

Multilingual Testing Constructs: Theoretical Foundations

By: [Micheline B. Chalhoub-Deville](#)

Micheline B. Chalhoub-Deville (2019) Multilingual Testing Constructs: Theoretical Foundations, *Language Assessment Quarterly*, 16:4-5, 472-480, DOI: [10.1080/15434303.2019.1671391](https://doi.org/10.1080/15434303.2019.1671391)

This is an Accepted Manuscript of an article published by Taylor & Francis in *Language Assessment Quarterly* on 11 October 2019, available online: <http://www.tandfonline.com/10.1080/15434303.2019.1671391>.

*****© 2019 Taylor & Francis. Reprinted with permission. No further reproduction is authorized without written permission from Taylor & Francis. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document. *****

Abstract:

The field of language testing has made great strides in measuring language use. It is a monolingual construct, however, that anchors standardized language testing operations and classroom practices. Language use and performance research, see for example the MLJ special issue of 2011, demands that language testing operations also consider multilingual constructs. The present paper, generally guided by the published research as well as contributions in this special issue, examines how to expand the field's theoretical foundations to consider integrated multilingual testing constructs and translanguaging pedagogies.

Keywords: multilingualism | language testing | second language acquisition (SLA) | translanguaging

Article:

Introduction

The field of language testing has made great strides in measuring language use and performance. It is a monolingual construct, however, that invariably anchors standardized second/foreign language testing (L2) programs. Documented interactions in the real-world demand that language testing operations also consider multilingual (ML) constructs and the inclusion of other languages/first language (L1) to facilitate better assessment of content and educational achievement as well as language use (see the Modern Language Journal special issue of 2011). Shohamy (2011) writes that the construct underlying L2 assessments

“is of language as a closed and finite system that does not enable other languages to ‘smuggle in.’ This view is in stark contrast to the current understanding of ML competencies for which various languages and aspects ‘bleed’ into one another in creative ways as manifested by a growing number of users, especially immigrants, who

are born into one language and acquire additional language(s), resulting in multilingual competencies” (p. 418).

Similarly, Garcia and Wei (2014) lament the fact that “standardized assessments are usually administered in one language only” (p. 133). They report that reluctance among test developers to engage in ML assessment is one of the main challenges for adopting a flexible approach to translanguaging in educational settings. Limerick (2017), Saville (2017), and Schissel, Leung, López-Gopar, and Davis (2017) maintain that current assessment systems are problematic because they fail to support, if not undermine ML, plurilinguistic practices.

The present commentary is an extension of my role as a discussant in a 2017 symposium, entitled *The construct of multilingualism and language policies in language testing*, I co-organized with Schissel at the Language Testing Research Colloquium (LTRC), Bogota (Schissel & Chalhoub-Deville, 2017). The papers from the LTRC symposium, also published in the current journal issue, explore ML testing in various countries and regions of the world. They probe into classroom, standardized, and certification testing systems. The papers also examine theoretical, operational, and policy challenges of an ML orientation. In line with my symposium discussion, the present commentary examines how to expand current language testing theory and practice to consider ML testing constructs.

Multilingual realities and orientation

The shift in demographics speaks to the spread of ML communities and the need to reconsider ML construct representation not only in terms of L2 pedagogy but also in our testing operations. “We live in an increasingly multilingual world, where two out of three children grow up in an environment where more than one language is spoken. And in the UK the linguistic landscape is becoming ever more diverse – in London almost half the primary school population speaks a language other than English at home” (<http://www.ahrc.ac.uk/research/readwatchlisten/features/livinginamultilingualworld/>). A similar situation exists in the US. Projections show that “by 2055 (non-Hispanic) whites will no longer be the majority population in the U.S. ... and Mexican “hegemony” of immigration is expected to change, with Asians projected to be the largest immigrant group in the country” (Moll, 2017, p. xi). This shift in US demographics and the attendant ML reality is observed not only in large urban environments but also in small rural towns. “Large cities like New York and Los Angeles have been known for their diverse language communities, but even small towns and rural areas across the country have seen growth in the presence of DLC’s [diverse language communities]” (Gonzalez & He, 2017, p. 3). In the area where I live and work, i.e., in central North Carolina, “over 120 languages [are] spoken in the homes of Guilford County school aged children, with the top five being Spanish, Vietnamese, Arabic, Jarai (from the Montagnard community), and Urdu (Center for New North Carolinians, n.d.) (Gonzalez & He, 2017, p. 3). These quotes basically assert that multilingualism is a phenomenon that represents our world’s reality today and in the future.

