance of language use to accomplish a specific purpose in a particular context—is central to the CLB. When instructors or assessors describe communicative language ability, they are describing a person's ability to accomplish communicative language tasks for particular contexts. Accomplishing communicative tasks provides the learner, instructor, or assessor with demonstrable and measurable performance outcomes. (p. ix)

For one example comparable to the reading standards for nursing English above, Reading Benchmark 3 in the CLB (Centre for Canadian Language Benchmarks, 2012, p. 81) is called "Getting Things Done," and it includes an overall sketch of the expected task abilities, as well as sample assessment tasks:

- Get information from simple formatted texts (such as forms, tables, charts, schedules, and directories).
  - Recognizes layout.

https://doi.org/10.1017/S0267190516000027 Published online by Cambridge University Press

  - Identifies layout and specific information.
  - Scans formatted text (such as a directory) to find specific information.
- Get information from short business or service texts (such as brochures, notices, form letters and flyers).
  - Gets overall meaning.
  - Interprets simple graphics.
  - Demonstrates strengths and limitations typical of Reading Benchmark 3, as listed in the Profile of Ability.

Sample Tasks

Read a simple product advertisement to make a decision about whether to buy it.
Check a bank account balance and make a payment online.
Read a short notice about staff kitchen rules.

Several hundred such task descriptions form the heart of the CLB standards across the four modalities and at several levels of ability, serving to make real the reason why language is (or should be) being learned and how it needs to be assessed.

For similar reasons, other large-scale language policies and standards have been founded on the basis of, or at least in substantial reference to, tasks. The *Common European Framework of Reference* (Council of Europe, 2001), *Hong Kong Target-Oriented Curriculum* (e.g., Curriculum Development Council, 1993), *Standards for Foreign Language Learning* (National Standards in Foreign Language Education Project, 1996), and *Standards for the Recognition of Initial TESOL Programs in P-12 ESL Teacher Education* (Teachers of English to Speakers of Other Languages, 2010) are additional examples of language policy documents that draw explicitly and heavily upon tasks as a means for depicting expectations about language learning and ability and, in particular, to provide a basis for language assessment development and validation. For tasks used in the creation of broad standards such as these, emphasis is placed both on the comprehensive range of representation (i.e., the scope of different kinds of language tasks expected) and the degree of task complexity (i.e., the level of ability required to accomplish various tasks). That is, standards seek to depict not only the types of tasks that learners can do, but also a scale of ability reflected in doing different tasks and in doing them to greater or lesser degrees of accomplishment.

Tasks in Proficiency Assessment

A next level of endeavor has to do with the critical role played by tasks in large-scale, standardized language proficiency assessments that seek, increasingly, to enable accurate interpretations about learners' abilities to actually use the target language to accomplish valued communicative goals. By and large, such assessments are used predictively to determine the extent to which second language users can demonstrate the kinds of language abilities that will be called upon in particular discourse domains (e.g., university study), and they lead most typically to decisions about program admissions and/or placement into opportunities for further language study. Thus, in one example, an examinee might be asked to first

https://doi.org/10.1017/S0267190516000027 Published online by Cambridge University Press

read a passage from a psychology textbook explaining a particular concept, then listen to a recorded lecture by a professor that illustrates the same concept, and finally to answer the following question orally in English: "Explain *flow* and how the example used by the professor illustrates the concept" (Educational Testing Service [ETS], 2016). This task, taken from the Test of English as a Foreign Language iBT (ETS, 2016) online practice materials, is a good example of the extent to which simulations of authentic communication tasks have come to be common practice on international tests of English for academic purposes proficiency. Indeed, in the original documents that framed the major redesign of the TOEFL test some 15 years ago (Jamieson, Jones, Kirsch, Mosenthal, & Taylor, 2000), the importance of such tasks was clearly stated: "The test will measure examinees' English language proficiency in situations and *tasks* reflective of university life" (p. 11, emphasis added). The resulting Internet-based test (the iBT) not only featured new writing and speaking performance tasks, but also incorporated integrated-skills tasks (see Chapelle, Enright, & Jamieson, 2011).

