Afterword: Meeting the Challenges of Workplace English Communication in the 21ˢᵗ Century

David H. Slomp, Editor, Special Issue, *University of Lethbridge*

María Elena Oliveri, Editor, Special Issue, *Buros Testing Center, University of Nebraska-Lincoln*

Norbert Elliot, Editor, Special Issue, *University of South Florida*

1.0 Introduction

When we began the work of this special issue, we were motivated by a concern for the consequences stemming from the use of contemporary assessments to meet the challenges of Workplace English Communication (WEC) in the 21ˢᵗ century. We were equally motivated by the potential and promise that comes from envisioning new assessment designs that enhance fairness and opportunity to learn. In this afterword, we look across this special issue to draw out the lessons learned from researchers and scholars involved in designing, using, and interpreting evidence from assessments of complex WEC tasks.
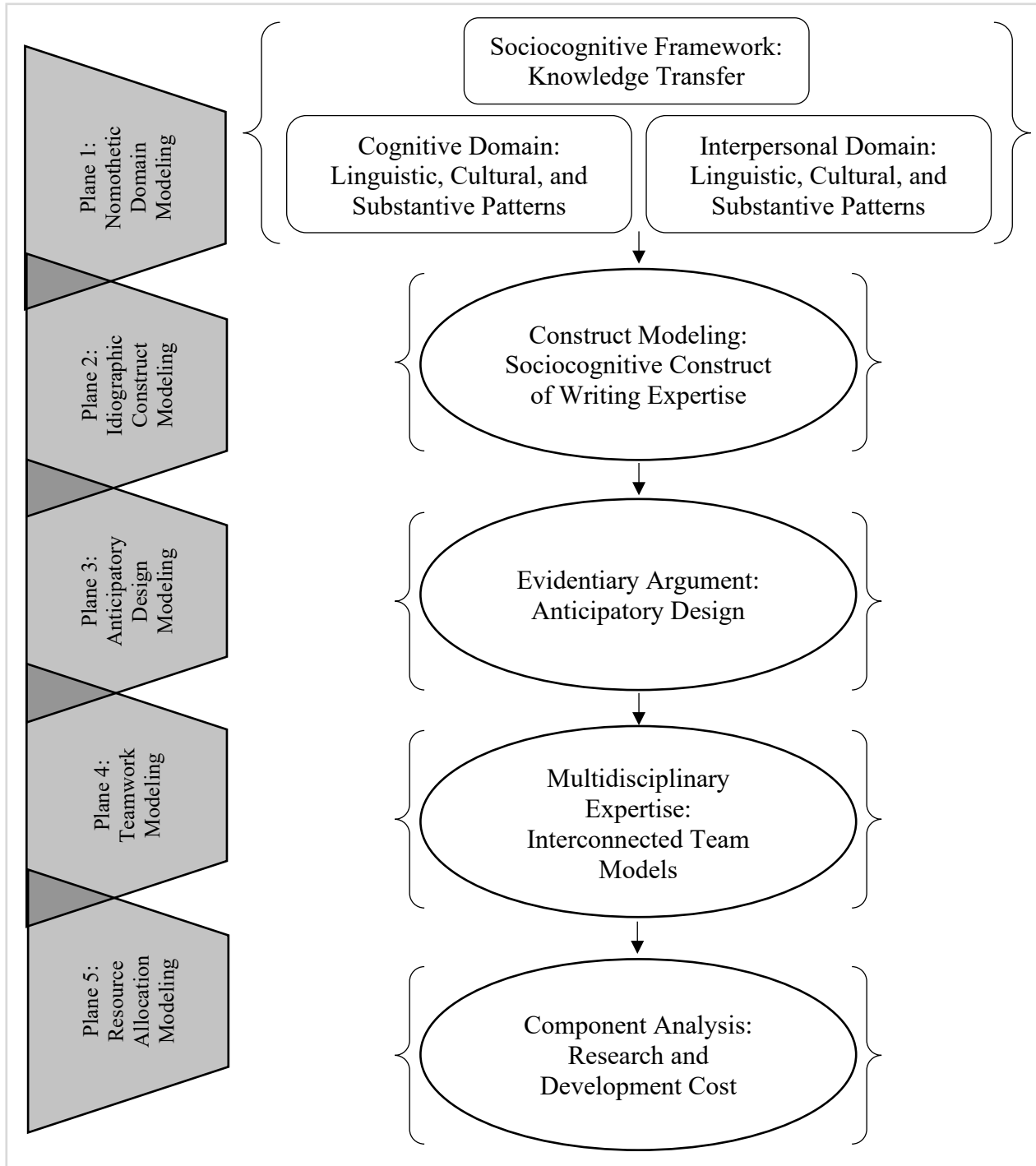
Along these lines, the reflections in this afterword are drawn from a slow, deliberate review of the final versions of each article contained in the special issue. Key observations, ideas, and insights from each article were highlighted and entered into a mind mapping platform by the first author. Coded by article of origin, each insight was linked thematically to other insights from within and across SI articles. Insights were then organized by themes. Once organized, coding was shared with the second and third authors for commentary and discussion. This discussion was used to create the afterword.

In the introduction to this special issue, we began with a general conceptual model of WEC described in §5.0 (Oliveri, Slomp, Elliot, et al., 2021). After the review of the special issue articles described above, we were able to create a conceptual model of lessons learned shown in Figure 1 below. The lessons are organized by five planes. Each of the planes constitute areas of deliberation on topics that are germane to Assessment *for* Learning (A*f*L) applied to the special case of WEC. A*f*L is a process where teachers use assessment information to adjust their

teaching strategies and students use information from assessments to inform their learning approaches. As such, assessment, teaching, and learning are viewed as tightly knit and complementary, as one informs the others.

Figure 1

Complex Tasks of Hard to Assess 21st Century Communication Skills: Lessons Learned from the Design of Workplace English Communication

Plane 1: Nomothetic Domain Modeling

Sociocognitive Framework: Knowledge Transfer

Cognitive Domain: Linguistic, Cultural, and Substantive Patterns

Interpersonal Domain: Linguistic, Cultural, and Substantive Patterns

Plane 2: Idiographic Construct Modeling

Construct Modeling: Sociocognitive Construct of Writing Expertise

Plane 3: Anticipatory Design Modeling

Evidentiary Argument: Anticipatory Design

Plane 4: Teamwork Modeling

Multidisciplinary Expertise: Interconnected Team Models

Plane 5: Resource Allocation Modeling

Component Analysis: Research and Development Cost

In considering Plane 1, the special issue has taught us that authors do indeed adopt a sociocognitive framework for domain analysis, domain modeling, and score interpretation, among other assessment development and assessment use practices. The sociocognitive framework encompasses cognitive and interpersonal domains, with analysis undertaken with emphasis on the linguistic, cultural, and substantiative (LCS) patterns of WEC. Its use is timely given the high levels of population (student) diversity in today's classrooms in the U.S. and the world over. Failing to adopt a sociocognitive approach can lead to students feeling marginalized or result in reduced opportunities to learn, as was described in the introduction and within other articles in the special issue. In §2.0 of this afterword, we pay particular attention to the application of the sociocognitive framework as it is related to transfer of knowledge, skills, and other attributes (KSAs) across academic and non-academic settings, with a focus on LCS patterns that can be used to target elements of transfer across settings.

Reflecting on Plane 2 considerations, we learned that construct-modeling activities are deeply situated (i.e., context-dependent) and that, in the case of WEC, authors in this special issue attended closely to the context in which communications occur while reflecting on the impact of instructional and assessment consequences. In §4.0, we examine what it means to design educational interventions with consequences in the foreground. As well, we found that authors focus on writing expertise as they begin the draw-down process from nomothetic to idiographic modeling (ranging from more general to more specific forms of construct modeling, respectively).

With regards to Plane 3, we learned that evidentiary arguments are realized in anticipatory, principled design practices—the rubber-meets-road practice of forming interpretation and use arguments (IUA). In §5.0, we describe the ways that our authors fleshed out mechanisms for systematically collecting and examining evidence while linking those findings to anticipated IUAs. We also learned that our authors noted that these design practices and the assessment processes rising from them are to be accompanied by new WEC pedagogies. Formative in nature as part of AƒL, we describe the need for these pedagogies in §3.0.