Multilingualism is the norm for many individuals and communities in the world, which necessitates that the nature of ML constructs be considered closely in language testing. Table 1 presents some key features that differentiate between two divergent ML representations. At one

end of the continuum, the ML construct denotes multiple languages as separate systems with potentially a conflictual relationship. Languages are siloed, i.e., considered as separate systems with independent internal processes, and L1 and/or other languages besides the target language of learning are typically considered a detriment to the development of that language. On the other end of the continuum, multilingualism references flexible, dynamic, and complex systems that allow users to deploy resources as needed to take control of their language performance, enhance language learning, and access content.

Table 1. Divergent perspectives on Multilingual assessment orientation.

Deficit approach	Strength approach
←	→
<ul style="list-style-type: none"> • Subtractive perspective 	<ul style="list-style-type: none"> • Additive outlook
<ul style="list-style-type: none"> • Siloed languages 	<ul style="list-style-type: none"> • Integrated languages
<ul style="list-style-type: none"> • Monolingual standards 	<ul style="list-style-type: none"> • Plurilinguistic system
<ul style="list-style-type: none"> • Idealized native speaker 	<ul style="list-style-type: none"> • Real-world communicators
<ul style="list-style-type: none"> • Interlanguage development 	<ul style="list-style-type: none"> • Heteroglossic acquisition
<ul style="list-style-type: none"> • Code switching 	<ul style="list-style-type: none"> • Translanguaging

Similar to traditional, subtractive bilingualism, the pedagogic goal on the left end of the continuum is to move individuals in their language development to attain higher levels of performance along some monolingual standard based on idealized native speakers. The other end of the continuum portrays ML users from a position of strength in terms of their repertoire of languages. The assortment of languages within an individual are valued and allowed to develop as needed. This ML perspective operates with a plurilinguistic reference point that questions notions of idealized native speakers and calls for standards based on language use in specific contexts and ML communities, e.g. as observed with heritage learners.

A concept long upheld in the second language acquisition (SLA) literature is interlanguage (IL). IL references stage(s) of development as learners progress from an L1 base to the targeted L2. A strong feature of interlanguage is the distinction it displays between\among languages and the favoritism for an *uncontaminated* L2 use goal. This interlanguage concept is contrasted with a single heteroglossic processing system. This heteroglossic, ML construct references a unified language repertoire versus two/three/etc. autonomous or even interrelated language systems. In an ML construct perspective, as depicted on the right end of the continuum, all languages are flexibly deployed as needed in a specific interaction or setting, including a language assessment situation.

Closely connected with IL, is the term code switching. Code switching, long discussed in SLA, views crossing language boundaries in communication negatively. Translanguaging, on the other

hand, upholds a favorable outlook on the ML repertoire and on language mixing.

“Translanguaging is a new term in bilingual education; it supports a heteroglossic language ideology, which views bilingualism as valuable in its own right (MacSwan, 2017).

Translanguaging is common practice among individuals who overlap in their ML repertoires. At a very basic level, translanguaging is a “process of going back and forth from one language to the other” (García, 2009, p. 49). Translanguaging refers to the act of moving between/among languages to communicate. Researchers such as Canagarajah (2015), García (2009), Garcia and Wei (2014), Rojas (2014), Wei (2018) note that, compared to a compensatory communication outlook, which is traditionally depicted in code switching, translanguaging represents a positive orientation whereby ML resources are strategically deployed to make meaning, enhance interactions, and articulate identities. In pedagogy, translanguaging utilizes learners’ unified ML system to advance their academic development and achievement. The ‘trans’ aspect of translanguaging disrupts a traditional, cognitive, separate language, and context independent conceptualization of language use and interaction (Ortega, 2019). Translanguaging embraces the linguistic and cultural norms in given contexts, honors persons’ identities, and promotes flexible language use, (Canagarajah, 2015; Ortega, 2019).