Other large-scale tests of similar scope have followed suit. For example, the International English Language Testing System (IELTS, 2012) "assesses the English language proficiency of people who want to study or work where English is used as the language of communication" (p. 1), hence "IELTS is a *task-based* test covering the four skills" (p. 5, emphasis added). Examples of other assessments from around the world that explicitly tap into examinees' abilities to do language performance tasks include:

- Australia: the Adult Migrant English Service assessments
- Flanders/Holland: the Certificate in Dutch as a Foreign Language
- Spain: the Diploma in Spanish as a Foreign Language[1]

The inclusion of increasingly performative and complex communication tasks in these and other standardized proficiency assessments is reflective of the desire by test developers to better approximate real-life contexts and constructs of academic (or other) language use (Gysen & Van Avermaet, 2005). That is, such assessments seek to capture the functional nature of communication that learners, teachers, and educational administrators are concerned with, and which is at best only indirectly related to previously predominant assessment formats that emphasized knowledge of, rather than ability with, the target language (e.g., multiple choice tests of vocabulary or grammar). For tasks used in standardized proficiency assessments, the emphasis is on prototypicality, comparability, and scalability; thus, individual assessment tasks are designed to represent task types or families, so that interpretations can be generalized to a broader domain of similar tasks, and to reflect distinct levels of language ability, so that the likelihood of accomplishing such tasks can be determined.

Tasks for Employment Certification

A third level of task-based assessment has to do with the use of tasks in the certification of profession-related language skills or competencies. Increasingly, it is clear

https://doi.org/10.1017/S0267190516000027 Published online by Cambridge University Press

that the language abilities required for successful communication within distinct workplace domains and genres are heavily influenced by the specific requirements of the actual tasks undertaken therein on a regular basis (Douglas, 2000; Elder et al., 2012). Assessments that are used to determine an employee's readiness for corresponding language demands must therefore replicate typical task procedures, content, situations, interlocutors, and other factors, in order to provide trustworthy indications of the extent to which learners can handle the real-world requirements of task performance on the job. For example, an examinee might be asked to first listen to an authentic recorded exchange between an English-language air traffic controller and a pilot who is making an approach for landing at a busy airport, after which the examinee could be prompted to explain in detail what happened in the interaction and whether the interlocutors made any procedural or content errors that might threaten the safety or effectiveness of the landing procedure, followed by other tasks that put the examinee in the position of the air traffic controller. It should be clear that (a) accomplishing this kind of task is not a reasonable expectation for most language learners, but (b) for those who expect to use language in high-stakes settings like that of air traffic or marine communication, task-based assessments of this sort are probably essential for tapping into the likelihood of a candidate's abilities to use the language on the job (Alderson, 2009). Indeed, according to the International Civil Aviation Organization PRICE Study Group (2007), "Because of the high stakes involved, pilots and controllers deserve to be tested in a context similar to that in which they work and test content should therefore be relevant to their roles in the work-place." For this reason, task-based assessments have been developed to replicate the demands of these and similar high-stakes professional environments that require language use under very specific circumstances (for one recent example of a task-based assessment for aviation English, see Park, 2015).

Another common example of profession-specific TBLA has to do with the workplace demands of medical communication for nurses, doctors, and others. The CELBAN, mentioned above, and the Occupational English Test (OET, n.d.) in Australia are good examples. The OET (n.d.) is advertised as follows: "OET has been developed specifically for 12 healthcare professions. Tests simulate real workplace tasks and test relevant language skills for each profession" (emphasis added). In the speaking subskill test, for example, candidates must complete several role-plays that present typical tasks and challenges associated with the interaction between a particular type of health care provider and a patient; performances are rated not only according to factors such as fluency, intelligibility, and appropriateness, but primarily for "overall communicative effectiveness," that is, the extent to which the task was deemed to be successfully accomplished (see OET, n.d.). Other examples of task-based assessments for certifying job-related language abilities range from tests of taxi drivers to call center employees to teachers (e.g., Elder, 2001; Lockwood, 2015). For tasks used in this kind of assessment, selection of critical or key job-related tasks—those that define the primary language-related challenges of the workplace—is essential, as is the careful simulation of authentic situational and interactional factors that define the actual demands of the tasks in real life. Additionally, it is critical to consider when and why such assessments

https://doi.org/10.1017/S0267190516000027 Published online by Cambridge University Press

are used. The extent to which learners or employees should be held accountable to the rigorous authentic demands of a specific task may depend on whether the assessment is being used as a screening device for potential employees versus an on-the-job accountability tool, the former calling for broadly representative tasks and the latter for highly specific job tasks (see discussion in Lockwood, 2015).

