Research Article

Principled Development of Workplace English Communication Part 2: Expanded Evidence-Centered Design and Theory of Action Frameworks

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

David Slomp, *University of Lethbridge*

André A. Rupp, *Mindful Measurement / Success Academy Charter Schools*

Robert J. Mislevy, *Educational Testing Service*

Structured Abstract

- **Background**: In today's rapidly evolving world, technological pressures coupled with changes in the nature of work increasingly require individuals to use advanced technologies to communicate and collaborate in geographically distributed multidisciplinary teams. These shifts present the need to teach and assess an expanded set of knowledge, skills, and attitudes, including how to communicate at work in collaborative environments using diverse forms of technology. They also present the opportunity to create novel forms of instructional materials and forms of assessment that extend the more traditionally used summative assessments to assessments used for learning and instruction. This design process can be facilitated through the use of conceptual frameworks employed to guide assessment design and development. Their use is important to support more expansive and complex design goals emerging in the design of

assessments of 21st century skills such as Workplace English Communication (WEC). In this article, we reflect on an evolving WEC construct needed for today's economy and discuss implications for expanding how we teach and assess it using formative assessments for learning. We then discuss the features of the expanded Evidence-Centered Design for learning and assessment systems (e-ECD) and Theory of Action (ToA) frameworks and illustrate their integrative application to inform the design and development of WEC training modules (or resources). We conclude with suggestions for next steps in this line of research.

- **Questions Addressed**: In reference to the e-ECD and ToA frameworks, our article addresses questions in two areas. We illustrate the benefits of using the ToA to explicitly identify the components of an assessment, its action mechanisms, stakeholders' needs, score-based decisions and their impact, and the services designed for test takers and users. We illustrate the benefits of using the e-ECD framework to guide design efforts in principled ways to enable consideration of both key elements that relate evidentiary elements relevant to the construct, aspects of learning and assessment, and measurement models. Consideration of these frameworks is important to design assessments and make sense of the evidence for meaningful interpretation of students' results.
- **Conclusions:** This article illustrates the application of conceptual (e.g., the e-ECD and ToA) frameworks that can be used to inform the design and development of similar modules for complex tasks of 21st century skills. This article contributes to the literature on WEC and complex assessments of hard-to-assess constructs more generally by offering a way of thinking about designing, assessing, and then evaluating the design and assessment of interactive educational modules for teaching complex communication knowledge and abilities while remaining attentive to (negative) consequences associated with the stakeholders designing, developing, and using the assessments.

Keywords: anticipatory design frameworks, blended learning, digital modules, Evidence-Centered Design (ECD), expanded Evidence Centered Design (e-ECD), Theory of Action (ToA), Workplace English Communication (WEC), writing analytics

1.0 Background

The fourth industrial revolution (Industry 4.0; Schwab, 2017) focuses on interconnectivity, automation, and the use of real-time data. At work, Industry 4.0 has brought changes to how people communicate with colleagues, clients, and supervisors, with associated increases in the use of digital technologies and environments used to communicate. These shifts have led to increases in the need to expand the skills necessary to communicate, which we refer to as

Workplace English Communication (WEC). We define WEC—a form of sophisticated discourse in which organizational and disciplinary norms for framing and communicating information are used for a variety of aims—as involving various forms of linguistic knowledge including

- rhetorical knowledge (knowledge of the audience and purpose of the communication),
- genre knowledge (knowledge of the types of linguistic and contextual features to use when communicating using various forms of communication),
- metacognitive knowledge or awareness (awareness and understanding of one's own responses within a communicative exchange), and
- substantive knowledge (knowledge of the types and level of complexity of language needed when communicating with different audiences such as technical and non-technical audiences).

In later sections of this article, we elaborate on our definition of WEC; we also refer the reader to Corrigan and Slomp (2021, this issue) for an elaboration of the WEC construct.

We suggest that shifts due to Industry 4.0 precipitate changes in the nature of work and how people communicate and collaborate at work, which present opportunities to expand the traditional construct of communication taught in academic contexts. Such expansion may involve teaching students a wider repertoire of genres, that is, not only focusing on the writing of essays or narrative forms of communication but also providing them with opportunities to learn other genres or forms of communication occurring at work (e.g., email writing, real-time chat-based communication, workplace documents, and technical documentation).

Shifts in assessment may also include expanding the forms of assessment used for decision-making and informing teaching and learning—from a main focus on using summative assessments to using interim and formative assessments. The goals are to provide students with feedback on their communication abilities (see Wetzel et al., 2021, this issue for the use of writing analytics to teach students various workplace communication genres). When assessments are used for formative purposes, to shape instruction and learning, the issue of fairness—with its attention to bias, disparate impact, equity of opportunity to demonstrate learning, and equity of opportunity to learn—in assessment design and use is brought to the fore. In such cases, inaccuracies in the information generated by an assessment cannot be buried in the aggregate; information must hold true for each learner. In this way, formative assessments become integral to each student's opportunity to learn, broadly defined as their access to the resources needed to learn (Moss et al., 2008).

These shifts result in new opportunities for researchers, educators, and students. For researchers, they enable opportunities to study the knowledge, skills, and attitudes, as well as novel genres of assessment for learning (see Slomp et al., 2021, this issue for more about assessment for learning). For teachers, they create opportunities to integrate digital learning resources into classroom practices to teach and assess students' learning. For students, they enable them to practice the needed skills to be better prepared for the evolving 21st century

academic and workplace settings. With an emphasis on fairness and opportunity to learn, we note that an expanded construct model can be beneficial in enhancing preparation, teaching, and assessment of WEC. In contrast, a focus on a narrower construct may reduce opportunity to learn, leading to underprepared students who lack needed skills for employability (Oliveri et al., 2020).

In this article, we discuss these issues in five sections. The present section provides an overview of the article. The second section elaborates on the need to teach and assess the WEC construct and discusses implications for expanding how we teach and assess it using formative assessments for learning. The third and fourth sections present two frameworks, the expanded Evidence-Centered Design for a learning and assessment system (e-ECD) framework, as well as the Theory of Action (ToA) framework, to guide the development of WEC training modules (or resources). We conclude with suggestions for next steps in this line of research.

1.1 Complementarity

To address the issues discussed in the roadmap, we illustrate our application of the e-ECD and ToA frameworks to the design of digital training modules we prototyped. The e-ECD framework is designed to guide design efforts in principled ways by jointly considering key elements that relate evidentiary elements relevant to the construct, aspects of learning and assessment, and measurement models. The ToA framework is designed to explicitly identify the components of an assessment, its action mechanisms, stakeholders' needs, score-based decisions and their impact, and the services designed for test takers and users. The ToA also allows us to explicitly identify the action mechanisms and the decisions, behaviors, and solutions sought from the assessments.