In an article intended for language testers (Chalhoub-Deville, 2003), I have characterized L2 testing constructs in ways that fit with the two ends of the ML continuum depicted in the above table. Similar to the position at the left end of the continuum, representations of L2 testing constructs as cognitive, generic, fixed conceptualization, tend to promote a separation between language ability-in language user and language-user-in context. A more socio-cognitive L2 depiction, portrays constructs as fluid, co-constructed, and context-dependent. This position, which can be said to represent the right end of the continuum embraces the interaction of language and context as one inseparable entity, i.e., language ability-in language user-in context. I provide more on these positions later in the article.

Multilingual assessment orientation

Researchers such as Schissel et al. (2017) call for assessments that embrace “fluid and dynamic language practices ... Weaving together theories and pedagogies of translanguaging and translanguing practices with formative assessment practices and task-based language assessment,” (p. 1). Saville (2017) makes the case for learning oriented assessment (LOA) to better implement plurilinguistic policies and promote “plurilingual skills needed in contemporary society” (p.1). Limerick (2017) details the distorted environment of language testing for educators in Ecuador as the system struggles with the diglossic Kichwa system. Fundamentally, such authors question the tenets our traditional language tests, which have paid little to no attention to ML constructs. The ML realities of the world and the state of knowledge in applied linguistics compel us to reconsider our testing practices and prompt engagement with ML constructs features especially as depicted on the right-hand side of Table 1. Since the early 2000s, I have been making the case for an alternative paradigm of L2 test operations (e.g., Chalhoub-Deville, 2003, 2009, 2016; Deville & Chalhoub-Deville, 2006; Martin, Patton, & Chalhoub-Deville, 2017). I believe that this alternative paradigm can better accommodate ML constructs.

Table 2 contrasts features embedded in traditional testing practices with those in a proposed alternative paradigm. The elements included in the table are key to the argument that we need to

reexamine the assumptions we hold with regard to language use constructs, including ML constructs as well as the design and development of integrated ML assessments.

Table 2. Prevailing paradigm versus recent thinking in language testing.

Established tests of L2 use in context	Tests of integrated L2 interactions
<ul style="list-style-type: none"> • Monolingual/siloed multilingual constructs 	<ul style="list-style-type: none"> • Integrated multilingual constructs
<ul style="list-style-type: none"> • Generalized performance 	<ul style="list-style-type: none"> • Emic orientation
<ul style="list-style-type: none"> • Static, fixed language use 	<ul style="list-style-type: none"> • Fluid language communication
<ul style="list-style-type: none"> • Stand-alone performance 	<ul style="list-style-type: none"> • Co-constructed interactions
<ul style="list-style-type: none"> • Language ability-in language user and language-user-in context 	<ul style="list-style-type: none"> • Language ability-in language user-in context
<ul style="list-style-type: none"> • General theory 	<ul style="list-style-type: none"> • Local theories
<ul style="list-style-type: none"> • Task is generic 	<ul style="list-style-type: none"> • Task is integral to construct
<ul style="list-style-type: none"> • Assessment of learning 	<ul style="list-style-type: none"> • Learning-oriented assessment
<ul style="list-style-type: none"> • Sustaining innovations 	<ul style="list-style-type: none"> • Disruptive innovations

The Table is adapted from Chalhoub-Deville (2009, p. 254) *Table 11.2. Differences in perspectives on context representation*. The domains solicited to formulate this table include language testing, applied linguistics, communication, education, and psychology (e.g., Erickson & Shultz, 1981; Kramsch, 1986; McNamara, 1997; Resnick, 1994; Snow, 1994; Swain, 2001; Young, 2000). Established tests, the left-side of Table 2, work with monolingual constructs and favor a compartmentalized view of languages. In this paradigm language use in context is portrayed as a static or fixed entity to make possible inferences of generalized performances. The alternative paradigm upholds integrated constructs where languages are integrated into a coherent repertoire. The paradigm featured on the right-side of the Table 2, depicts fluid, emic communication where interactions are bound by specific, local sociocultural norms. Language testing professionals have tended to favor general, cognitive models that represent the L2 construct across situations. Also general, cognitive models uphold the concept of individualized abilities, which enable us to award individual scores. The alternative approach presented, advances socio-cognitive views, which posit interactions as bound to and driven by given contexts.