In considering Plane 4, we learned that assessment design activities requiring collaborative efforts did not necessarily fall along the lines of distributed leadership (or a top-down leader approach) but rather benefitted from a more interconnected team model wherein experts from various disciplines contributing to the project were able to work together contributing from their own disciplinary perspectives. Rather, as we elaborate in §4.2, multidisciplinary collaboration can be more effectively organized around principled design frameworks and the associated use of specific artifacts (e.g., construct models and Q-matrices) used within design frameworks as focal points. These design artifacts and psychometric models are identified in §4.1.

In Plane 5, we provide details on the costs of developing the Kitchen Design (KD) prototype embodying the operationalized WEC model. (For more on this prototype, see §3.0 of the introduction). In the special issue articles, surprisingly little attention was paid to the costs of

developing WEC interventions. The actual cost of $750,000, defined in detail in §6.0, reveals the need for the detailed planning identified in Planes 1 to 4.

Section 7.0 of this afterword demonstrates the usefulness of Figure 1. There, we discuss four projects that are currently underway based on the design of a digitally based, formative assessment program for teaching WEC.

As we did in our introduction, we conclude this afterword (§8.0) with thanks to those who made this special issue possible. Our hope is that this afterword will distill for our readers a set of frameworks and insights that will motivate and guide future iterations of WEC research.

2.0 Implications of Using a Sociocognitive Lens

Using frameworks of 21st century communication skills, many contributions to this special issue focus on the challenge of designing assessments that explicitly support the transfer of KSAs across contexts. This emphasis on transfer is a marked departure from traditional models of large-scale writing assessment that are often drop-from-the-sky and disconnected from the pedagogical goals of the classroom. Traditionally, writing assessments have focused on measuring written products and making inferences from those products to the underlying expertise that shaped them. When the focus shifts away from measuring products toward supporting transfer, the nature of the assessment needs to change as well. As noted in the introduction to this special issue, A*f*L, with its emphasis on augmenting opportunity to learn and utility to inform integrated teaching and assessment synergies, is well suited to enable and inform these new directions.

2.1 Application of a Sociocognitive Approach to Support the Transfer of Learning

Why is there a need for assessment systems and assessment designs that focus on transfer?

The 21st century workplace is marked by flux and transience (International Organization for Migration, 2019). To thrive in this environment, people need to develop the capacity to adapt to new roles, contexts, and tasks on an ongoing basis. Accordingly, our assessments need to both reflect this new reality, and they need to measure the constructs that underpin this capacity.

Key to this necessary adaptability is the capacity for transfer. A number of the articles in this special issue pay particular attention to examining the mechanisms associated with transfer. Authors observe that the capacity to perceive critical features and parallels between situations is foundational to transfer. Applied to the context of workplace communications, this means that students need to learn how to identify parallels and critical differences between communicative situations, and to leverage that analysis to inform the choices they are making in new situations. Haigler (2021, this issue) points to three factors that facilitate transfer: knowledge of how a discourse community (or community of practice) functions, participants' self-identity, and participants' gradual integration into the targeted discourse community. Key to understanding how a discourse community works is understanding the community's communication purposes, values, and associated expectations (Corrigan & Slomp, 2021, this issue). With integration into

that community—including its social and linguistic practices and LCS patterns—students develop richer understandings of those values, expectations, and purposes. Significantly, students then have the capacity to shape communications practices within that community.

Viewed through a sociocognitive lens, transfer involves developing an understanding of the LCS patterns that shape communications within and between groups. These patterns reflect the values and purposes of discourse communities. Understanding LCS patterns and how they are similar and different across groups is key to supporting transfer (Oliveri, Mislevy, & Slomp, 2021, this issue).

LCS patterns develop over time and across a range of contexts. With respect to language use, some of these patterns are more universal than others. Conventions around sentence structure and punctuation use, for example, are relatively stable and universal. However, variations on these patterns arise within different cultural, racial, or linguistic communities, and within different communication genres. Punctuation use in texting, for example, follows a different set of conventions and patterns than would be expected in academic writing, while punctuation in emails might share features of use in both contexts.

Through experience, we gain access to these patterns, learning to apply them to the texts we create. The challenge of language learning is to develop the capacity to identify and internalize these patterns—both universal and context-specific—to understand how these patterns relate to the contexts in which they are used, as well as to understand how to employ these patterns within those contexts. To meet these challenges, students need formative, timely feedback to help them understand how well they are acquiring, internalizing, and utilizing these patterns.

Focus on transfer has oriented the work in this special issue in three ways. First, focus on transfer has foregrounded the importance of metacognition and of the need to design instruction that supports growth in metacognitive capacity, as well as the need to develop assessments that provide formative assessment data that can guide teaching and learning of metacognition.

Second, the capacity to transfer knowledge highlights the need for new pedagogies that emphasize deeper learning involving metacognition and that provide students with opportunities to identify the LCS patterns that shape textual conventions and authors' choices within and across discourse communities. Wetzel et al. (2021, this issue) demonstrate a data-driven pedagogy that utilized writing analytics to render transparent the word- and sentence-level decisions writers make. The information provided by DocuScope Classroom—a computer-assisted rhetorical analysis tool—enables writers to understand how LCS patterns within and across discourse communities shape a writer's word- and sentence-level choices. This form of information can help writers better understand the values and expectations of the communities within and for which they are writing. At the same time, this pedagogy promotes metacognition as writers use text analytics to better understand the rhetorical situation in which they are writing, monitor, and critically review their own rhetorical choices as they craft text to deliver on their own rhetorical intents. In enacting such pedagogies, our authors note that attention needs to be paid to narrowing instructional gaps among the following: what is taught in the classroom, how

learning is guided in the classroom, and how learning occurs in the workplace. One of the challenges of learning in the workplace is that one must function independently, which exacerbates the criticality of being able to problem-solve and solve the communications tasks arising on a daily basis also independently. Therefore classroom instruction needs to focus on teaching students how to problem-solve communications tasks with lesser degrees of scaffolding as students move forward through their educational years. In this light, a problem-solving model of writing instruction is preferred. A problem-solving model provides students a metacognitive scaffold that principally emphasizes transfer and metacognitive awareness that empowers students to select how to vary discourse vis-à-vis the audiences and purposes for which communications are used. This model differs from traditional forms of writing instruction (e.g., academic writing) in which context and audiences do not extend beyond the classroom and the teacher.

Third, emphasis on transfer highlights the need for transfer-oriented assessments. Just as instruction needs to narrow the gap between the learning that occurs in the classroom and the learning that occurs in the workplace, assessments also need to better align with a workplace context. This realization drives the design of WEC toward the use of performance tasks embedded within authentic contexts and task-based scenarios. The goal of such new efforts is to develop scenario variants that help students explore different rhetorical purposes, structures, and situations. Informed by genre theory, this approach encourages students to experience comparative genre learning (Wentzel et al., 2021, this issue) that will prepare them for transfer.

Framing transfer through a sociocognitive lens carries implications for assessment design, analysis of performance data, and score reporting. Assessment tasks need to be carefully mapped to LCS patterns and their variations, and they need to be embedded in rich, authentic contexts so that they can provide information on how well test-takers understand these patterns and how appropriately they utilize these patterns within and across a range of contexts. When assessment design is approached this way, assessment designers and users are able to go beyond coding for the presence and absence of correct or incorrect (0, 1) responses so that we can track students' use of those patterns and report back to them how their performance is impacted by their use of those patterns (Oliveri, Mislevy, & Slomp, 2021, this issue). The emphasis on analysis of performance data and score reporting, therefore, is not so much on providing students with a summative score, but more importantly, on providing them with information that will help them understand strengths and weaknesses in language use, in their access of LCS patterns, and in their appropriate application of those patterns within specific and varied contexts. Score reports, too, necessarily need to be designed to provide test-takers with clear information regarding areas of strength and areas for growth with respect to their use of LCS patterns (Zapata-Rivera et al., 2021, this issue).