We posit that these two frameworks (e-ECD and ToA) function in complementary ways and that their joint use allows designers, researchers, and developers to consider the array of central aspects of assessment that are needed when designing, developing, and using assessments of complex constructs. These issues involve considering the needs of stakeholders, test takers, and users of the data for instructional and/or assessment purposes; these processes are facilitated through the use of the ToA. Moreover, these issues involve considering those needs and their implications for design and development of the specific assessments used for learning; these processes can be facilitated through the use of the e-ECD framework. Further, these two frameworks complement each other by enabling consideration of different aspects relevant to assessments at different levels of abstraction. The ToA provides an overview of the needs of stakeholders, test takers, and users of the data for instructional and/or assessment purposes. The e-ECD framework is more concerned with fine-level details of the construction of assessments themselves, such as the development of the items, scores, scoring guides, and the like.

1.2 Modules

We illustrate the use of these two frameworks for the design of digitally distributed scenario-based modules. The modules were designed for use by high school and college students in a *blended learning environment* (e.g., Dziuban et al., 2018). A blended learning environment involves combining digitally delivered online interactive educational materials and instructor-led classroom methods. The modules present opportunities for students to learn WEC using engaging scenarios where they work in a fictitious company that specializes in designing and overseeing the construction of commercial and private kitchens. No specialized knowledge of design or kitchens is needed to participate in the modules. A first prototype we created was shared with the contributors of this special issue. It illustrates considerations for the design for a singular communication mode—email writing—and its associated WEC construct components for illustrative purposes.

Note that in this article, we do not elaborate on specifics of the modules at a fine level of detail such as the types of embedded prompts / items or scoring approaches used. Instead, we provide enough details to give a broader conceptual context to explain our application of the two frameworks for the design of the modules. Other articles in this special issue (see Oliveri, Mislevy, & Slomp, this issue, and Oliveri, Slomp, et al., 2021, this issue) provide more details and key considerations for designing the items, feedback, and scoring mechanisms. Moreover, we do not advocate a prescriptive approach to design; rather, we suggest careful deliberation of key design aspects that are malleable depending on the context of local classrooms or the populations participating in the assessments. Such decisions are best informed through dialogues with all constituents composing the ecosystem where the modules would be used and the types of decisions that would be informed through the use of the modules (see Oliveri et al., 2020 for an overview of a multilevel design framework to inform design decisions within higher-education contexts).

2.0 Industry 4.0 Impact on the WEC Construct and its Assessment

With respect to workplace communication, Industry 4.0 has led to a need to expand the WEC construct and the assessment forms used to assess it. WEC skills are one of the top five most important skill clusters needed for work, yet students are often underprepared to communicate at work with different audiences and different forms of communication (Casner-Lotto & Barrington, 2006; Hart Research Associates, 2015). Results of employer surveys consistently report that incoming employees are lacking competency in communication (National Association of Colleges and Employers [NACE], 2018).

According to a survey conducted by Hart Research Associates (2015, p. 4), the following college learning outcomes were rated as "very important" by more than 80% of employers: (1) "the ability to effectively communicate orally" (85%), (2) "the ability to work effectively with others in teams" (83%), (3) "the ability to effectively communicate in writing" (82%), and (4) "the ability to apply knowledge and skills to real-world settings (80%). Nevertheless, only about

a quarter to a third of employers (between 23-37%) felt that recent graduates are well-prepared in these four areas. Differences between employers' beliefs with respect to incoming employees' preparation and the degree to which students think they are prepared is large, with employers often reporting that employees are underprepared for the workplace (Hart Research Associates, 2015; LeFebvre & Clark, 2013; Ortiz & MacDermott, 2017).

Notably, Hart Research Associates (2015) reported that while 74 percent of students believe colleges and universities are "doing a good job" in preparing them for entry-level positions, only 42 percent of employers share this belief. Similarly, according to a poll of business leaders conducted in 2013, only 11 percent strongly agreed and an additional 22 percent agreed that "higher education institutions in this country are graduating students with the skills and competencies that my business needs" (Gallup & Lumina, 2014, p. 25). Low proficiency in communication has consequences for employers, employees, instructors, and students.

2.1 Importance of WEC in a Shifting Workplace Environment

WEC is important for various stakeholders across countries, businesses (both employers and employees), and educational institutions (both instructors and students). For a country, low levels of literacy can limit the opportunities available to individuals and create a non-diversified workforce for jobs that require further education (Kirsch et al., 2007).

For employers and employees, the costs associated with low proficiency in communication include difficulty in communicating with external clients or vendors, which leads to negative consequences such as restricting the range of customers, suppliers, and other business partners. Additional costs include miscommunication among employees collaborating on projects as well as a reduced ability to transfer knowledge across organizational units and to expand international networks (Neeley, 2012; Piekkari, 2006). In addition, employees may fail to develop a sense of corporate identity, or to identify with the company for which they work (Ojanperä, 2014).

For instructors and their institutions, improving workplace preparation is important so that educational institutions such as high schools, vocational schools, and community colleges remain valuable avenues for formal training to prepare students for careers that require postsecondary education degrees. In turn, the benefits of improved student preparation are increased workplace competence, greater productivity, and more opportunity for advancement and growth (College Board, 2004).

Improved preparation in communication is needed in academic settings to infuse instruction with the needed opportunities for students to demonstrate competency in WEC (Oliveri & Tannenbaum, 2019). Avenues for improving preparation in WEC may include strategies such as providing students with opportunities to learn an array of communication forms that are connected to the workplace and providing them with more frequent feedback on their use of the new forms of communication. Moreover, it may include using assessments for learning that provide more targeted feedback and leverage digital technology to integrate digital tasks into classroom learning (Hundleby & Allen, 2017).

2.2 Expanding Integrated Forms of Assessment to Support Instruction of WEC

In this section, we focus on the assessment aspect of improving WEC preparation. We start with a recap of how a sociocognitive perspective on learning and assessment moves beyond the familiar dichotomy of formative and summative assessment. (For a fuller discussion, see Oliveri, Mislevy, & Slomp, 2021, this issue; Oliveri, Slomp, et al., 2021, this issue; and Oliveri, Slomp, Elliot, et al., 2021, this issue.) Then, we elaborate on our description of the WEC construct.

2.2.1 Forms of Assessment

The design processes that guide our development of the WEC assessments call for robust construct articulation at the earliest stages of design. The need for high degrees of alignment between the assessment design and the construct measured is even greater in the contexts of formative assessments where the information produced by the assessment is used to teach individual students. In common usage, formative assessments provide feedback to guide students' learning along the way, while summative assessments provide information outside the learning context, for purposes such as evaluating student outcomes or surveying achievement in groups of students.

As we describe in Oliveri, Mislevy, and Slomp (2021, this issue), we suggest surpassing binaries that polarize formative and summative assessments and to focus on the commonalities between these two forms of assessment. The key distinctions as to purpose and integration with learning remain important, but other secondary characteristics associated with the terms can be seen to share common features. It is often thought that formative assessment, but not summative, provides contextualized feedback to individuals that is immediately useful for learning; tasks can be richer and reporting can be more individualized, although results are hard to compare across tasks, times, or places. In contrast, it is thought that summative assessment, but not formative, provides information that is more comparable across tasks, times, or places by way of a common reporting framework; these properties are achieved by means of more complex modeling procedures and potentially lead to tasks that are less authentic and less connected with individuals' conditions of learning.