In the early 2000s, (Chalhoub-Deville, 2003), I presented the concept of ability-in language-user-in context to denote that “the ability components the language user brings to the situation or context interact with situational facets to change those facets as well as to be changed by them. The facet aspects of the context the language user attends to dynamically influences the ability features activated and vice versa” (Chalhoub-Deville, 2003, p. 372). Such characterization is contrary to the prevailing cognitive view in language testing, represented in Bachman’s Communicative Language Ability (CLA) model (Bachman, 1990; Bachman & Palmer, 1996, 2010). The CLA model acknowledges the importance of context and interaction yet separates language ability-in language-user from language-user-in context. A sociocognitive orientation calls for the elaboration of local theories that detail specific constructs in given contexts of use, and the awarding of joint scores (Chalhoub-Deville, 2003). This approach involves diligent documentation of how particular language interactions take place within specific socio-cultural settings from the perspective of those involved. Such documentation considers participants’ negotiation of identities and cultural heritage and the entanglement of internal resources and processes with tasks at hand.

In language testing, a task (an item, or a question) is traditionally viewed as a generic test method, intended to elicit construct attributes. A test method is characterized broadly, including in educational measurement literature, as a *necessary evil* to access targeted construct features (see Chalhoub-Deville, 2009, 2016; Deville & Chalhoub-Deville, 2006). This characterization is supported by established and widely-used multitrait-multimethod (MTMM) investigations (Campbell & Fisk, 1959). Test method/task specific variance in MTMM is considered as error or construct-irrelevant (See Bachman, 1990). Tasks are essentially a tool to access construct attributes and not considered integral to the participants’ experiences or to the L2 construct. Such portrayal of tasks is contrasted with a contention that ability and task are a single interacting structure, which denotes an L2 construct. An example of this person and task interaction is depicted in game-based assessment (GBA) (Lay, Patton, & Chalhoub-Deville, 2017) where the L2 construct is an entanglement of the test taker and the task.

Closely associated with this single, dynamic task perspective is learning-oriented assessment (LOA) where assessment data is used to support students in their learning and language development. For example, in GBA, players are enticed to engage in the game, scaffolded while exploring, routed according to their levels, provided with immediate and tailored feedback that facilitates progress to the next level, and presented with challenges to develop their abilities and skills. GBA can support complex interactions, facilitate collaborations, and promote extended time on task. Established assessments, especially large-scale tests, on the other hand, favor assessment of learning (AOL). AOL assessment differs markedly in its intent and processes from LOA. As Hamp-Lyons (2017) explains, with AOL student scaffolding is not permitted, assessment is authority-based with a judgement-focus, feedback is delayed or not provided, and a feed-forward to serve learning is typically absent. With AOL, students tend to work independently with rigid types of interactions and short task engagement.

Features proposed in the alternative test paradigm, on the right side of Table 2, help us push open scholarly boundaries to allow new constructs such as integrated MLs into our operations. Not all ML advances, however, are of comparable orientation. In an earlier article (Chalhoub-Deville, 2010), I have differentiated between *sustaining* versus *disruptive* technology to better

understand the nature of innovations computer-based tests introduced into our field. *Sustaining* innovations refer to advances that help simplify operations, enhance access to products and services, improve testing experiences, and enrich the measurement of construct features, among other things. Changes that amend our decidedly monolingual-based L2 assessment operations can be characterized as *sustaining* knowledge innovations. While these sustaining innovations are important and worthy of commendation, they do not alter the field at the core. *Disruptive* innovations denote fundamental knowledge transformations especially to construct representation. An integrated ML construct, which emphasizes features such as plurilinguistic communication and translanguaging requires that we transform our value systems, rethink test takers' experiences, reconsider L2 constructs, and overhaul testing operations.