In this special issue, we have only begun to scratch the surface of what LCS patterns enable. As WEC prototypes become operational, and as large-scale tryouts of those assessments are conducted, opportunity to examine patterns of test-taker response can further clarify the role that

access to LCS patterns play in test-taker performance. As Oliveri, Mislevy, and Slomp (2021, this issue) observe, when designing assessments and tasks, it is important to determine which LCS patterns are linked to the learning targets (evidence of construct-related validity) and which patterns are linked to student characteristics (potential sources of construct-irrelevant variance). The line between the two can be thin, and, in the case of formative assessments, even the boundaries are new. Larger-scale analysis of performance data, especially data derived from items focusing on test-taker metacognition, will help to further elucidate how test-taker access to a range of LCS patterns supports success on WEC tasks. This analysis will also help to elucidate which LCS patterns can be linked to sources of construct-irrelevant variance that are not explicitly part of the learning targets.

2.2 Application of a Sociocognitive Approach to Digitally Informed Writing Pedagogies

Much work still needs to be done to better define the LCS patterns that support success in WEC. Writing analytics, as exemplified by Aull and Aull (2021, this issue) and Wetzel et al. (2021, this issue), can be a powerful tool for supporting this work. Writing analytics can be used to reveal patterns in email communications—such as politeness, style, and topics (Aull and Aull, 2021, this issue), as well as genre moves (Wetzel et al., 2021, this issue)—that can inform scoring criteria, feedback to learners, and instructional interventions to support learning.

In contrast to formulaic or prescriptive guidelines related to language learning, a focus on LCS patterns supports the goals of students learning to organize information in a manner that helps them organize new knowledge about language into meaningful patterns (Ambrose et al., 2010). Thus, instead of providing students prescriptive guidelines divorced from context (e.g., "avoid passive voice sentences" or "don't use run-on sentences"), a focus on LCS patterns provides students with a descriptive and contextualized text visualization environment to enable them to engage with the learning of language in deeper ways over time to engage with more complex and nuanced aspects of genres. Moreover, students are guided to self-assess their own composing decisions according to foundational principles of language use they learn in class through integrated learning and assessment tools described in this special issue. The use of the visualization tools and just-in-time feedback provided through the tools will support students' writing process in low-stakes environments through self- and peer-review activities.

More broadly, as Oliveri, Mislevy, and Slomp (2021, this issue) note, developing assessments of complex constructs requires a richer understanding of a targeted domain and learning principles associated with mastering it. Meeting the challenges of 21st century assessments involves aligning learning and assessment situations with complex scenario-based tasks that include opportunities to track students' performance. Such are the foundations for the kinds of ambitious assessments now needed. As part of this foundational work, analyzing both communications tasks and test-taker responses to those tasks through the lens of LCS patterns, as Wetzel et al. (2021, this issue) demonstrate, can provide richer understandings of the targeted domain and of student ability with respect to that domain. The title of Aull and Aull's (2021,

this issue) study, "*Write a Greeting for Your Email Here*," playfully points to the traditional limitations of pedagogy and assessment in the field of workplace communication, where historically the focus has been on surface features of a very complex genre. In contrast, our use of LCS patterns in the design of the WEC modules enables the design of more complex tasks, assessed through criteria that more closely connect to the underlying complexities inherent in workplace communications.

2.3 Application of a Sociocognitive Approach for Diverse Populations

Considering LCS patterns is also important to support the teaching and learning of culturally and linguistically diverse populations in meaningful ways. Mislevy and Duran (2014) highlight the importance of acknowledging broadly conceived social contexts to address linguistic and cultural patterns of behavior across subpopulations, as there may be knowledge and activity patterns that are unique to specific cultures or domains. Lewis (2006) highlights differences in discourse when conducting business transactions internationally across cross-cultural teams. For instance, differences may emerge with regards to values, with one culture valuing bottom-line results and expediency and other cultures valuing idealism, imagination, and a lesser focus on planning and details. Along this line, the use of LCS patterns may also vary, with the former providing straight-to-the-point information and the latter using LCS patterns that are less concise and more open-ended, for instance. Developing the competency to write for the different cultures would involve developing an understanding of variability in LCS pattern uses across the two cultures; alternatively, the differences in communication styles may lead to unintended consequences, such as delays in business transactions or misunderstandings across groups. As such, an understanding of LCS patterns and their variation are important when teaching and doing business with culturally and linguistically diverse groups (Oliveri, Lawless, & Mislevy, 2019).

Attention to LCS patterns is not only important when communicating with individuals and organizations in different countries, but it also has relevance when communicating with individuals within a single country in cases when there is within-country population heterogeneity. As an example, in the U.S., Baker-Bell (2020) highlights the importance of acknowledging, valuing, and normalizing linguistic variations (e.g., African American vernacular), which differ linguistically from mainstream English. The goal of understanding and using LCS patterns is thus to equip learners with an understanding of linguistic variations and enable learners to be in the driver's seat able to determine which linguistic features to use, how, and when. In the receptive end, the goal is to better understand the writer's intent, the linguistic choices writers made, and to be better equipped to interpret the written compositions.

Attentive to such ends, Oliveri and Ercikan (2011) conducted research to investigate the comparability of problem-solving items administered in the Programme for International Student Assessment (PISA) administered to students in French and English in Canada. Performance data from PISA are used to compare student achievement; inform curricula, program development, and evaluation; and make decisions concerning educational policies. There is an implicit

assumption that the assessments and items within the assessments are comparable across groups speaking different languages (English and French) and across versions of an assessment (English or French) administered across English- and French-speaking countries or regions within countries. The validity of score-based uses and inferences made based upon the results critically depends upon the items' comparability. However, prior research has shown that there are differences in test adaptation practices that lead to differences in the performance on test items across groups within countries and regions within countries because of dialectal differences within variants of English and French, for instance. These dialectal differences or differences in the use of LCS patterns need to be considered in the translation and adaptation of tests to make meaningful score-based differences. Along this line, we suggest that LCS patterns across groups need to be considered to meaningfully interpret LCS patterns across groups within and across countries.

3.0 Formative Assessment and Data-Driven Writing Pedagogies

The intentional use of LCS patterns to guide assessment planning, task design, and analysis of performance data described in the previous section highlights an important difference between formative and summative assessments. In their review of large-scale literacy assessment programs in Canada, Slomp, Corrigan, and Sugimoto (2014) found that there was a positive relationship between proximity of assessment use to classroom research/uses: the closer to the classroom the research was situated, the more negative its findings were with respect to consequences of assessment design and use. Conversely, the further from the classroom the research was situated, the more positive the findings were. Put differently, policy leaders and school district leaders were far more positive about the use of assessment data to guide teaching and learning than were classroom teachers and school leaders. This difference in perspective largely came down to perceptions of how useful assessment data were to guiding teaching and learning. Policy leaders and school district superintendents were very positive about using assessment data to guide resource allocation (e.g., they advanced that performance data could inform the level of support districts, schools, and classrooms received). On the other hand, classroom teachers often observed that the quality of information provided by the assessments was either poor or insufficient for providing meaningful information for informing instructional decisions. These findings are sobering. When those closest to the data, the test-takers, and the environment in which the data are generated and used are skeptical of the quality and utility of the information provided by the assessment, the enthusiasm toward the value of those data and their potential use by those more distant from the assessment is called into question. To address these issues, Oliveri, Mislevy, Elliot, et al. (2020) present a multilevel design model to minimize the unintended effects of test use when scores are used primarily (or uniquely) to inform summative decisions. The model considers both consequences of test use and the meaning of score-based interpretations to help reduce sources of construct-irrelevant variance and support valid score-based inferences when assessment data needs to be presented at different grain sizes.

Unlike summative assessment data, formative assessment data are primarily designed to provide information that can be used in the classroom. Consequently, formative assessment programs are more useful when they are closely aligned with the classroom and when the data they provide have a high degree of fidelity with classroom instruction. Such an approach requires more flexibility in design and use of the data in order to be used in different classrooms with learners with diverse background characteristics and learning styles and approaches. As we elaborate later, approaches such as the use of DocuScope has a more flexible design to enhance classroom integration.