2.2.2 WEC Construct

Corrigan and Slomp (2021, this issue) articulate in detail the construct model (their Figure 2) that informs the design of our WEC modules. The focus of this construct model is on seven domains of expertise required for success in workplace communication. These domains include metacognitive knowledge, critical discourse knowledge, discourse community knowledge, rhetorical aim knowledge, genre knowledge, substantive knowledge, and communication task process knowledge. Drawing on a new-literacies orientation (Leu et al., 2016), this construct model embeds web and technology-based knowledge and skills within each domain.

Drawing on a sociocognitive framework, Oliveri, Mislevy, and Slomp (2021, this issue) use this model to frame communication as a problem-solving exercise: As people take up

communication tasks in new contexts and discourse communities, they need to learn how to acquire and utilize the patterns of language use that define success within these rhetorical situations. Our WEC modules are designed to provide learners with information on how well they are acquiring and utilizing these patterns of language, and to provide guidance on the learning path that will lead to greater success, resilience, and independence with respect to the varied communication tasks students will face in the workplace.

As discussed more fully in Oliveri, Mislevy, and Slomp (2021, this issue), we can now start from a more precise understanding of the nature of the constructs of interest, of the situations in which people use the corresponding knowledge and skills, the aspects of performance drawn upon in such situations, and the evidence needed to make inferences about the development of the corresponding capabilities. This understanding is the common foundation of integrated forms of instruction and assessment, from which we can draw on the full range of techniques to design activities and analyses that emphasize aspects of learning and of who needs what information, when, and in what form.

A given contextualized activity may seem the same on its surface, but serve instructional, formative, summative, or a combination of purposes, depending on contextualization, analytics and reporting, and integration with learning. The examples discussed below and those further illustrated with our WEC prototype illustrate how this can happen.

2.2.3 Examples of Digital Tasks

With these concepts in mind, we consider recent examples of digital tasks that focus on expanded literacy forms. One example is the *Online Research Comprehension Assessment* (ORCA) developed by the New Literacies Research Lab at the University of Connecticut (Leu et al., 2009). The ORCA provides students with a searchable online platform for acquiring and evaluating source material to be used in completing online writing tasks. Another example is Alberta, Canada's *Student Learning Assessment* (SLA) program, which provides students with a digital portal within which they read and listen to source material as they answer a series of questions by clicking on responses or dragging textboxes (Alberta Education, 2021).

As compared to paper-and-pencil assessments, these types of assessments present items and content in more immersive, engaging, and authentic ways. However, they are not without challenges, such as increased complexities in scoring more complex data and increased potential for sources of construct-irrelevant variance (e.g., understanding digital interfaces, interacting with additional stimuli; Oliveri et al., 2019). Assessments that seamlessly integrate learning and assessment may overcome some of the challenges mentioned.

2.2.4 Integrating Learning and Assessment

Andrade et al. (2019) propose that better connections between assessment and instruction can have a positive impact on learning by providing feedback mechanisms, learning strategies, and information about students' strengths. Teachers may use such information to support students'

learning by incorporating notions that learning is an active process of constructing knowledge and making sense, and by providing feedback to students about their learning processes.

Integrated (formative) assessments may provide essential sources of information that may be used by teachers to guide their instructional practices, examine their own instructional strategies, and obtain information about students' knowledge and understanding. Then, teachers may then feed *back* to students, parents, or administrators among others for progress monitoring. Increased use of formative assessments would be beneficial to inform decision making at school, district, or state levels.

At school levels, these decisions may include identifying curricular areas of strength and others needing improvement as well as making judgments about teaching quality. At the district and state levels, these decisions may entail evaluating school effectiveness, allocating resources to schools or programs, and making long-term curriculum plans. An expanded use of formative assessments employed within a learning framework also would represent an important shift in the literacy assessment culture.

This expansion in the form of assessment is responsive to research conducted over the past two decades that champions the role of formative assessment in promoting learning and achievement gains in schools (Chappuis et al., 2009; Popham, 2008; Stiggins, 1999). Many of the claims for the utility of formative assessments are based on Black and Wiliam's (1998) seminal research, although it should be noted that this work was later critiqued by Bennett (2011), who observed that such claims were often overstated and not supported by rigorous empirical research.

Minimizing the challenges of using assessments for learning and augmenting their benefits requires richer understanding of principles of evidence-based design for effective use in a blended learning curriculum to enhance the utility for users including students and teachers. In this article, we suggest that the use of a framework such as e-ECD can guide such design efforts in principled ways. The framework allows for the joint consideration of key elements that relate to evidentiary elements relevant to the construct, aspects of learning and assessment as well as measurement models associated with making sense of the evidence for meaningful interpretation of students' results. Consequently, in what follows, we discuss the e-ECD framework in relation to an expanded WEC construct and integrated learning-assessment approach.

3.0 The Extended Evidence-Centered Design Framework

Recently, Arieli-Attali et al. (2019) published the e-ECD framework, which builds on the oft cited *Evidence-Centered Design* (ECD) framework (Mislevy et al., 2003). In the context of large-scale assessments, ECD recently has been formally utilized in the development of assessments such as the *Test of English for International Communication* (Schmidgall et al., 2019), although there are numerous examples in the literature referencing ECD for assessment design, development, and implementation purposes.

Informed by principles of ECD, e-ECD extends the typical cross-sectional perspective on assessment information (i.e., designing, delivering, analyzing, and reporting on a single assessment at one point in time to capture proficiency states and provide summative feedback) to a longitudinal perspective on learning (i.e., designing, delivering, analyzing, and reporting on multiple learning activities and assessments over time to capture learning gains and provide formative feedback).

Formally, e-ECD is a principled design approach that helps identify, map, and categorize activity patterns associated with a particular context or practice to render test takers' implicit behaviors and attitudes observable and assessable in an operational assessment. As with ECD, the expanded framework allows for the design of systems for learning that are principled, valid, and learner-focused. As with ECD, the process includes identifying the *knowledge*, *skills*, *and attitudes* (KSAs) that should be assessed, the tasks that could be designed to collect the associated evidence of the possession of the targeted construct elements, and the evidence used to evaluate the KSAs.

Key questions raised by e-ECD include the following:

- What are the central aspects of the construct targeted for assessment?
- What types of tasks / activities could be used to assess the KSAs?
- What types of measurement / learning analytics models could be used to assess the KSAs?
- What kinds of interpretations could be made from behavioral observations and model-based summaries?
- What kinds of misconceptions might be identified in advance so that feedback mechanisms and learning aids are provided to students to provide guidance for stakeholder actions?

In line with these questions, we call out three extended dimensions and associated foci of the e-ECD framework that set it apart from ECD: (a) learning and assessment (the conceptual and construct-related elements of students and targeted proficiencies over time), (b) digital instructional content (the behavioral tasks that students must complete and the designs used to teach and assess them), and (c) measurement models for learning (the statistical components and the evidence they provide, such as the scoring rules used to provide automated feedback to students). We discuss these three dimensions in the following subsections and connect them to WEC training.