Tasks for Language Educational Assessment

At a fourth and final level of endeavor, task-based assessments have found their way increasingly into a variety of uses in language education classes and programs. Fundamentally, a primary motivation for incorporating TBLA into language education has to do with the need to align assessment, curriculum, and instruction such that they complement each other in supporting effective language learning. That is, as language teaching moves persistently to embrace notions such as communicative competence, task-based instruction, ability for language use, and so on, it is a clear imperative that assessments will need to support rather than contradict or undermine the corresponding educational values and emphases. A focus on discrete-point tests of knowledge about language, when instruction emphasizes ability to use language, may result in confusion and a misalignment of teaching and learning with the outcomes that are actually assessed. Accordingly, assessments that require learners to perform with the language in order to accomplish communicative goals associated with valued tasks have been adopted increasingly as the assessments of choice in language educational settings around the world, for both summative and formative purposes (Norris, 2009).

For a summative example, in a project funded by the European Union—the Guidelines for University Language Testing (GULT) project—Fischer, Chouissa, Dugovičová, and Virkkunen-Fullenwider (2011) described a rationale for greater incorporation of task-based assessments for determining learner achievement within university language education in Europe:

> By using a task-based approach to an LSP [language for specific purposes] test in a specific realistic situation based on facts and authentic documents, the test takers can show to what extent they are able to communicate in a foreign language, which makes the results more useful and relevant for them. This type of test is also more transparent for the test takers, as they themselves can judge more easily how well they managed, linguistically, a realistic communication situation. The overall aim of the GULT Project is therefore to develop guidelines and models for transparent task-based LSP testing at university. (p. 15)

Fischer et al. (2011) went on to provide numerous illustrations of how such assessments were being used in practice across European universities, primarily for outcomes assessment at the end of sequences of instruction. Such an approach, they argued, not only supports the communicative instruction that is being implemented and encourages learner awareness raising, but it also provides a better indication of the value and extent of learning that has occurred. For another example, where task-based assessment is being used to trace learning outcomes in L2 writing

https://doi.org/10.1017/S0267190516000027 Published online by Cambridge University Press

abilities for both young learners and adults, see Alanen, Huhta, and Tarnanen (2010).

From a formative perspective, task-based assessment is being put into practice with the primary aims of raising learner and teacher awareness about language learning processes and outcomes, and in particular to provide a rich basis for feedback in various directions (for early work along these lines, see Byrnes, 2002). For example, the Integrated Performance Assessment project in the United States (e.g., Adair-Hauck, Glisan, Koda, Swender, & Sandrock, 2006) provided foreign language teachers with a systematic approach to eliciting learner performance on tasks in the target language, assessing performance according to learning expectations, and providing feedback that both teachers and learners can act upon to improve learning outcomes. In a concrete illustration of this assessment approach, learners might first be asked to read about health and nutrition in Spanish and from the perspective of the Spanish-speaking world, and then to self-assess their own understanding of the content and language in the reading (thereby raising their awareness about their reading and task abilities). Next, two learners might have a discussion in Spanish about what they found out in the readings, and each learner would then assess the partner's abilities to express meaning in a clear manner, according to a rubric describing important dimensions of oral interactional competence (thereby offering feedback from an external perspective on individual language ability). Finally, the learners might each produce a letter of application describing why they should be selected for a sports camp in Spain, offering the teacher a more formalized opportunity to provide feedback on their ability to use language effectively in the performance of a particular type of task.[2] The introduction of similar types of primarily formative task-based assessments into language classes, programs, and school systems has been documented and explored worldwide, for example, in Belgium (e.g., Colpin & Gysen, 2006; Van Gorp & Deygers, 2013), Hong Kong (e.g., Hong Kong SAR, n.d.), Japan (e.g., Weaver, 2013), and New Zealand (e.g., East, 2012).