In conclusion, the prevailing test paradigm has enabled us to develop the language testing field over many decades. We need to remain critical and be guided by new thinking and innovations to continue to advance and accommodate the changing linguistic landscape in the world. In the next section, I examine the nature of ML innovations discussed by language testing researchers. I focus on three papers, Saville, Schissel et al., and Limerick, presented at the 2017 LTRC ML symposium and appear in the present issue. These papers represent pioneering ML assessment explorations in a variety of domains – policy, achievement, and certification testing.

Multilingual assessments: sustaining versus disruptive innovations

The papers, Saville, Schissel et al., and Limerick underscore a representation to the ML construct, which can decidedly be characterized as a *disruptive* innovation. Integrated ML constructs require that we transform our theories and practices to accommodate features such as plurilinguistic interactions, translanguaging, non-standard language varieties. Such transformations are challenging but needed. However, a number of the other innovations these three papers call for can be more easily considered within our current operations.

In his paper, Saville (2017) points to the ML communities in Europe and the policy landscape that seeks to promote plurilingualism. The term plurilingualism refers to individuals with linguistic repertoires of multiple languages, which are deployed in given cultural contexts. Saville references the need for multilingual agendas that promote the concept of linguistic repertoires. “The concept of linguistic repertoires is especially useful since it does not focus on mastering isolated languages, but rather on developing different levels of ability in a number of languages which are used in different spheres of life and for different purposes” (Saville, 2017, p. 6). Saville maintains that assessment, including large-scale assessment, is at odds with goals to develop students' plurilingual abilities. Saville adds that LOA can be used “to address this problem” (Saville, 2017, p. 1). LOA, he argues, needs to be the centerpiece of not only classroom-based but also large-scale language assessment. “LOA aims to ensure that the large-scale assessments impact positively on learning and on educational systems in the local contexts where they are used” (Saville, 2017, p. 2). In summary, while Saville's notion of ML represents a *disruptive* innovation, his push for LOA can be characterized more as *sustaining* innovation primarily because the call for LOA is not unique to ML assessment. Learning-based assessment and the challenges they introduce, for example in terms of validity, are increasingly discussed in mainstream assessment as well (Moss, 2016). Such *sustaining* innovations advance theory and practice but do not reconstitute our L2 construct outlook.

Positioning assessments to serve learning is a challenge that permeates, not only ML-related assessments, but all testing practices, including accountability testing. For example, in the US, educational testing systems are increasingly required to link assessment to learning. A critique of these testing systems, while informative, is not central to the present argument (Those interested in the topic, are referred to Chalhoub-Deville, 2016). Of relevance is that monolingual initiatives such as Race to the Top (see U.S. Department of Education, 2009) and Every Student Succeeds (Darling-Hammond et al., 2016). Act require test publishers to explicitly attend to learning and growth. Thus the call by Saville (2017) to locate “learning at the heart of all assessment contexts – from classroom tasks to progress tests ... to high stakes exams” (Saville, 2017, p. 2) illustrates common grounds for ML and monolingual test developers. LOA is worthy of consideration in ML and in mainstream testing operations. The bottom line is that the challenges of adopting LOA in ML testing theory and practice are shared with monolingual testing systems. Such wide support for LOA can engender critical changes in the language testing field. However, such changes, while very challenging, do not in and of themselves engender ML *disruptive* thinking. In addition to embracing LOA, we need to continue to push for innovations at the construct level. The shift in value systems and constructs needed to embrace ML communities and plurilingualism is not necessarily an outcome of LOA adoption. LOA is an important but not sufficient condition for promoting ML assessments. Language Testing professionals need, as Saville himself articulates, to push for policies at the macro and micro levels that promote ML communities and advance plurilingual capabilities.