3.1 e-ECD Dimension 1: Learning and Assessment

This dimension focuses on expected changes of skills and proficiency over time (i.e., over learning occasions). With regards to teaching and assessing WEC, to develop greater WEC competency, we suggest that pedagogy and assessment may benefit from developing the problem-solving, analytic, and metacognitive strategies needed to (a) recognize critical features

of a communicative situation and ways to act/respond within each of those situations in relation to audience, genre, and context of the problem presented; (b) identify parallels between that situation and previous ones in which interactants engaged or exemplars presented in earlier situations; and (c) enact a plan to apply that analysis to the completion of new tasks in new contexts. For WEC learning and assessment, these strategies translate into

- focusing on providing learners with opportunities to problem-solve the communicative situation while using analytic and metacognitive strategies,
- presenting learners with situated and contextualized (case) problems to engage with, and
- exposing learners to more authentic workplace-like situations.

We elaborate on each of these strategies next.

3.1.1 Problem-Solving, Analytic, and Metacognitive Strategies

We posit that efficiency in WEC involves the ability to problem-solve, analyze the communicative situation, and use metacognitive strategies and skills to identify potential solutions. To illustrate, early research by Flower and Hayes (1980, 1981) in the area of writing demonstrated that, when faced with a novel writing task, less proficient writers often assign themselves the wrong problem or task. Less proficient communicators tend to use previous tasks as a template, however imperfect, for completing a subsequent task. This approach assumes that the qualities that constitute good writing are both stable and universal.

More recent work by Hayes and his colleagues (Leijten et al., 2014) furthermore reveals that construct models for writing are related to the sites where writing takes place. As they remind us, a writer in the workplace creates documents and utilizes resources in different ways than do students. As such, existing writing models, which are commonly derived from student writing in the academy, must be expanded to include writing in nonacademic contexts. Another implication is to focus on the role of metacognition in the assessment, as Flower (2017) observed that 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.

3.1.2 Situated and Contextualized Communication

Another aspect of learning we advance is that, due to the deeply contextual nature of communication, there is no universal set of standards that define what "good" communication is. Instead, we put forth that communication is situated, and the qualities that define effective communication vary in relation to features of the communicative situation. Understanding that communication is situated and contextualized is important as it relates to *teaching-for-transfer*, that is, the capacity to take knowledge of writing developed in one context and apply it or

repurpose it to a new context. This notion is important because it informs our assessment models and the types of scenarios we develop, as we illustrate later.

Increases in the complexities of a 21st century knowledge-based economy has led to an associated rise in the number of roles, genres, modalities, audiences, and contexts that students need to engage with to communicate effectively in the workplace. As a result, it appears very difficult within the academy to prepare future employees for all specific sets of tasks and communicative contexts they will face in the workplace. Thus, the true challenge of preparing employees for the contemporary workplace is providing them with the KSAs needed to independently problem-solve the ever-evolving communication tasks they are faced with (Leu et al., 2016), which requires targeted training.

3.1.3 Engagement with More Authentic (Workplace-Like) Contexts

We support the notion that increased digitization creates opportunities to bring learning about the real world to the classroom and to narrow the disjuncture between the kind of learning that occurs in the classroom as compared to the more hands-on and applied learning that often occurs in the workplace. As some researchers argue, classroom learning is somewhat more theoretical, generic, and teacher-driven, while the learning of workplace skills and their workplace contextualization is more focused on specific work outcomes as they apply to a unique work context. The latter context requires employees to interact with more diverse audiences than in a traditional classroom context and to follow the specific norms, protocols, and general company procedures (Dias et al., 2013). The use of digital training materials for WEC, such as the authentic digital modules mentioned at the outset, supports a closer alignment of training workplace contexts; we discuss this in more detail in the following subsection.

3.2 e-ECD Dimension 2: Aspects of Interactive and Digital Instructional Content

Interactive and digital instructional content should be carefully designed to support students' learning. This support may include (1) curating teaching and learning content by designing and integrating teaching and learning systems, (2) providing immersive experiences based on participative modeling practices, or (3) using simulations or scenario-based assessments. We start our discussion by providing a more general description of the benefits of using an integrated learning and assessment system in a blended learning environment and then discuss the types of digital environments that may be used to teach and assess WEC.

3.2.1 Integrated Learning and Assessment Systems Within a Blended Learning Approach

Advances in digitization create opportunities for better integrating teaching, learning, and assessment—a practice that is well supported by research (e.g., Bennett, 2014; Gordon Commission, 2013; Partnership for 21st Century Skills, 2008; Pellegrino et al., 2001). For instance, the National Research Council (2012) articulates the need for developing richer assessments that measure learning outcomes and student progress, inform instruction, give

students feedback on their learning, and provide models for instruction. Wylie (2017) and Wylie and Lyon (2017) similarly highlight the advantages of formative, ongoing classroom assessments that are well integrated within high-quality classroom instruction to improve discipline-specific learning outcomes and encourage students to become increasingly independent. A critical component of formative assessment is ongoing feedback to students about their learning, which involves evaluating students' work more frequently (e.g., Shepard et al., 2005; Shute, 2008; Shute et al., 2008).

Now emerging (but still limited in number) are comprehensive and engaging learning systems with embedded assessments for WEC that go beyond selected-response questions. Several candidate systems are currently available in the educational technology ("edtech") market but are still relatively limited from a metacognitive or diagnostic perspective. For example, they often do not contain automated real-time feedback mechanisms for *constructed responses* (e.g., short or extended responses using written, spoken, or multiple modalities jointly) and do not provide students with opportunities for deeper engagement with the constructs through the inclusion of metacognitive self-reflection questions. This is due, in part, to the technological development efforts that are required to create self-contained systems with a variety of key desired functionalities (see Yan et al., 2020). Integrated teaching, learning, and assessment modules with automated feedback mechanisms for WEC training present several attractive opportunities for students and teachers. For example, it is time-consuming for instructors to evaluate written compositions, drafts, or planning documents for every student while planning and adapting lessons on-the-fly as the needs of students become apparent. The development of digitally delivered training environments with integrated real-time feedback that is adaptive to students' needs can help alleviate teachers' workloads while providing diagnostic feedback to students as they learn. These kinds of systems contrast with more traditional forms of assessments that often fail to provide opportunities for students to deepen their ability to think critically, problem-solve meaningfully, and communicate interactively, or they may have limited washback (Oliveri & Wendler, 2017).

With respect to language assessment specifically, Martin (2004) illustrates how tests used for language instruction may have *negative washback* (i.e., the effects of testing on curriculum design, student learning behaviors, and teaching practices such as teaching to the test or students learning only material that will be on the test). Specifically, Martin describes the struggles of English language students who, despite the resources invested in learning English, are only able to engage in marginal dialogs and interactive communication with English speakers regardless of the time, effort, and money they spent on studying English.

Martin (2004) describes that such failures in interactive communication stem from decontextualized language instruction focused on learning language rules and conventions. Instead, it would be more desirable to learn language in interactive ways that involve speaking in diverse contexts and situations, interacting with a wider range of genres in more authentic ways. This is true for general language instruction and WEC skill instruction, which should include

formative learning and assessment opportunities in authentic, scenario-based simulations, as we discuss next.