In the case of WEC, minimizing stakeholders' distrust in the use of WEC data (e.g., employers, instructors, or students) would involve extensive planning and consultation with end users and experts in substantive aspects of the WEC construct and instruction, such as writing pedagogy and workplace communication. See §6.0 of the introduction for more information on technical and instructional challenges. Collaborative teams involving individuals with expertise in classroom instruction, the construct, and assessment design processes can work together to guide the assessment design process. When applied to assessment design and score reporting, experts can use a model such as the multilevel design model noted above to work together with end users and experts in writing and workplace communications to develop items, measurement models, and score reports to provide meaningful information to inform instruction and guide learning. Central to achieving trust in assessment and the use of evidence-based practices is to help ensure that the evidentiary-seeking practices and the tools used for assessment are closely coupled with addressing instructors' instructional needs. Various practices or approaches to integrate teaching and assessment and use AfL in the classroom exist. Examples include the use of scenario-based tasks, as was shown in various articles in the special issue, and the use of digital tools such as DocuScope, which is described in the next section.

3.1 Formative Assessment and Innovative Writing Pedagogies

Manuscripts across the special issue highlight the need to address the goal of improving classroom-based assessment practices and integrating the use of AfL in classrooms. For instance, these manuscripts highlight the relationship between innovative formative assessment design and the associated need for new writing pedagogies, and emphasize the importance of teaching for transfer (e.g., learning that goes beyond the classroom into workplace readiness and preparation by equipping students with 21st century skills needed for work and life, as noted in § 2.0).

Examining the issue of teaching for transfer through the lens of innovative digital technologies and LCS patterns creates opportunities to innovate from an instructional perspective. Such innovations are important with regards to the students taught (e.g., to acknowledge their diversity with respect to ways of learning and communicating), the subjects taught (e.g., with regards to expanding the linguistic elements students are asked to learn), and the tools they are provided (e.g., through the integration of digital technologies for learning). This expansion is important because not all students use LCS patterns in the same way; thus, the

use of more flexible assessment tools that integrate data visualization, which allows learners to view their own compositional choices, enables them to identify the extent to which their use of structure, style, and content is similar or different to the genre in which students are writing (e.g., workplace or academic writing).

Moreover, the use of LCS patterns in assessment and instruction helps to clarify central challenges at the heart of the transfer issue. Because our expectations related to structure, style, and content of any given text are mediated by LCS patterns, the capacity to access and make sense of the LCS patterns that govern a specific rhetorical context can be seen as a key to transfer. As Oliveri, Slomp, Rupp, and Mislevy (2021a, this issue) observe, "It is not *knowledge* that most productively transfers, but, rather, it is the metacognitive processes around *knowing*—the ability to perceive critical features and parallels between situations—that best transfer" (p. 81). It logically follows, then, that a writing pedagogy supporting transfer will focus primarily on teaching students how to analyze rhetorical situations, how to draw parallels between those situations, and how to identify critical features that reflect both similarities and critical differences between situations. Such is the power of DocuScope, which enables students in classrooms to autonomously evaluate their own compositional choices and self-modify texts to better approximate them to a desired model of effective composition within specific genres.

Aull and Aull (2021, this issue) demonstrate the power of corpus analytics to reveal patterns across contexts and within complex genres. While Wetzel et al. (2021, this issue) demonstrate the application of these analytic techniques to help foster metacognitive awareness and growth in students, their work also demonstrates the power of formative assessment to support growth and development of student writers. As student writing is processed using the same analytic techniques applied to a corpus of similar texts, students are able to use the feedback from this analysis to critically reflect on their own authorial choices, making decisions to revise their texts to achieve their rhetorical aims within the confines of the rhetorical context in which they are working.

While the focus of DocuScope Classroom is on micro-level choices and patterns, the analysis this tool enables can be used to open up conversations with students related to the values and expectations of the discourse communities in which students are learning to write; the variations within genres brought about by differences in values and expectations of different communities; and, the differences in how content is handled across these communities. This approach may help facilitate the use of a writing-to-learn approach that blends traditional, declarative instruction on core rhetorical genre principles relevant to writing with a hands-on portion for students to visualize their composing decisions in constructed-response assignments (Connolly & Vilardi, 1989; Fry & Villagomez, 2012). Pedagogies advancing such conversations can help students develop the critical awareness about differences, similarities, and patterns across rhetorical contexts that are key to transfer.

4.0 Designing with Consequences in the Foreground

Authors in this special issue envisioned the teaching and assessment of WEC from a formative assessment perspective. This design choice foregrounds the necessity of attending, from the outset of the design process, to the consequences of assessment design and use. Every design choice needs to advance the goal of providing information to students and teachers that will drive learning forward. So, while consideration of evidence related to validity, reliability, and fairness remains necessary and important to the design of these assessments, alignment with learning needs, and the instructional practices that support them, becomes even more important. In A*f*L, documentation of alignment is an important new source of evidence to be used in IUAs.

To inform their design, authors in this special issue identified principled design frameworks that would focus attention on a number of key design considerations including the importance of developing a robust understanding of the targeted construct; ensuring effective construct sampling across assessment tasks; and, alignment between the types of performances required by those tasks and the classroom environment in which learning is to occur and the workplace environment in which this learning will be applied. Authors also identified principled design frameworks that helped identify the types of information that needed to be collected and the measurement and analysis models necessary for transforming that performance data into actionable information. Finally, authors identified frameworks that would challenge us to be critically aware of the link between assessment design features, stakeholder actions, and the intended outcomes that motivated the design of the assessment program. While there are many promising models, special issue authors settled on three complementary principled design (anticipatory) frameworks: expanded Evidence Centered Design (e-ECD), Theory of Action (ToA; Oliveri, Slomp, Rupp, & Mislevy, 2021a), and Integrated Design and Appraisal Framework (IDAF; Oliveri, Slomp, Rupp, & Mislevy, 2021b).

The precursor to all of these models is Evidence-Centered Design (ECD), which Mislevy, Almond, and Lukas (2003) describe as a framework that applies evidentiary reasoning to the design of assessment programs. The purpose of the framework is to ensure alignment between the assessment argument, the construct being measured, the tasks being designed, and the scoring and reporting of evidence collected. The e-ECD framework extends the focus of ECD on clearly articulating the links between measurement models, assessment tasks, and construct elements by adding three considerations to design challenges: learning over time, the data collection opportunities connected to interactive and digital instructional content, and the need for measurement models that incorporate possibilities for learning over time alongside the complexities of digital instruction and data features (Arieli-Attali et al., 2019). Its integrated focus on these elements is designed to support meaningful IUAs of performance data for the purpose of supporting teaching and learning within an A*f*L perspective. From a sociocognitive perspective, the expanded ECD model described by Mislevy and Oliveri (2019) and Oliveri, Lawless, and Mislevy (2019) outlines ECD considerations for the assessment of diverse populations.

The ToA similarly compels assessment designers to articulate the path that leads from features of an assessment's design to the intended intermediate and long-term outcomes the assessment was designed to achieve. This process enables assessment designers and users to question the assumptions linking design choices to envisioned outcomes. The articulated pathway also provides a roadmap for systematic interrogation of these assumptions, which can help ensure that envisioned outcomes are indeed achieved. Applied to formative assessment programs, a ToA model signals careful attention to how educators and students are using assessment data, and it enables an investigation into the social and environmental factors that might mitigate or support the meaningful use of assessment data to guide teaching and learning.

The IDAF provides assessment designers with a heuristic that draws attention to the consequences of an assessment program on students, educators, and systems of education at every stage of an assessment's design and use. This focus on consequences of design choices elevates concern for fairness in the design and appraisal of assessment systems. The IDAF is ideally suited to the design of formative assessment programs because this attention to consequences compels assessment programs to generate meaningful, actionable data for each student and classroom that can inform instruction and support learning.

A further strength of using these three principled design frameworks when designing assessment programs is that they facilitate communication across specialists involved in the design and implementation of these assessment programs. Additionally, each of the frameworks emphasizes the importance of clearly articulating construct models in advance of an assessment's design, as the construct model is essential to task design, assessment blueprinting, the choice of measurement models, and the structure and nature of educational interventions. We would argue that robust, well-articulated construct models, developed in advance of an assessment's design, are essential to any principled approach to assessment design.