Similar to Saville (2017), Schissel et al. (2017) promote ML assessments that advance a *disruptive* representation of the construct and push for incorporating multilingual resources as part of testing. The authors draw on translanguaging pedagogies to formulate their ML arguments. The Schissel et al. research study underscores assessments, which closely align with learning. They anchor their research in classroom-based assessment and utilize ML task-based language assessment (ML-TBLA) tasks. Methodologically, the researchers presented students with reading materials in English and Spanish and students provided their responses in English. Such ML-TBLAs introduce innovations that loosen the confines of a monolingual construct in terms of access to information and represents *disruptive* language testing innovation. Such innovation is not carried into responses and scales. The responses/data captured for analysis does not allow students to utilize their ML repertoires. Additionally, the scales employed to analyze test takers’ ML-TBLA performance focus on traditional language features, e.g., lexical knowledge in one language. These features are adopted to maintain alignment with instruction. Alignment between assessment and learning/instruction, as stated earlier with Saville, is an important, *sustaining* form of innovation. Moving forward, it is important to explore TBLAs that embrace integrated ML in all aspects of design and development, including student responses and rating scales.

The ML-TBLA design in the Schissel et al. research study also shares concerns raised with monolingual integrated test tasks. For example, the plagiarism concerns expressed in the Schissel et al. study (2017) because of the close transfer of information from the reading texts to the writing task is an integrated task design issue deliberated by the English only TOEFL iBT developers. The plagiarism concern is a key reason why TOEFL iBT avoids this specific reading-writing modality sequence. Listening always precedes the written/verbal output to

minimize content transfer. The point here is that monolingual and ML test tasks share challenges and solutions/knowledge accumulated can transfer from one area to another. We do not need to discuss ML theory and practice in isolation. We can consider relevant lessons learned with traditional assessments. In conclusion, the Schissel et al. ML-TBLA study challenges current testing practices that shy away from ML constructs. The authors call for *disruptive* innovation in terms of embracing “multilingualism and dynamic translanguaging practices ... [that] include ... Indigenous languages” (p. 7). Their work also points to *sustaining* innovations, e.g., how current language testing operations can accommodate ML-TBLA to move beyond strict adherence to monolingual constructs and learning-testing alignment.

The Limerick (2017) paper explores the complexities of language policies in Ecuador, which mandate that educators pass exams to document their speaking abilities in Spanish and an Indigenous language. Limerick, similar to the Saville (2017) and Schissel et al. (2017) papers, assert the need for the language testing community to push for a *disruptive* approach to constructs. He expresses concern about language constructs that are “pure, bounded, and stable, especially in the form of a single standardized variety” (p. 3). Limerick’s study explicates how individuals mix standard and everyday language varieties to access knowledge, make meaning, mediate social identities, advance their careers, etc., which requires that we fundamentally transform the constructs depicted in our language testing models and commonly assessed in our language testing programs. Limerick’s descriptive account of the diglossic Quechua language use in Ecuador highlights the type of messy but necessary engagement with ML constructs. He maintains that sidestepping this messiness, for example by choosing some language variety to develop Kichwa tests and/or mandating ML ability across modalities, can serve to sabotage policy goals intended to revitalize a language and empower traditionally marginalized individuals. Limerick’s paper is a call to engage with ML constructs that emphasize the importance of understanding how translanguaging takes place, who is involved, and in what modality.

Limerick also questions the notion of fixed expectations and “standards” in exams. He writes: “That the exam changes depending on the number of people who arrive is helpful, as it allows the exam to incorporate more Andean notions of working and traveling together into the context of giving the exam” (p. 11). While the other two papers discussed call for a learning-testing alignment, which is commonly expressed by now in the field, this paper goes further and calls for alignment with local sociocultural norms. The development of local theories that include not only localized ML construct delineation but also flexible test administration practices is a challenge and calls for *disruptive* thinking and research. An issue to consider in this regard is how to balance fairness issues that jeopardize score interpretation and use with expected flexible administrative practices. Scholars need to engage to address such challenges.