3.2.2 Participative Modeling

Given the importance of WEC skills (Oliveri et al., 2017), their malleability (Aguinis & Kraiger, 2009), and trainability (e.g., Arthur et al., 2003; Delise et al., 2010; Salas et al., 2008), various organizational training programs have been developed to teach WEC to compensate for insufficient academy-based training. In a study of training program effectiveness, Arthur et al. (2003) discuss the ways in which programs differ along various dimensions such as their design and their predominant inclusion of immersive experiences, lectures, or opportunities to discuss, reflect on, and experience the use of WEC. The authors also note variations in delivery methods such as lectures, discussion, self-instruction, audiovisual, programmed, or computer-assisted instruction.

Dell'Aquila et al. (2015) suggest other dimensions of variability including a focus on *symbolic* (e.g., lectures and presentations) or participative modeling processes such as role plays, with the former ranking lower in cognitive and behavioral involvement. Study results suggest that immersive experiences can transform the way people think and how they engage with the materials taught. Moreover, people become motivated to learn on their own when they can relate to what is presented through more participative approaches such as through the use of simulations. For instance, de Freitas and Oliver (2006) and Michael and Chen (2005) discuss that more effective workplace training programs allow students to engage with real-world ("authentic-like") situations using active methods such as group discussions, interactive exercises, role-play simulations, and case studies. Such programs typically include a focus on attitudinal and cognitive processes through simulated interactions, metacognitive and reflection exercises, and (a)synchronous feedback from the digital system itself as well as teachers or peers.

Effective systems often support "learning-by-doing" practice opportunities and ongoing constructive feedback (see Bennett, 2016). Advantages of such immersive teaching and learning approaches involve their potential intrinsic power for engaging and motivating students. They also allow for broader, less superficial coverage of the constructs to better support students in developing skills related to "thinking like" the professionals they will become upon entry to the workplace and the completion of their studies.

3.2.3 Scenarios and Simulation-Based Learning and Assessment Activities

O'Reilly and Sabatini (2013) note that contextually-integrated scenarios offer benefits such as the ability to present students with plausible purposes for interacting with thematically related sources composed of different exemplars (e.g., webpages, emails, or blogs), which can be used to expose students to a range of genres in the workplace, for instance. These design features can facilitate the development of an expanded literacy construct to go beyond traditional assessments. Hauck et al. (2013) assert that scenario-based modules allow for the assessment of

language in more authentic situations through the inclusion of avatars that communicate with specific audiences and purposes.

Deane et al. (2015) describe how the use of scenarios can help increase student engagement and motivation by depicting realistic contexts and purposes that may lead to positive washback effects. This can be accomplished when instructional strategies are used to create deeper and structured opportunities for assessment by breaking down complex tasks into more manageable steps and supporting learning using scaffolds. Kolb (2014) and Woo and Reeves (2007) suggest that advances in technology and the development of online scenarios enable more meaningful learning opportunities, which allow individuals to interact around complex problems in more authentic and interesting environments with increased levels of engagement.

3.3 e-ECD Dimension 3: Measurement Models for Learning and Assessment

With respect to WEC, developments have often focused on conceptual models of the skills and subskills rather than their empirical evaluation. Customarily, such evaluation focuses on more traditional forms of communication such as essay writing. As an alternative, we suggest a focus on using a *sociocognitive perspective* (Mislevy, 2018; Mislevy & Oliveri, 2019), which recognizes that *linguistic, cultural, and substantive patterns* (LCS) shape how we approach communicative situations, how we understand or interpret those situations, and how we act within those situations.

3.3.1 LCS Patterns

This technical term refers to regularities in the way people act in different situations, both behaviorally and communicatively. These regularities are shaped both by factors internal to the person (e.g., general cognitive ability, general language ability, situation-specific motivation, situation-specific subject-matter knowledge) and external to the person (e.g., cultural norms, communicative norms, situational norms). Paying attention to LCS patterns is thus akin to fine-tuning one's understanding about how a person acts through deeply contextualized analyses of the driving factors behind the actions (see Oliveri, Mislevy, & Slomp, 2021, this issue).

The use of LCS patterns for assessing students' responses to WEC is important because each of the communicative activities students engage in when interacting with various workplace communicative activities requires them to understand LCS variations associated with each genre (form) of communication. The advantages of using an approach based on LCS patterns for analyzing and modeling communication is that for each individual, layers of LCS patterns differentially shape understandings of each communicative situation.

On the one hand, LCS patterns determine underlying understandings and expectations of what is appropriate within each communicative situation that are common to people sharing similar cultural, social, linguistic, and temporal contexts. On the other hand, subgroups within these broad contexts establish their own set of LCS patterns that uniquely shape underlying understandings and expectations of what is appropriate within each communicative situation

common to these groups. Additionally, each individual's unique experiences shape their access to both broader and subgroup LCS patterns, and to the expectations and understandings that accompany them.

We argue that key to developing transferable communicative competence and enhancing fairness and opportunity to learn for diverse populations is the ability to experience and analyze LCS patterns and to extrapolate from them principles for action in new situations with populations of diverse backgrounds. Beyond LCS patterns, we also underscore the evidentiary-reasoning benefits of derivatives of multidimensional design tables or extended Q-matrices, which can be used to capture the relationship between the design features and expected statistical properties of the measurement opportunities for the underlying WEC skills that are provided during learning (see Oliveri, Mislevy, & Slomp, 2021, this issue).

3.3.2 Q-Matrices and Data Cubes

In its simplest form, a prototypical Q-matrix is a two-dimensional relational table that has one row for each discrete *measurement opportunity* (e.g., a test question or item in a traditional assessment or a log file sequence for a digital activity) and a column for each *unobserved (latent) characteristic* of the learner that can be measured; these are typically the KSAs we mentioned at the outset. Each learner characteristic typically corresponds to a variable in a measurement model that is designed to capture inter-individual variation. This can be done dichotomously (e.g., 0-1 / low-high / "fails standard" - "meets standard") or polytomously (e.g., 0-1-2 / low-medium-high proficiency / successive levels in a learning progression) for criterion-referenced classification purposes as well as approximately continuously (i.e., using a real number scale) for norm-referenced comparison purposes.

Cell entries in the Q-matrix reflect the degree to which each measurement opportunity provides information about each learner characteristic. In the simplest Q-matrices, these entries are binary (i.e., 1-0 / yes-no) reflecting that a learner characteristic is measured / required, but they could also be polytomous or real-valued, reflecting a minimum required level of a characteristic (e.g., Ding et al., 2016; von Davier, 2005). Observable learner performances are then scored dichotomously (e.g., 0-1 / incorrect-correct) or polytomously (e.g., 0-1-2 / incorrect-partially correct-correct), depending on the task type and the degree to which evidence can be reliably captured. A prototypical Q-matrix thus assumes a one-to-one relationship between measurement opportunities, learner characteristics, and observable performance.

Adopting a sociocognitive perspective encourages us to consider, and subsequently encode, the influence of LCS patterns more carefully. It has several implications for the design architecture of Q-matrices. For example, Q-matrices can be extended to capture additional task design variables that were inspired by the underlying design patterns for the learning tasks. This information can be used to sensitize task designers to the design choices they make, to evaluate the mixture of design features across a set of activities, and to statistically model the relationship

between the design features and observable task metrics (e.g., difficulty, discrimination) from a measurement model.