For these classroom- or curriculum-embedded assessments, tasks tend to emphasize extended, complex, skills-integrated, and multistage performances with the language, such that learners have ample opportunity to demonstrate what they have learned and in order to provide repeated and rich occasions for learning- and instruction-related feedback. Along these lines, recent interest in project-based, scenario-based, and learning-oriented assessment has spurred considerable additional test development on the basis of extended language task performances, the objectives of which seem to be (a) to create a more seamless instruction-assessment-feedback relationship within language teaching and (b) to wash back onto language teaching and learning practice by providing meaningful and rich assessment information related to the functional uses of language for communication purposes (see, e.g., Purpura & Turner, 2014). Also interesting here are developments in technology that allow for largely or entirely computer-mediated task-based assessments that employ virtual realities, avatars, and other affordances—in conjunction with automated scoring and feedback mechanisms—to embed learners increasingly within "real" language use environments where they have to engage

https://doi.org/10.1017/S0267190516000027 Published online by Cambridge University Press

in purposeful, interactive, and individualized communication to a degree that was heretofore very challenging to realize in the traditional classroom environment. There is little doubt that the rapid escalation of technological capabilities coupled with a major focus on learning-oriented assessment will only increase the use of authentic communication tasks as a foundation for language assessments into the future.

In Summary

Clearly, tasks seem to play an increasingly prominent role in many levels of and uses for language assessment: descriptions of language tasks are useful for conveying valued learning and ability targets in policies and standards; prototypical tasks offer a meaningful basis for eliciting and generalizing about learners' functional language proficiency, rather than mere knowledge, in order to make admissions, placement, and other high-stakes decisions; high-fidelity simulations of specific tasks enable the certification of language users' abilities within critical occupational domains; language performance tasks provide well-aligned exemplars of learning outcomes for determining achievements in educational programs that focus on developing learners' ability to use the language; and extended, complex, and integrated tasks establish a rich space within which learners and teachers can collaboratively examine various aspects of language development and, critically, generate meaningful feedback to further improve learning and instruction. While the characteristics and qualities of tasks vary in apparent and intentional ways across these distinct levels, what is common to all is the idea that language assessment—in these cases at least—should be based in no small part on tasks that require functional language use for the meaningful conveyance of ideas within a clearly defined communicative setting. Given that commonality, it is not inaccurate to suggest that these are all examples, to varying degrees, of what might be termed TBLA.

WHY BOTHER WITH TASK-BASED LANGUAGE ASSESSMENT?

The task-based reorientation of various types and levels of language assessment, as indicated by the examples above, is probably one of the major, if relatively unacknowledged, developments in language testing of the recent era. Nevertheless, it is also the case that designing, implementing, and evaluating TBLAs are by no means simple tasks. Numerous overviews and critique articles have been written about task-based assessment over the past two decades, and they have tended to emphasize challenges and concerns with TBLA (e.g., Bachman, 2002, 2007; Brindley, 2013; Brown, 2004; Chalhoub-Deville, 2003; Mislevy et al., 2002; Norris et al., 1998; Wigglesworth, 2008). Prominent issues raised include the extent to which tasks and associated language use contexts can be adequately simulated within controlled testing conditions, the authenticity of criteria for evaluating task performances, the reliability of raters in judging task performances, the

https://doi.org/10.1017/S0267190516000027 Published online by Cambridge University Press

comparability of tasks in terms of construct representation as well as level of difficulty, the generalizability and extrapolation limitations of tasks as a basis for assessment design, and the practicality and expense involved in creating and using task-based tests. While these are all legitimate points of consideration, they also tend to reflect a rather monolithic notion of assessments as measures of individuals' underlying abilities or traits—that is, from this perspective, the primary role of tests is to give us an adequate snapshot of examinees' knowledge and abilities, to the extent that we can reliably infer about their underlying language competence.

What I believe the array of examples provided above suggests, by contrast, is that task-based assessments are used for a variety of purposes, each emphasizing quite distinct qualities for assessment design (and, indeed, validation). While some assessments, such as standardized proficiency tests, may call for a theory of individual ability traits as the measurement construct underlying task selection and operationalization, other assessments, such as those found in occupational testing, may tend to focus much more squarely on the particular requirements of a given task and language use situation, because accomplishing that specific task is essential to the job at hand. Others assessments may prioritize features of tasks not typical of traditional assessment regimes, such as flexibility or manipulability of the task in eliciting rich performances from diverse examinees in order that meaningful feedback can be provided to all, or highly engaging task content that maximizes the learner's motivation and attention, washing back on their desire to learn language. Task-based assessment, then, even though all varieties share a core interest in functional language use, appeals to distinct assessment needs through quite diverse operational approaches. While this dimension of TBLA also presents a challenge to language testers, in that it may require them to think outside of the traditional test development and validation box, it also presents a critical opportunity for taking the intended uses of assessments seriously as a starting point for assessment design. Along these lines, in a special issue of *Language Testing* on TBLA, published in 2002, I posed and answered this question: "Why should language testers bother with trying to design tests that simulate language use contexts, the communication tasks that occur within them and the criteria whereby performances on such tasks may be judged? The answers lie, I think, in the actual variety of uses for language tests." (Norris, 2002, p. 343).