4.1 Construct Models

Each of the principled design frameworks chosen to guide the design of the WEC modules emphasized the importance of robust construct modeling as a foundational aspect of an assessment's design and use. As Corrigan and Slomp (2021, this issue) observe, construct modeling is an intensive, recursive exercise. As such, construct models, at both the nomothetic and idiographic levels of specification, provide a rich focal point for multidisciplinary collaboration. Table 1 illustrates the central role of construct modeling to the WEC prototype design process and the multidisciplinary expertise that engaged in conversations around this modeling work. As can be seen from Table 1, the construct model informed discussions of workplace communications, task design, global design features, score report design, and the ToA.

Table 1

Construct-Informed Design Artifacts

SI manuscript	Artifact	Artifact description
Oliveri, Mislevy & Slomp (2021)	Table 2	Construct dimensions for task design
	Table 4	Q-matrix
Oliveri, Slomp, Rupp, & Mislevy (2021a)	Figure 2	ToA model
Oliveri, Slomp, Rupp, & Mislevy (2021b)	Table 2	Higher-level design pattern
Corrigan & Slomp (2021)	Figure 2	Sociocognitive construct of writing expertise
	Table 2	Overview of domains of expertise
	Table 3	Relation of domains of expertise to curriculum, instruction, and assessment
	Table 4	Idiographic model of writing relative to WEC modules
Aull & Aull (2021)	Figure 1	Sociocognitive construct model of writing expertise
	Figure 2	Competencies involved in workplace email communication
	Figure 3	Rhetorical knowledge and genre knowledge in workplace email communication
Wetzel et al. (2021)	Figure 1	Mapping clusters of genre moves
Zapata-Rivera et al. (2021)	Table 1	Dimensions of the targeted construct
	Figure 2	Sample prospective score report for teachers
	Figure 3	Sample prospective score report for students

While the construct model informing this project predated it (Corrigan, 2019), the many conversations about the application of the model to the WEC KD prototype design led to a series of refinements of the model. In particular, conversations about operationalizing the construct at the idiographic level resulted in reflections on how the domains were described and linked to one another at the idiographic level. For example, in the original model, borrowing from Beaufort's (2007) rendering of expertise in writing, the domain of discourse community knowledge was considered superordinate to the domains of rhetorical knowledge, genre knowledge, subject matter knowledge, and writing process knowledge. However, as we engaged with the work of Wetzel et al. (2021, this issue), the rhetorical intention of the writer moved further into the foreground. As a consequence, the model was changed to accommodate this change in thinking. In the current iteration of the model, discourse community knowledge and rhetorical aim knowledge are placed at the same level within the model, and are situated within dynamic tension of one another. Similarly, Beaufort's classification of process-related knowledge as generalized writing process knowledge was revisited during discussions of the task design and

scoring. Through conversations with item designers, score report designers, psychometricians, and subject matter experts, it was eventually determined that we could not measure a generalized writing process, but that we could measure and compare processes with which test-takers engaged to complete each task. Research literature on second language writing processes confirmed the value of thinking in terms of task processes rather than generalized writing processes (Tardy, 2009). Consequently, the model was further refined to reflect these new insights.

4.2 Multidisciplinary Collaborations Guided by Principled Design Frameworks

Given the complexity involved in developing digital assessments of workplace communication skills, a diverse set of expertise is required for successful design. Teams envisioned by the authors contributing to this special issue include people with expertise in the following areas: assessment design, cognitive science, curriculum and instructional design, human-computer interaction, information visualization, subject matter (applied linguistics, workplace communications, writing), task design, psychometrics, and score report design.

Oliveri, Mislevy, and Slomp, (2021, this issue) observe that multidisciplinary teams need to work in non-hierarchical ways, thereby ensuring that one set of expertise does not drive the entire design process. Instead, each set of expertise drives the process, as needed, at different phases of design and validation. A design model based on distributed leadership might be difficult to manage. Principled design frameworks, however, can provide a structure for these collaborations by highlighting key questions, concerns, and required expertise at each stage of design, implementation, appraisal, and use.

Table 2 below exemplifies the ways that principled design frameworks can help organize research expertise. This reconceptualization of multidisciplinarity described in §5.4 of the introduction to this special issue shows how Classification of Instructional Programs (CIP) codes—a federal system that provides a taxonomic scheme that supports the accurate tracking and reporting of fields of study—may be integrated in terms of design frameworks (Oliveri, Slomp, Elliot, et al., 2021).

Table 2

How Principled Design Frameworks Help Organize Expertise

Design frameworks			Principled design questions	Required expertise									
e-ECD	ToA	IDAF		Assessment design	Cognitive science	Curriculum and instructional design	Educational policy	Human-computer interaction	Information visualization	Subject matter	Task design	Psychometrics	Score report design
	X		1. What are the desired outcomes from the implementation and use of this assessment program?	X		X	X			X			
	X		2. What program elements will support achieving these desired outcomes?	X	X	X	X		X	X			X
X			3. What kind of misconceptions might be identified in advance so that feedback mechanisms and learning aids are provided to students to provide guidance for stakeholder actions?	X	X	X	X	X	X	X	X	X	X
	X		4. What action mechanisms will be needed to mediate between program elements and desired outcomes?	X	X	X	X	X		X		X	X
		X	5. How will the development and use of the assessment support the aims and the targeted populations?	X	X	X	X			X		X	X
X		X	6. What are the constructs and content domains targeted for assessment?	X	X	X	X			X		X	
X		X	7. What are the tasks, types, settings, contexts, that may be used to collect the information/data sought?	X	X	X		X		X	X	X	X
X			8. What types of measurement/learning analytics models could be used to assess the targeted construct and knowledge domains?	X	X			X	X		X	X	X
		X	9. How should information be presented to stakeholders to optimize achievement of desired outcomes?	X	X	X		X		X		X	X
		X	10. How will scoring criteria and procedures support achievement of desired outcomes?	X		X				X		X	X
X			11. What kinds of interpretations can be made from behavioral observations and model-based summaries?	X	X		X					X	X
		X	12. What is the evidence for the validity, reliability, and fairness of the assessment program?	X	X	X	X	X		X	X	X	
		X	13. Have the desired outcomes motivating the assessment program been achieved?			X	X			X		X	
		X	14. Taken collectively, does the evidence provide an understanding of the positive or negative unintended impact from the assessment program?	X		X	X			X		X	

Oliveri, Slomp, Rupp, and Mislevy (2021a, this issue) observe that in addition to relying on design frameworks to guide collaborations, multidisciplinary teams can focus their work around shared artifacts shown in Table 1 that provide a focal point for the above questions. For example, a ToA framework (see Oliveri, Slomp, Rupp, & Mislevy, 2021a, this issue) motivates questions 1, 2, and 4 in Table 2 above; as such, the framework itself can provide a focal point around which all 14 questions can be explored. Slomp and Elliot (in press) demonstrate how curriculum and instructional designers, and subject matter experts can collaborate around a ToA model to advocate for assessments that support identified learning goals.

5.0 Interpretation and Use Arguments

Each of these frameworks complement and extend contemporary models of validation focused on IUAs, fleshing out mechanisms for more systematically collecting and examining evidence linking facets of an assessment's design to the IUA of the data that assessment program generates. While validity historically has been concerned with the IUAs of assessment data, there has long been a recognition that the quality of an assessment instrument is integral to the quality of the data generated through the use of that instrument. While it is true that poor inferences can be drawn from high-quality and accurate data, it is virtually certain that poor-quality data leads to poor inferences. For this reason, Tannenbaum and Katz (2021, this issue) draw a helpful distinction between design validity and interpretation validity—the former providing the foundation for the latter. This view is consistent with the e-ECD, ToA, and IDAF models that have informed the design of the WEC KD prototype.

5.1 Design Validity

The concept of design validity is not often associated with the IUA validation model. It might, therefore, be helpful to think of design validity—and evidence related to it—in terms of technical design considerations that are anticipatory to providing warrants for the scoring, generalization, extrapolation, and decision inferences that are necessary elements of an IUA.