Rather than assuming a global one-to-one relationship between these design features for all learners, under a sociocognitive perspective, it would likely be more appropriate to include different feature value combinations for learners from different subgroups. The most meaningful definition of these subgroups should be derived from an understanding of how the characteristics of learners within each subgroup interact with the design features of the measurement opportunity (e.g., shared learning experiences, micro-cultural backgrounds, or personal histories), rather than be determined through coarse groupings that exist mostly for legal purposes (e.g., ethnicity or gender).

As a result of a sociocognitive perspective that considers LCS patterns, a simple two-dimensional representation of a Q-matrix is no longer the most effective choice. Instead, we are faced with the inherent multidimensionality of a data cube (e.g., von Davier et al., 2019) where the Q-matrix part of the data cube functions like a design cube within the larger data architecture. This is best represented through software architectures that use appropriate *metatags* for all aspects of the data, which facilitates querying of the data cube at multiple levels to inspect and utilize the interaction of task design characteristics, learner characteristics, and scoring characteristics.

This helps to connect these augmented Q-matrix data cube components with other parts of the data cube in data warehouses. For example, if a particular Q-matrix instantiation is viewed as specific to a particular task, activity, or scenario, then a sociocognitive perspective would suggest the existence of alternative, parallel tasks, activities, or scenarios in a data warehouse with associated Q-matrices from which the desired variants for a particular implementation could be pulled.

Importantly, however, a Q-matrix does not prescribe a particular measurement model that an analyst should use. This aligns relatively naturally with current thinking around modern measurement models, which are perhaps best viewed as part of a componential evidence accumulation machinery fueled by appropriate model construction given the data complexities at hand. The design decision of which psychometric modeling family is most appropriate for a given use context typically depends on both the properties of the data and the stakeholders' needs.

Given the complexity of the interactions involved in this general modeling setup, a flexible measurement modeling approach would likely be required to perform robust model-data fit evaluations at scale. Possible modeling families might include Bayesian inference networks (e.g., Almond et al., 2015), diagnostic classification models (e.g., Rupp et al., 2010; von Davier & Lee, 2019), or mixture item response theory models (e.g., von Davier, 2008; von Davier & Yamamoto, 2004). This also underscores a common tension in modern measurement modeling—more fine-tuned, contextually-sensitive modeling approaches are generally data-hungry and thus most appropriate for applications at large scales.

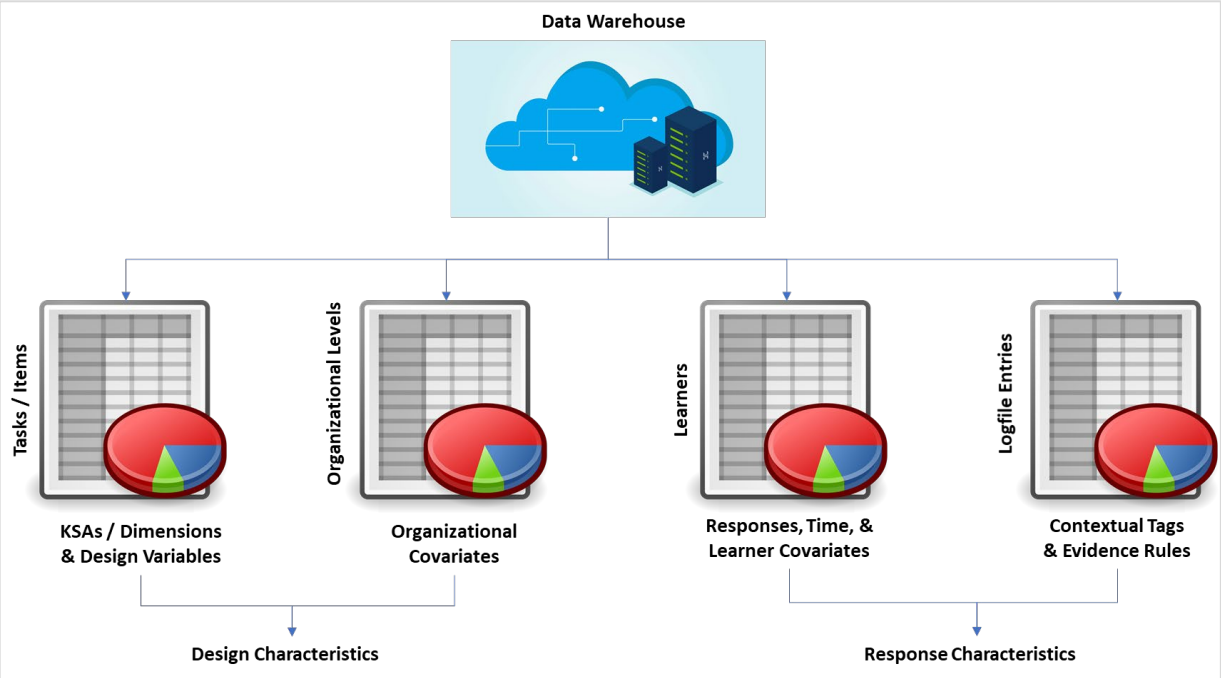
Utilizing a sociocognitive perspective around WEC using these kinds of architecturally-instantiated Q-matrices implies that variants of the contextualized scenarios would have to be created through the instantiation of global design patterns and associated design variables. It means that a close collaboration with instructors and learners is necessary to better understand which learner subgroups are most relevant to define and what the influence of the scenario design features on the activation and utilization of the LCS patterns is and how these interactions can be leveraged to create robust scoring protocols.

Figure 1 illustrates the conceptual idea of a multidimensional data warehouse from which different kinds of representations can be pulled with appropriate queries, each one representing different kinds of activities and considerations during the design, modeling, and reporting phases. From left to right, we have four representations displaying four different sets of relationships that are commonly used in modeling that would be important for doing sociocognitively-grounded WEC work.

First, we have a table showing the relationship between items / tasks / scenarios and the core skills that they are intended to assess along with the key features / variables that guided their design. This is useful for determining the relationship between the construct model and the task model in particular. It is frequently referred to as the augmented Q-matrix noted above. Second, we have a table showing the relationship between organizational layers of the data (e.g., classrooms, schools, or districts) and the characteristics that distinguish them. This information is useful for performing appropriate explanatory analyses that explain performance differences at different levels of aggregation. The third table shows the relationship between learners and items / tasks / scenarios (or parts of these), which is typically how we think of a response file for straightforward analyses. The fourth table shows the relationship between logfile entries, either for an individual learner or multiple (stacked) learners, and contextual information for some (or all) entries in the logfile that can be used by algorithms to create evidence rules (i.e., new indicators based on observable patterns). These four tables are illustrative, and other combinations or structures can be queried from the warehouse under a proper setup.