This line of response requires expansion, also in light of the variety of assessments that seem to be invoking task-based ideas in the contemporary era. To be clear, in my point of view, TBLAs may not (should not) meet all purposes for assessment within language educational, occupational, or other domains of endeavor, but they probably do offer a lot of advantages for certain purposes. Tasks per se seem particularly useful in relation to contemporary notions of language teaching and learning:

- Task-based performances align well with curriculum and instruction that focuses on functional language use.
- Task-based performances tell us things that other item types do not (e.g., ability to accomplish specific tasks under particular conditions).

https://doi.org/10.1017/S0267190516000027 Published online by Cambridge University Press

- Task-based performances offer rich opportunities for eliciting and observing language use with meaningful content in context; there is always more to communicative language use than just language, and tasks capture these other dimensions rather well.
- Task-based performances can reveal multiple aspects of language ability and/or development within a single instance (e.g., accuracy, complexity, and fluency; content knowledge; procedural knowledge; expertise; and pragmatic sensibility).
- Task-based performances are in many cases the most indicative targeted expectations of language classes and programs, thereby serving as useful representations of learning outcomes. (And a corollary concern to this assumption is that if they do not reflect our expectations, then what exactly are we teaching in language education?)

Furthermore, adopting tasks as the unit of analysis for language assessments enables the prioritization and realization of a variety of assessment dimensions that may otherwise be marginalized:

- Task-based assessments provide relevant and comprehensive frames of reference for learners and teachers to generate and act upon feedback related to form-function-meaning relationships, supporting increasingly important formative, learning-oriented uses that are called for in language education and related testing.
- Task-based assessments communicate to teachers and learners and others what is valued within a given educational setting, raising awareness about what is to be learned, why, and how.
- Task-based assessments introduced as innovations may impact curriculum and educational systems, causing a rethinking of what and how teaching and learning is occurring (i.e., leading to intentional positive washback).
- Task-based assessments generate interpretable and publically meaningful data regarding the worthwhile outcomes of language learning programs (i.e., summative uses for accountability, certification, program evaluation).

In the end, then, it may be that TBLA is worth the effort. Where assessments are put to use for purposes that require the qualities described above, it would seem irresponsible in many cases not to engage in something akin to TBLA. Yes, it is possible to create quite reliable, psychometrically sound, and highly efficient shortcut measures that correlate highly with other criteria for language ability (e.g., Bernstein, Van Moere, & Cheng, 2010). However, what is possible is not necessarily always what is needed. Indeed, it is important to ask whether such assessments possess other qualities that are called for in educational assessments that seek to do something more than or different from just measuring well—sound measures on their own may not respond to the intended uses and consequences of a given assessment. Hence, where assessments are interpreted to indicate ability to use the language functionally, where assessments are used to predict the accomplishment of specific target tasks, where assessments must offer insights into complex dimensions of interactional or intercultural competence, where assessments enable meaningful formative feedback, where assessments should align with and potentially wash back on curriculum and instruction that emphasize language for

https://doi.org/10.1017/S0267190516000027 Published online by Cambridge University Press

communication purposes—under these circumstances, TBLA makes a lot of sense, and the effort to design task-based assessments to fit these and related intended uses might itself be worth the effort. Indeed, accumulating examples from diverse settings around the world (e.g., Adair-Hauck et al., 2006; Davison, 2007; East, 2015; Van Gorp & Deygers, 2013) have begun to provide evidence that task-based assessment innovations can exert a considerable positive influence on language teachers, learners, and educational systems, as they are gradually and carefully put into practice. Perhaps those kinds of consequences are reason enough to bother engaging with TBLA.