Tannenbaum and Katz (2021, this issue) suggest that the foundation of design validity is a carefully articulated ToA, linking assessment design to intended outcomes because it provides the blueprint for designing and constructing an assessment program including task design, scoring approaches, and reporting mechanisms. Here, we draw attention to three sets of technical design considerations that shaped our development of WEC KD modules.

5.1.1 Q-Matrix

The bridge between the construct model and the inferences to be made from assessment data based on that model is the Q-matrix—a cognitive diagnostic model used to connect measurement models, tasks (often referred to as items or questions), and students' performances. The Q-matrix provides a blueprint for the assessment which helps to ensure construct representation across the

assessment and helps connect evidence of the construct collected from test users to aggregates of their performance on the various aspects of the construct.

The design of a Q-matrix is an iterative process. Construct features are identified, key elements of the content domain are foregrounded, and the range of contextual or situational factors that bridge the learning environment and the target environment are articulated. Item types are identified. These are then plotted against the construct, content, and situational dimensions identified. The goal of this mapping exercise is to ensure both balance and breadth in how these dimensions are sampled by the assessment so that information on students' performance can be tracked and communicated to users via score reports, as we elaborate later. Tannenbaum and Katz (2021, this issue) observe,

> For complex performance tasks, given their many interacting activities and the richness of the potential data to be collected, consistent documentation of the task features and how they relate to the construct is particularly important. Such articulation of the underlying argument of the tasks and how they fit together into the larger assessment form key elements of design validity that ultimately support intended inferences about learner competencies. (pp. 205-206)

They also observe that because performance tasks often provide assessment designers with fewer opportunities to collect evidence (than say a traditional selected-response assessment), it is important to map each task to the matrix of construct facets it is designed to measure. With a view to assessments that support transfer, it is equally important to map these tasks to the features of the target environment in which test-takers will be expected to apply the knowledge and skills being measured.

The iterative nature of Q-matrix design involves the constant tweaking of how items are mapped as items are reviewed by experts, and as data from play studies, usability studies, cognitive studies, and small-scale tryouts reveal how well each task in fact aligns with the facets to which it was mapped. These data need to inform either item redesign or modifications of the Q-matrix. The goal of this iterative process is to ensure robust construct sampling and representation in the final assessment.

5.1.2 Assessment Tasks

Haigler (2021, this issue) observes that affinity between tasks in the learning and the assessment context and those in the workplace is important to supporting the transfer of learning from educational to the workplace environments. This observation necessarily guides our WEC prototype design with a focus on the verisimilitude of tasks to the types of activities that will likely need to be carried out in the workplace and the need to expose students to such tasks to enhance workplace readiness and preparation.

Aull and Aull (2021, this issue) support this work with a detailed analysis of workplace email communication tasks—the focus of the WEC KD prototype. They observe that there is great

variability within the email genre. For example, emails share features of both spoken and written communication, and they range from the highly formal to the highly informal. How this variability is expressed is often a function of the social environment in which they are crafted. These social factors include the following: complex networks of hierarchical relationships between senders and receivers of emails; the multiple, intertwined purposes that motivate email communications; and the cultural, gendered, and intrapersonal factors that shape interpretations of the appropriateness of communications between colleagues. All of this complexity is situated with the ecological contexts—including the values and purposes of the organizations—within or between which the email communications take place.

Drawing from Aull and Aull's analysis (2021, this issue), we recognize that the assessment tasks incorporated into the WEC modules needed to reflect this range of social factors. If the goal of the assessment is to measure test-takers' ability to problem-solve complex writing tasks, test-takers also need opportunities to demonstrate their capacity to navigate these sets of social and ecological complexities while also performing the range of design options available to them within the scope of the genre.

This range of complexities underscores the importance of a well-designed Q-matrix that enables item designers to ensure that items sample both the construct facets required, and that they reflect the range of ecological, social, and intrapersonal factors that mediate this form of communication.

5.1.3 Scoring

Tannenbaum and Katz (2021, this issue) observed that performance tasks provide richer information about learner knowledge, skills, and metacognitive capacity. At the same time, however, the range of complexities described above fosters real challenges in scoring. While there have been advances in automated scoring capabilities, these systems still lack the capacity to provide the rich information needed for the assessment data to meaningfully inform teaching and learning.

Human scoring, too, remains challenging, as complex tasks require the design of complex rubrics linked to construct domains. In the WEC prototype, multiple rubrics, tailored to each assessment task and keyed to different constellations of construct domains, needed to be developed. Similarly, human scorers need to be trained, and their scoring needs to be monitored for issues of construct-irrelevant variance (traits of the scorer influencing scores).

Once developed, rubrics need to be critically reviewed. We advocate the use of Broad's (2003) Dynamic Criteria mapping as a process for assessing the rubrics. Using this process, we may compare expert reviews of a series of test-taker-generated emails to determine what features of email communications experts valued in the genre and mapped these against the rubrics used to score the emails. This activity enables teachers and learners to identify, through the use of reflection questions and rubrics, which aspects of their email composition choices approximated

effective use of linguistic features and which ones could be tweaked/refined for closer approximation and increased effectiveness in email writing.

5.1.4 Usability Studies

Essential to evidence of design validity are usability studies through which the assessment tasks under development are critically examined. The purpose of these studies is to examine if the assumptions about the tasks are valid. As Tannenbaum and Katz (2021, this issue) suggest, these usability studies occur in three stages: early prototype stage, initial assessment stage, and deployed assessment stage. Table 3 summarizes the purpose of each of these sets of studies, the key questions asked during these studies, and the evidence collected to address these questions.

Table 3

Overview of Usability Studies

Stage	Purpose	Key questions	Evidence collected
Early prototype	Examine the link between test-taker behavior and anticipated interpretations and claims about test-takers	What information is each task eliciting from test-takers? Does information from each task support intended inferences and claims about test-takers?	Observations of learner performance on tasks
Initial assessment	Examine if the tasks and scoring systems address the construct of interest	Do the tasks address the construct of interest? Do the tasks elicit expected responses?	Think-aloud data, keystroke logs, eye-tracking data
		Are the tasks consistent with expert opinion about what is needed to assure relevance, authenticity, and generalizability?	Expert coding of assessment tasks
		Does the scoring system provide construct-relevant data while minimizing construct-irrelevant variance?	Expert coding of scoring criteria, analysis of rater behaviors, and examination of performance data by subgroups
Deployed assessment	Examine construct representation, construct-irrelevant variance, uses of assessment data, and their consequences	How reliable are the assessment tasks?	Statistical analysis of performance data focusing on task difficulty and discrimination, differential task functioning, differential performance by subgroups, analysis of internal structure, and relationship between performance data and external criterion
		How are decision-makers using performance data? What are the consequences of how performance data are used?	Interviews with decision-makers

Use of the ToA framework and IDAF complement Tannenbaum and Katz's (2021, this issue) recommendations for usability studies. By articulating upfront the intended outcomes stemming from the use of the assessment program and the design features that are anticipated to support those outcomes, the ToA provides a transparent frame for the usability studies that follow. The IDAF maps well onto this set of envisioned studies, adding to them a consistent focus on the consequences of design choices.

5.2 Interpretive Validity

While design validity focuses on the assessment design, and on the capacity of the assessment to provide data that can inform inferences and decisions about test-takers, interpretive validity focuses on critically examining the IUAs drawn from those data. We divide this category of validation into two focal areas: critical examination of the measurement models used to bridge raw performance data and key inferences about test-takers, and critical review of score reporting.

5.2.1 Measurement Models

Measurement models, Mislevy (2018) explains, draw a link between performance on a set of assessment tasks and the meaning to be made from that performance. Measurement models include variables to analyze test-takers' capabilities (e.g., proficiency variables) and associated background variables (e.g., covariates), variables to evaluate what test-takers do (i.e., observable variables such as students' responses on tests), and variables for features of situations (e.g., task variables and associated parameters), among others. Reasoning backwards through a link function, probabilistic reasoning creates a pathway from observable performance, through person and situation variables, to inferences about test-takers' capabilities. An IUA provides the rationale and justification for the inferences made from test-taker performance data based on students' performance on tests.