Figure 1

Common Two-Dimensional Representations as Derivatives from Data Warehouse Architecture



It is in this complex space of multiple interactions that measurement approaches would have to be designed or selected in order to accumulate the evidence. Moreover, if automated scoring routines for complex performances or automated detectors for emotional states were to be created, these indicators might also have to be included in such models as contextual variables. Most importantly, meaningful modeling decisions would involve understanding the design, uses, and interpretations of the data, and the contexts in which the data will be used and analyzed, which are better attained through dialogs with stakeholders comprising the user ecosystem as Zapata-Rivera et al. elaborate in their article (2021, this issue).

3.4 An Application of the e-ECD Framework to the WEC Modules

In the previous subsections, we described the e-ECD framework and its dimensions with respect to a more comprehensive WEC construct. In what follows, we illustrate its application to our WEC prototype module that targets email writing, which remains a widely used form of communication at work despite the surge of real-time communication tools such as Slack, GitHub, or Microsoft Teams. A more comprehensive capturing of the WEC construct would require further design and development efforts than the ones we have advanced thus far in our research.

Table 1 lists the dimensions of the e-ECD framework, provides a description of the components we included in WEC, and lists associated examples. We highlight our emphasis on expanding the construct traditionally taught for WEC through two design features: (1) the use of scenarios to help increase student engagement for learning WEC and (2) the analysis of LCS patterns to provide feedback to students on their use of language in various workplace situations to communicate with diverse audiences and carry out diverse communicative tasks.

Table 1

Dimensions of e-ECD and Their Instantiation in WEC Modules

Name/Label	Description of components included	Examples
Learning and assessment (the conceptual and construct-related elements of students and targeted proficiencies over time)	Inclusion of underrepresented workplace-relevant genres of communication, metacognition, and problem-solving opportunities	The modules include emails as a form of workplace communication.
Digital instructional content (the behavioral tasks that students must complete and the designs used to teach and assess them)	Use of a scenario-based approach: Scenarios augment the level of authenticity and verisimilitude of the modules to workplace environments	The modules enable students to interact with different avatars such as a virtual boss, coworkers, clients, and a mentor.
Measurement models for learning (Statistical)	Inclusion of automated feedback and use of LCS patterns for coding students' responses	Automated feedback in some items is immediate and actionable. Rubrics are designed to capture meaningful variation in the use of LCS patterns by communicative context; feedback is provided to students on their use of the patterns. Q-matrices capture the key skills measured.

3.4.1 Learning and Assessment: Modules Teach and Assess an Expanded WEC Construct

In the kind of WEC training we envision, students engage in various work-related problems through simulated business situations and engage in completing key communicative activities. The communication activities and genres included in the modules were informed by previous research conducted by Oliveri and McCulla (2019) based on data from the *Occupational Network* (O*NET) database. The O*NET is the United States' primary source of occupational information, which contains information on over 900 jobs; see also an elaboration of the construct in Corrigan and Slomp (2019) and Beaufort (1999, 2008).

The authors discuss key construct elements of WEC including linguistic knowledge (i.e., rhetorical knowledge, which entails comprehending the audience and purpose of the

communication, and genre knowledge, or the types of features to use depending on the form of communication used). It also includes metacognitive knowledge (i.e., awareness about one's understanding of their own responses and one's ability to articulate this understanding). Moreover, it includes substantive knowledge, or the ability to understand the content of the communication and the type and level of complexity of specialized vocabulary needed in the communication.

This dimension involves making choices related to which type of vocabulary to use, depending, for example, on the extent to which the communication is directed to a technical or non-technical audience. Further, this dimension involves understanding the content of the communication, the level of detail that should be used to convey the information, and the context and organizational structure in which the communication occurs. Learners are able to use this context to understand, incorporate, and respond to the information conveyed in the communication in ways that are appropriate for the task at hand. These subdimensions are all inherently metacognitive in nature, providing a framework for learners to analyze, respond to, and monitor their completion of communication tasks.

3.4.2 Digital Instructional Content

Gorin and Mislevy (2013) note that there is limited value and utility associated with decontextualized assessments that seem to "drop-from-the-sky" and instead, have advocated the use of more contextualized tasks that more closely align with classroom instruction. Using digital technology in a blended environment may be helpful for students who are in remote locations or who are underserved students with limited access to project-based activities or reduced access to individualized tutoring or other learning supports (see Mehta, 2014). That being said, it is important to recognize that without a carefully designed blended learning approach there is a parallel risk of having "drop-from-the-sky" digital activities with similar limitations to decontextualized assessments.

In Oliveri, Mislevy, and Slomp (2021, this issue), we elaborate on the various design choices made in the design of the WEC modules we mentioned before. Here, we briefly note that we used a scenario-based approach to increase the level of authenticity of the modules to workplace environments and more finely evaluate uses of LCS patterns. Specifically, the modules expose students to a fictitious company and ask them to use emails to communicate with virtual coworkers, their boss, virtual clients, and a mentor to provide students with opportunities to practice email writing. Students play the role of a project manager at a medium-sized company that designs and sells modular kitchens to home owners, construction companies, and home improvement stores. As the project manager, students are charged with coordinating the activities of three equal-ranking virtual teammates who work together to increase the company's visibility and growth.

Beyond allowing us to analyze contextualized writing, these situations are used to raise students' awareness of the types of workplace environments they may eventually face, and thus

empower them to see themselves in careers and professional situations that may be unfamiliar to them. This objective requires a careful design of scenario variants that reflect the realistic and aspirational goals of the diverse learner populations we aim to support, which we discuss in more detail in the next sections. Moreover, to better support learners, the modules include learning resources such as checklists to help them identify, master, and evaluate their own writing.

3.4.3 Measurement Models for Learning

The design of WEC's rubrics is critical to capture meaningful variation in students' uses of LCS patterns. Paying attention to LCS patterns means developing differentiated and situated ways to score students' responses and providing more nuanced feedback on particular LCS patterns across different communicative situations, as well as asking specific metacognitive questions about the drivers behind their use of LCS patterns. This means, for example, that it is necessary to evaluate the extent to which students' LCS patterns in a particular communicative context (e.g., writing an email to a particular boss in which a particular request is made) are aligned with the broader communicative conventions of the associated genre as well as the more specific communicative conventions that are grounded in their cultural or personal histories.

For our sample WEC module prototype, a simple, prototypical Q-matrix was created to explicate the link between items/tasks (the stimuli used to collect students' responses) and the elements of the measured constructs. Learning supports were also integrated into the modules, and the prototype included worked-out examples, automated feedback that was dependent upon students' responses, and opportunities for students to revise their work.

To conclude this section, we assert that increased digitization has various implications, including the need to expand the skillset targeted for assessment; the potential for enabling opportunities to better integrate teaching, learning, and assessment; and the increasing possibilities for teaching and learning using contextualized scenarios. These scenarios can show students more authentic and real-life situations through avatars or models of more effective instruction. We suggest that to fully benefit from such affordances and to mitigate any potential negative unintended consequences (e.g., narrowing the curriculum, teaching to the test), we need to analyze potential affordances considering a ToA.

4.0 Theory of Action

We now discuss the ToA and structure this section in two parts. In the first part, we describe general aspects of a ToA such as its purpose, its components, and the general questions asked in ToAs. In the second part, we apply the ToA to WEC and describe the module components and the short- and long-term objectives for teachers and students who use them.