NOTES

1. For more on these assessments see AMES Australia, https://www.ames.net.au; CNaVT, http://www.cnavt.org; and Diplomas of Spanish as a Foreign Language, http://www.dele.org.

2. For more on this and related integrated performance assessment (IPA) tasks, see Foreign Language Educators of New Jersey, Consortium for Assessing Performance Standards, http://flenj.org/CAPS/?page=parent.

REFERENCES

Adair-Hauck, B., Glisan, E., Koda, K., Swender, E., & Sandrock, P. (2006). The Integrated Performance Assessment (IPA): Connecting assessment to instruction and learning. *Foreign Language Annals*, *39,* 359–382.

Alanen, R., Huhta, A., & Tarnanen, M. (2010). Designing and assessing L2 writing tasks across CEFR proficiency levels. In I. Bartning, M. Martin, & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between language testing and SLA research* (pp. 21–56). EUROSLA Monograph. Retrieved from http://www.eurosla.org/eurosla-monograph-series-2/

Alderson, J. C. (2009). Air safety, language assessment policy, and policy implementation: The case of aviation English. *Annual Review of Applied Linguistics, 29,* 168–187.

Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing, 19*(4), 453–476.

Bachman, L. F. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 41–71). Ottawa, Canada: University of Ottawa Press.

Bernstein, J., Van Moere, A., & Cheng, J. (2010). Validating automated speaking tests. *Language Testing, 27*(3), 355–377.

Brindley, G. (1994). *Task-centered assessment in language learning: The promise and the challenge*. Retrieved from ERIC database. (ED386045)

Brindley, G. (2013). Task-based assessment. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. London, UK: Blackwell. doi: 10.1002/9781405198431.wbeal1141

Brown, J. D. (2004). Performance assessment: Existing literature and directions for research. *Second Language Studies*, *22*(2), 91–139.

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, *32*(4), 653–675.

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

Byrnes, H. (2002). The role of task and task-based assessment in a content-oriented collegiate foreign language curriculum. *Language Testing*, *19*(4), 419–437.

Centre for Canadian Language Benchmarks. (2012). *Canadian language benchmarks: English as a second language for adults*. Ottawa, Canada: Retrieved from http://www.cic.gc.ca/english/pdf/pub/language-benchmarks.pdf

Chalhoub-Deville, M. (2001). Task-based assessment: Characteristics and validity evidence. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 210–228). Harlow, UK: Longman.

Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language Testing*, *20*(4), 369–383.

https://doi.org/10.1017/S0267190516000027 Published online by Cambridge University Press