Adopting a sociocognitive perspective, Mislevy (2018) observes the challenges of developing accurate measurement models linking performance on an assessment of a complex construct to inferences about test-takers' performances in non-testing situations: "even examinees who have given identical responses construct them from personal resources, developed through their own unique histories. These would almost surely lead to different behavior in some other potential situations" (p. 155). The solution, he posits, is that assessment developers have had to argue that the model used is "close enough" to support the inferences being made. This raises the question, for each assessment program, how close is close enough? And, from a conditional sense of fairness perspective, what if some inferences are closer for some test-takers, such as the ones that are more closely aligned with the culture and background of the test developers, and are more distant for others (Mislevy et al., 2013)?

At the heart of interpretive validity, then, is a critical review of the models that support the inferences and uses drawn from test-taker performance data. Gathering evidence of interpretative validity thus begins with a review of the construct model underpinning the assessment program.

How well does the construct model account for the range of KSAs that support expertise in the area being assessed? With respect to WEC prototypes, this process involves a critical review of the KSAs that enable success in workplace email communication, an examination of how well the assessment captures those construct elements, and the degree to which those construct elements are reflected in the scoring criteria.

Gathering such evidence also involves an analysis of the resonance between the assessment tasks and the real-world contexts to which inferences will be drawn, and highlights the importance of collaboration from various stakeholders including employers, employees, instructors, and students to name but a few (see Haigler, 2021, this issue for examples of such initiatives conducted by states such as the Texas Workforce Commission project, which evolved under Haigler's leadership). In relation to the WEC prototype, addressing this objective involves critically examining how well the performance tasks reflect features of workplace email writing tasks in the real world. Are key aspects of real-world settings that impact performance accounted for in the assessment's design? Aull and Aull (2021, this issue) link concerns for construct modeling with consideration of the resonance between assessment tasks and real-world communications tasks. Construct evidence forms a precondition to and ongoing part of the definition of communication in a given context; construct evidence would include an indication from learners and employers that these tasks practiced in the KD matched tasks expected of learners in their workplaces. This analysis, Aull and Aull further observe, provides the basis for supporting both decision and extrapolation inferences.

In addition to accounting for the resonance between assessment tasks and their counterparts in the real world, interpretive arguments require—within the sociocogntive frame the WEC KD prototype has been developed—an analysis of the LCS patterns that shape test-takers' approaches to the assessment tasks. This analysis can be used to explain both their performance on each assessment task and their performances in criterion situations. These LCS patterns need to be accounted for in the data collected (both performance data and supplemental data) through the assessment program. Oliveri, Mislevy, and Slomp (2021, this issue) observe that it is important to establish the Q-matrix noted above, drawing on LCS patterning, that links individual and combined skills with facets of assessment tasks and their counterparts in the real world, to help provide further warrants for the extrapolation inference.

Finally, interpretive validity within a sociocognitive framework involves a critical review of the measurement models used to draw inferences from the performance data. Because each measurement model achieves different purposes and allows for different types of inferences, Oliveri, Mislevy, and Slomp (2021, this issue) observe that it is critically important to choose measurement models that are best suited for the measurement goals. Mislevy (2018) cautions that "we must keep checking back and forth between our reasoning in the model space and intended counterpart in the real world, drawing on our theories and experience to reason about our reasoning through the model" (p 163). Especially germane to the development of the WEC

KD modules, Mislevy cautions that the more complex the construct being measured, the more challenging it is to build models that provide valid interpretations.

5.2.2 Score Reporting

Score reports are the interface between performance data, inferences made from that data, and decisions made from those inferences. Score reports need to reflect the complexity of the models that inform them, taking into consideration the variables for test-taker's capabilities, variables for what test-takers do, and variables for features of situations contained in the measurement model for that assessment program. In addition, score reports need to consider the audiences for the data, ensuring the data are presented in a manner that enables justifiable inferences and sound decisions.

Zapata-Rivera et al. (2021, this issue) suggest two score reporting strategies: simple description of learner behaviors and score profiling. When score reports are provided as simple descriptions of learner behaviors, instructors and test-takers are provided with frequencies or counts of test-taker engagements with targeted behaviors. When score profiling is used, test-takers are clustered in groups according to similar behavior profiles. The characteristics of the behavior patterns that inform these clusters are then shared with test-takers and instructors.

Similar to the development of test items and scoring rubrics, the development of score reports requires its own program of research. Through an iterative cycle, design frameworks, mock-up score reports, and functional reports are developed and shared with end users. These, too, can be scaled up from individual think-aloud studies, to focus groups, to large-scale studies that examine comprehension, usability, and appropriate use of information presented through the score reports.

6.0 Cost

In the varied frameworks our authors adopted, little or no mention was made of the costs involved in developing the WEC prototype. As Table 4 below shows, the total development cost was approximately $750,000.

Table 4

Funding for Kitchen Design Prototype

Funder	Grant Title	Amount	Year
Educational Testing Service (ETS) Research & Development (R&D)	Employee Preparedness for Employment in Career and Technical Education	~$50,000	2019-2020
ETS R&D	Formative Assessment of Workplace English Communication & Collaboration Skills	~$100,000 per year	2018-2020
ETS R&D	Digital Training Modules to Teach and Assess Workplace Communications & Collaboration Skills	~ $50,000	2018-2019
ETS Test of English for International Communication (TOEIC)	Validity of Workplace English Communications Test Scores and Consequences of Score Use	~ $50,000 per year	2015-2018
ETS R&D	Defining Workplace English Communication Construct Elements	~ $50,000	2017-2018
Texas Workforce Commission	Literacy Texas Standards Alignment to Industry Clusters Initiative	~$50,000	2013-2014
Institute of Education Sciences	UTSA Educational Research Training Program: P-20 Pipeline Issues – Pathways to Education Sciences	~$100,000 per year	2016-2018

The method of cost effectiveness analysis described in §5.0 of the introduction was, of course, impossible to employ: When the project was terminated in 2021, there were no effect size measures to use. The reported cost includes tasks involved with prototyping; included in those costs are the costs of equipment, facilities, administration, teacher training, and maintenance; however, a finer grain breakdown of costs for specific activities remains unknown.

7.0 Looking to the Future

The work contained in this special issue has provided the foundation for a program of research focused on the design of a digitally based, formative assessment program for teaching WEC. To date, four projects are either proposed or underway. A major shift in process as we proceed with the development of a new set of WEC modules is the involvement of key stakeholder groups associated with diverse populations of learners from the earliest stages of the design process.

In the first project, we are consulting with faculty in business and management programs at two universities on the design of an initial set of modules set in an international business

scenario. In the second project, we are partnering with post-secondary career services units at two universities, collaborating with their student advisors, corporate partners, and students on the design of modules and formative feedback processes. In the third project, we are partnering with high school programs dedicated to preparing minority students for successful transitions into post-secondary or workplace contexts. Through a codesign collaborative model informed by a commitment to culturally sustaining pedagogies, we are examining how the modules that will be developed, and the pedagogical and assessment processes that will be developed, can be tailored to diverse populations of students. This project, along with the fourth project, links our sociocognitive model of writing to a linguistic justice framework (Randall et al., 2021). Our fourth project partners with minority-serving colleges and universities to further develop, refine, and expand this platform for instruction at the post-secondary level with faculty in business, education, and writing studies programs, and with diverse populations of undergraduate students, and pre-service teachers.

As this program of work evolves, the partnerships needed to bring it to fruition are further expanding. In addition to the disciplinary expertise listed in Table 2 of the introduction to this special issue (Oliveri, Slomp, Elliot, et al., 2021), we have brought in experts in fields of linguistic justice and culturally sustaining pedagogies. More significantly, we have brought teachers and instructors into this program of research, not as subjects, but as partners who have expert knowledge both on the diverse populations they are serving, and on the pedagogical strategies that best serve these populations. As we move forward, we hope to formalize partnerships with corporate entities so that we can better tailor modules and instruction to relevant and evolving workplace communications tasks, and so that we can begin to measure the long-term impact of this intervention.