Conclusion

Integrated ML constructs disrupt established construct representation and prevailing assessment paradigm. A dynamic, integrated ML construct engagement presents serious challenges, for example, in terms of how to design tasks that invite test takers to utilize their full linguistic repertoires and how to score performances that include translanguaging. Such challenges are not easily accommodated in our prevailing assessment paradigm. “Indeed, multiple challenged (sp.)

need to be addressed before multilingual/translanguage tests can be designed and used” (Shohamy, 2017). However, the field cannot ignore the ML realities of today’s world. We have to advance our knowledge base to accommodate real-world discursive practices. Research to pursue includes questions such as:

- What frameworks/models are useful to operationalize an integrated ML construct for assessment purposes?
- What does the integrated ML construct look like in different subject areas such as math, history, English language arts, etc.?
- How do we determine what is appropriate translanguaging at different ML proficiency levels?
- What are salient ML ability and translanguaging features when interacting with specific language assessment tasks?
- What should language test scales focus on in an ML assessment beyond traditional language components?
- What should our expectations entail for test takers from different age groups when engaging in integrated ML assessments? Are we to expect different patterns of translanguaging for different age groups?
- Who should be engaged as raters, i.e., ML qualifications, and what type of rater training is needed?
- What validity frameworks are useful to accommodate integrated ML constructs?

Given the nature of assessment, it is understandable that the field would lag behind other disciplines. However, we have to meaningfully engage. As indicated earlier, not all ML considerations will require overhaul of our L2 assessment systems. The issues discussed point to areas where current practices can be expanded/rethought to accommodate *sustaining* ML changes. Changes can be introduced while we develop the scholarly infrastructure that facilitates integrated ML construct assessments.

Disclosure statement

No potential conflict of interest was reported by the author.

References

Bachman, L. (1990). *Language testing in practice*. Oxford, UK: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford university press.

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

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.

Canagarajah, S. (2015). Clarifying the relationship between translanguaging practice and L2 writing: Addressing learner identities. Based on the invited plenary at the 2014 Maryland conference on academic and professional writing, University of Maryland, October 10-11, 2014. *Published in the Applied Linguistics Review*, 2015(6), 415–440. doi:10.1515/applirev-2015-0020

Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language Testing*, 20, 369–383. doi:10.1191/0265532203lt264oa

Chalhoub-Deville, M. (2009). Content validity considerations in language testing contexts. In R. Lissitz (Ed.), *The concept of validity* (pp. 241–263). Charlotte, NC: Information Age Publishing.

Chalhoub-Deville, M. (2010). Technology in standardized language assessments. In R. Kaplan (Ed.), *Oxford handbook of applied linguistics* (2nd ed. pp. 511–538). Oxford, UK: Oxford University Press.

Chalhoub-Deville, M. (2016). Validity theory: Reform policies, accountability testing, and consequences. *Language Testing*, 33, 453–472. doi:10.1177/0265532215593312

Darling-Hammond, L., Bae, S., Cook-Harvey, C. M., Lam, L., Mercer, C., Podolsky, A., & Stosich, E. L. (2016). *Pathways to new accountability through the Every Student Succeeds Act*. Palo Alto, CA: Learning Policy Institute.

Deville, C., & Chalhoub-Deville, M. (2006). Test score variability: Implications for reliability and validity. In M. Chalhoub-Deville, C. Chapelle, & P. Duff (Eds.), *Inference and generalizability in applied linguistics: Multiple research perspectives* (pp. 9–25). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Erickson, F., & Shultz, J. (1981). When is a context? Some issues in the analysis of social competence. In J. Green & C. Wallat (Eds.), *Ethnography and language in educational settings* (pp. 147–160). Norwood, NJ: Ablex.

García, O. (2009). *Bilingual education in the 21st century: A global perspective*. Malden, MA: John Wiley & Sons.

Garcia, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism, and education*. London, UK: Palgrave Macmillan.

Gonzalez, L. M., & He, Y. (2017). *Engaging with diverse language communities: Models and approaches*. Coalition for Diverse Language Communities Working Papers. Greensboro: University of North Carolina at Greensboro.