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.). (2011). *Building a validity argument for the Test of Teaching English as a Foreign Language*. London, UK: Routledge.
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. 151–174). Cambridge, UK: Cambridge University Press.
Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.
Curriculum Development Council. (1993). *Guide to the primary school curriculum*. Hong Kong: Curriculum Development Institute, Education Department.
Davison, C. (2007). Views from the chalkface: English language school-based assessment in Hong Kong. *Language Assessment Quarterly, 4,* 37–68.
Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge, UK: Cambridge University Press.
East, M. (2012). *Task-based language teaching from the teachers' perspective: Insights from New Zealand*. Amsterdam, The Netherlands: John Benjamins.
East, M. (2015). Coming to terms with innovative high-stakes assessment practice: Teachers' viewpoints on assessment reform. *Language Testing, 32,* 101–120.
Elder, C. (2001). Assessing the language proficiency of teachers: Are there any border controls? *Language Testing*, *18*(2), 149–170.
Elder, C., Pill, J., Woodward-Kron, R., McNamara, T., Manias, E., Webb, G., & McColl, G. (2012). Health professionals' views of communication: Implications for assessing performance on a health-specific English language test. *TESOL Quarterly*, *46*(2), 409–419.
Educational Testing Service (ETS). (2016). *TOEFL iBT® test questions* [Electronic document]. Retrieved from https://www.ets.org/toefl/ibt/prepare/test_questions
Fischer, J., Chouissa, C., Dugovičová, S., & Virkkunen-Fullenwider, A. (2011). *Guidelines for task-based university language testing*. Graz, Austria: European Center for Modern Languages.
Gysen, S., & Van Avermaet, P. (2005). Issues in functional language performance assessment: The case of the certificate Dutch as a foreign language. *Language Assessment Quarterly*, *2*(1), 51–68.
Hong Kong SAR. (n.d.). *Task-based assessment for English language learning at secondary level* [Electronic document]. Retrieved from http://cd1.edb.hkedcity.net/cd/eng/TBA_Eng_Sec/index.html
International Civil Aviation Organization PRICE Study Group. (2007). *ICAO policy on language proficiency testing*. Montreal, Canada: Author.
International English Language Testing System (IELTS). (2012). *IELTS guide for teachers*. Manchester, UK: British Council.
Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (2000). *TOEFL 2000 framework*. Princeton, NJ: Educational Testing Service.
Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice*, *18*(2), 5–17.
Lockwood, J. (2015). Language for specific purpose (LSP) performance assessment in Asian call centres: Strong and weak definitions. *Language Testing in Asia*, *5*(3). doi: 10.1186/s40468-014-0009-6
Long, M. H., & Norris, J. M. (2000). Task-based teaching and assessment. In M. Byram (Ed.), *Encyclopedia of language teaching* (pp. 597–603). London, UK: Routledge.
McNamara, T. (1996). *Measuring second language performance*. New York, NY: Longman.
Mislevy, R., Steinberg, L., & Almond, R. (2002). Design and analysis in task-based language assessment. *Language Testing*, *19*(4), 477–496.
National Standards in Foreign Language Education Project. (1996). *Standards for foreign language learning: Preparing for the 21st century*. Yonkers, NY: Author.
Norris, J. M. (2002). Interpretations, intended uses and designs in task-based language assessment. *Language Testing*, *19*(4), 337–346.
Norris, J. M. (2009). Task-based teaching and testing. In M. Long & C. Doughty (Eds.), *Handbook of language teaching* (pp. 578–594). Cambridge, MA: Blackwell.
Norris, J. M., Brown, J. D., Hudson, T. D., & Bonk, W. (2002). Examinee abilities and task difficulty in task-based L2 performance assessment. *Language Testing*, *19*(4), 395–418.
Norris, J. M., Brown, J. D., Hudson, T. D., & Yoshioka, J. K. (1998). *Designing second language performance assessment*. Honolulu: University of Hawai'i Press.
Norton, B., & Stewart, G. (1999). Accountability in language assessment of adult immigrants in Canada. *Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes*, *56*(2), 223–244.
Occupational English Test (OET). (n.d.). *The English language test for the healthcare sector*. Retrieved from https://www.occupationalenglishtest.org/

https://doi.org/10.1017/S0267190516000027 Published online by Cambridge University Press

Park, M. (2015). *Development and validation of virtual interactive tasks for an aviation English assessment* (Unpublished doctoral dissertation). Iowa State University, Ames, IA.

Purpura, J. E., & Turner, C. E. (2014). *Learning-oriented assessment in language classrooms: Using assessment to gauge and promote language learning*. New York, NY: Routledge.

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

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. 167–185). Harlow, UK: Longman.

Teachers of English to Speakers of Other Languages. (2010). *Standards for the recognition of initial TESOL programs in P-12 ESL teacher education*. Alexandria, VA: TESOL.

Van den Branden, K., Depauw, V., & Gysen, S. (2002). A computerized task-based test of L2 Dutch for vocational training purposes. *Language Testing, 19,* 438–452.

Van Gorp, K., & Deygers, B. (2013). Task-based language assessment. In A. J. Kunnan (Ed.), *The companion to language assessment: Vol. II. Approaches and development* (pp. 578–593). West Sussex, UK: Wiley.

Weaver, C. (2013). Incorporating a formative assessment cycle into task-based language teaching. In A. Shehadeh & C. Coombe (Eds.), *Researching and implementing task-based language learning and teaching in EFL contexts* (pp. 287–312). Amsterdam, The Netherlands: John Benjamins.

Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass.

Wigglesworth, G. (2001). Influences on performance in task-based oral assessments. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 186–209). Harlow, UK: Longman.

Wigglesworth, G. (2008). Task and performance based assessment. In N. Hornberger (Ed.), *Encyclopedia of language education* (pp. 2251–2262). New York, NY: Springer.

https://doi.org/10.1017/S0267190516000027 Published online by Cambridge University Press