Through this program of research, we hope to achieve four long-term outcomes:

1. the development and refinement of a digital platform that provides high-quality formative feedback to diverse populations of students and teachers working to develop WEC;
2. further the development and refinement of principled design processes for large-scale formative assessment programs of WEC and other complex constructs;
3. the ability to enable and empower a cohort of teachers equipped to lead innovation in the teaching and assessment of WEC in secondary, post-secondary, and workplace contexts; and
4. the ability to empower diverse populations of students to be strong, independent writers who achieve enhanced outcomes in workplace communications that enable them to thrive in their personal, professional, and academic lives.

These goals are as ambitious as they are bold, but we believe that the current moment demands new approaches, better outcomes for all, and a courageous vision. We hope this unfolding project can contribute to that future.

8.0 Appreciation

It is hard to capture in the pages of this special issue and in this summary afterword how extensive and demanding the collaboration described in this afterword really was. As design teams, authors, and editors, we lived the challenges around transfer and accessing LCS patterns that are described in this paper. Designing the WEC modules at the core of this special issue required multiple meetings a week between team members spread across North America, working out of different research traditions and areas of expertise. In the midst of completing this work, we were hit by the COVID-19 global pandemic that severely disrupted our work lives and our capacities to push forward as we experienced losses we could not bear. In addition to the layers of expertise required to advance this program of work, the structure provided by principled design frameworks, the focus enabled by an orienting theoretical framework, and a common goal provided by a shared concern for the immediate and long-term consequences flowing out of this work, the success of this project was made possible by visionary, ethical, principled, and flexible multidisciplinary collaboration. Our special thanks to *The Journal of Writing Analytics* for patiently supporting this project, to the Board of Reviewers, to Editor-in-Chief Susan Lang for her guidance as we brought the project to completion, to colleagues in the Department of English at the University of South Florida, and to Michael Palmquist, our publisher.

Our authors and reviewers endured with us while the pandemic saddened their lives. Because they prevailed with the publication of this special issue, their story is one of hope.

References

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M., & Norman, M. K. (2010). How *learning works*. Jossey-Bass.

Arieli-Attali, M., Ward, S., Thomas, J., Deonovic, B., & Von Davier, A. A. (2019). The expanded evidence-centered design (e-ECD) for learning and assessment systems: A framework for incorporating learning goals and processes within assessment design. *Frontiers in Psychology*, *10*, 853.

Aull, B., & Aull, L. (2021). "Write a greeting for your email here:" Principles for assessing interpersonal workplace email communication. *The Journal of Writing Analytic*s, *5*.

Baker-Bell, A. (2020). Dismantling anti-black linguistic racism in English language arts classrooms: Toward an anti-racist black language pedagogy. *Theory Into Practice*, *59*(1), 8-21. https://doi.org/10.1080/00405841.2019.1665415

Beaufort, A. (2007). *College writing and beyond: A new framework for university writing instruction*. University Press of Colorado.

Broad, B. (2003). *What we really value: Beyond rubrics in teaching and assessing writing*. University Press of Colorado.

Connolly, P., & Vilardi, T. (1989). *Writing to learn mathematics and science*. Teachers College Press.

Corrigan, J. A. (2019). *Writing in times of deixis: A validation study of a large-scale assessment of new literacies* [Doctoral dissertation, University of Ottawa]. uO Research. http://dx.doi.org/10.20381/ruor-22948

Corrigan, J. A., & Slomp, D. H. (2021). Articulating a sociocognitive construct of writing expertise for the digital age. *The Journal of Writing Analytic*s, *5*.

Fry, S. W., & Villagomez, A. (2012). Writing to learn: Benefits and limitations. *College Teaching*, *60*(4), 170-175.

Haigler, K. (2021). The role of occupational research in complex assessments: Aligning educational practices with workplace requirements. *The Journal of Writing Analytic*s, *5*.

International Organization for Migration. (2019). *World migration report 2020*. https://publications.iom.int/system/files/pdf/wmr_2020.pdf

Lewis, R. D. (2006). *When cultures collide: Leading across cultures*. Nicholas Brealey Publishing.

Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement.* Routledge.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design* (Research Report RR-03-16). Educational Testing Service. https://www.ets.org/Media/Research/pdf/RR-03-16.pdf

Mislevy, R. J., & Duran, R. P. (2014). A sociocognitive perspective on assessing EL students in the age of Common Core and Next Generation Science Standards. *TESOL Quarterly*, *48*(3), 560-585. https://doi.org/10.1002/tesq.177

Mislevy, R. J., Haertel, G., Cheng, B. H., Ructtinger, L., DeBarger, A., Murray, E., Rose, D., Gravel, J., Colker, A. M., Rutstein, D., & Vendlinski, T. (2013). A "conditional" sense of fairness in assessment. *Educational Research and Evaluation*, *19*(2-3), 121-140.

Mislevy, R. J., & Oliveri, M. E. (2019). *Digital module 09: Sociocognitive assessment for diverse populations (Version 2.0)*. https://ncme.elevate.commpartners.com

Oliveri, M. E., & Ercikan, K. (2011). Do different approaches to examining construct comparability lead to similar conclusions? *Applied Measurement in Education*, *24*, 1-18.

Oliveri, M. E., Lawless, R. R., & Mislevy, R. J. (2019). Using evidence-centered design to support the development of culturally and linguistically sensitive collaborative problem-solving assessments. *International Journal of Testing*, *19*(3), 270-300. https://doi.org/10.1080/15305058.2018.1543308

Oliveri, M. E., Mislevy, R., & Elliot, N. (2020). New horizons for postsecondary placement and admission practices in the United States. In M. E. Oliveri & C. Wendler (Eds.), *Higher education admission practices: An international perspective* (pp. 347-375). Cambridge University Press.

Oliveri, M. E., Mislevy, R. J., & Slomp, D. H. (2021). Principled development of Workplace English Communication Part 1: A sociocognitive framework. *The Journal of Writing Analytic*s, *5*.

Oliveri, M. E., Slomp, D. H., Elliot, N., Rupp, A. A., Mislevy, R. J., Vezzu, M., Tackitt, A., Nastal, J., Phelps, J., & Osborn, M. (2021). Introduction: Meeting the challenges of Workplace English Communication in the 21st century. *The Journal of Writing Analytic*s, *5*.

Oliveri, M. E., Slomp, D. H., Rupp, A. A., & Mislevy, R. J. (2021a). Principled development of Workplace English Communication Part 2: Expanded Evidence-Centered Design and Theory of Action frameworks. *The Journal of Writing Analytics*, 5.

Oliveri, M. E., Slomp, D. H., Rupp, A. A., & Mislevy, R. J. (2021b). Principled development of Workplace English Communication Part 3: An integrated design and appraisal framework. *The Journal of Writing Analytics*, *5*.

Randall, J., Poe, M., & Slomp, D. (2021). Ain't oughta be in the dictionary: Getting to justice by dismantling anti-black literacy assessment practices. *Journal of Adolescent and Adult Literacy. 64*(5). 594-599.

Slomp, D., Corrigan, J. A., & Sugimoto, T. (2014). A framework for using consequential validity evidence in evaluating large-scale writing assessments: A Canadian study. *Research in the Teaching of English*, *48*(3), 276 -302.

Slomp, D., & Elliot, N. (2021). What's your theory of action: Making good trouble with literacy assessment. *Journal of Adolescent and Adult Literacy, (64)*4. 468-475.

Tannenbaum, R. J., & Katz, I. R. (2021). Validity considerations in complex task design. *The Journal of Writing Analytics*, *5*.

Tardy, C. M. (2009). *Building genre knowledge*. Parlor Press.

Wetzel, D., Brown, D., Werner, N., Ishizaki, S., & Kaufer, D. (2021). Computer-assisted rhetorical analysis: Instructional design and formative assessment using DocuScope. *The Journal of Writing Analytics*, *5*.

Zapata-Rivera, D., Andrews-Todd, J., & Oliveri, M. E. (2021). Communicating assessment information in the context of a workplace formative task. *The Journal of Writing Analytics*, *5*.