Hamp-Lyons, L. (2017). Language assessment literacy for learning-oriented language assessment. *Papers in Language Testing and Assessment*, 6, 88–111.

- Kramersch, C. (1986). From language proficiency to interactional competence. *The Modern Language Journal*, 70, 366–372. doi:10.1111/modl.1986.70.issue-4
- Limerick, N. (2017). *Indigenous language revitalization and the difficulties of testing native Kichwa-speakers in Ecuador*. Paper presented at the Language Testing Research Colloquium (LTRC), Bogota, Colombia.
- MacSwan, J. (2017). A multilingual perspective on translanguaging. *American Educational Research Journal*, 54, 167–201. doi:10.3102/0002831216683935
- Martin, A., Patton, E., & Chalhoub-Deville, M. (2017). A case for the use of the ability-in language user-in context orientation in game based assessment: Language testing and measurement considerations. *Language Testing Asia*, 7, 16. doi:10.1186/s40468-017-0045-0
- McNamara, T. (1997). ‘Interaction’ in second language performance assessment: Whose performance? *Applied Linguistic*, 18, 446–466. doi:10.1093/applin/18.4.446
- Moll, L. C. (2017). Foreword. In S. Spencer & P. Portes (Eds.), *US latinization: Education and the new Latino South* (pp. xi–xiv). Albany: State University of New York Press.
- Moss, P. (2016). Shifting the focus of validity for test use. *Assessment in Education: Principles, Policy & Practice*, 32, 236–251.
- Ortega, L. (2019 *Supplement*). SLA and the study of equitable multilingualism. *The Modern Language Journal*, 103, 23–38. doi:10.1111/modl.12525
- Resnick, L. B. (1994). Situated rationalism: Biological and social preparation for learning. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 474–493). Cambridge, UK: Cambridge University Press.
- Rojas, V. (2014). “Translanguaging”: A CUNY-NYSIEB guide for educators. Retrieved from https://cpb-us-w2.wpmucdn.com/share.sis.org.cn/dist/9/608/files/2014/11/TRANSLANGUAGING-PPT-14_2-zbydj9.pdf
- Saville, N. (2017). *How can multilingualism be supported through language education in Europe?* Paper presented at the Language Testing Research Colloquium (LTRC), Bogota, Colombia.
- Schissel, J., & Chalhoub-Deville, M. (2017). *The construct of multilingualism and language policies in language testing*. Symposium organizers. The Language Testing Research Colloquium (LTRC), Bogota, Colombia.
- Schissel, J., Leung, C., López-Gopar, M., & Davis, J. (2017). *Task-based translingual assessments: A framework for classroom-based language assessments for linguistically diverse*

communities. Paper presented at the Language Testing Research Colloquium (LTRC), Bogota, Colombia.

Shohamy, E. (2011). Assessing multilingual competencies: Adopting construct valid assessment policies. *The Modern Language Journal*, 95, 418–429. doi:10.1111/modl.2011.95.issue-3

Shohamy, E. (2017). *Towards an understanding of multilingual and translanguaging assessment*. Paper presented at the Language Testing Research Colloquium (LTRC), Bogota, Colombia.

Snow, R. E. (1994). Abilities in academic tasks. In R. J. Sternberg & R. K. Wagner (Eds.), *Mind in context: Interactionist perspectives on human intelligence* (pp. 3–37). Cambridge, UK: Cambridge University Press.

Swain, M. (2001). Examining dialogue: Another approach to content specification and to validating inferences drawn from test scores. *Language Testing*, 18, 275–302. doi:10.1177/026553220101800302

U.S. Department of Education. (2009). *Fact sheet: The race to the top*. Retrieved from www.whitehouse.gov/the-press-office/fact-sheet-race-top

Wei, L. (2018). Translanguaging as a practical theory of language. *Applied Linguistics*, 39, 9–30. doi:10.1093/applin/amx039

Young, R. (2000, March). *Interactional competence: Challenges for validity*. Paper presented at the annual meeting of the Language Testing Research Colloquium, Vancouver, Canada.