4.1 General Aspects of a ToA

At a high level, a ToA provides empirically testable claims related to the interpretation and use of information from a particular educational program or social service, connecting program

actions with program outcomes (Morell, 2005). Moreover, a ToA helps identify where in the assessment process unintended, negative consequences may occur and what may be done to address them. For decades, ToAs have been conceptualized and applied to organizational change efforts and to the evaluation of social and educational programs (e.g., Foster, 1972; Patton, 1978; Slomp & Elliot, in press).

A ToA specifies through graphical and textual modes the intended relationships between the components associated with the implementation of an assessment program and the desired outcomes of that implementation. Bennett (2010) notes that the ToA clarifies the stakeholders involved in the assessment and the associated actions to be taken by the stakeholders. More specifically, it describes the hypothesized relationships between program components, hypothesized action mechanisms, intended intermediate effects, and intended ultimate effects.

Applied to assessment, a ToA is a mechanism that can be used to identify the intended aims of an assessment program and trace the mechanisms through which those envisioned aims will be realized. A ToA, then, provides a blueprint for a program of research on the intended and unintended outcomes of an assessment program focused on the proposed action mechanisms. At a basic level, one asks a set of generalized research questions such as

1. What are the program elements included in the assessment?
2. Have the hypothesized action mechanisms been identified? What factors (ecological, interpersonal, intrapersonal) are supporting or creating barriers to the effective implementation of the hypothesized action mechanisms? What unanticipated action mechanisms are being developed and implemented in response to these factors?
3. What are the intermediate outcomes of both the hypothesized and unanticipated action mechanisms?
4. What are the long-term outcomes of both the hypothesized and unanticipated action mechanisms?

The use of the ToA requires us to formalize and make explicit expectations and assumptions about the test and testing outcomes, and, by doing so, presents opportunities to consider counterfactuals and unintended, negative consequences. In the absence of a ToA, testing issues may be too abstract or open to too much unnecessary variability, preventing us from considering unintended consequences systematically.

Examples of a ToA include the ToA designed for the *Cognitively Based Assessment of, for, and as Learning* (CBAL). CBAL is intended to be used to document student achievement (of learning), inform instructional planning (for learning), and serve as a worthwhile educational experience for students and teachers in and of itself (as learning). CBAL assesses reading, writing, and mathematics for middle-school students. The ToA that Bennett (2010) exemplifies has four aspects (program components, hypothesized action mechanisms, intended intermediate, and intended ultimate effects on test users). The program components typically include the test, its scores, and services designed for both test takers and score users. The action mechanisms capture the types of decisions, behaviors, and solutions expected from different stakeholders

using the scores and other information from the assessment program. The last two aspects indicate the theorized impact directly occurring from program use and the ultimate intended impact of the assessment program, which captures the societal values and goals as outlined in Messick (1989). Arrows within ToA models capture expected connections between each of the columns, which may be considered as claims.

4.2 A ToA for the WEC Modules

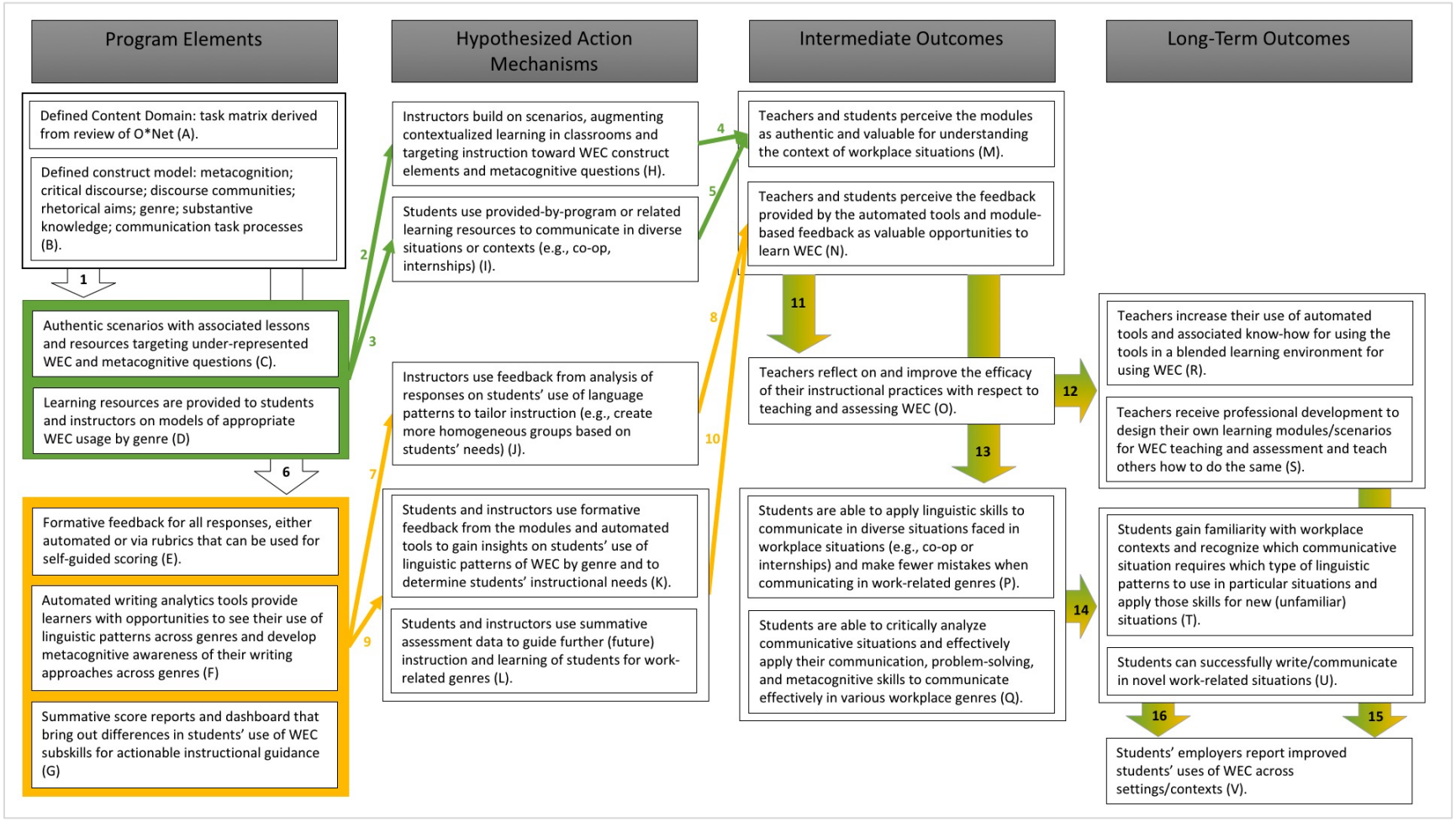
In alignment with the previous subsection, our initial articulation of a ToA for digital WEC training modules describes cause-effect relationships among (a) program elements, (b) hypothesized action mechanisms, (c) intermediate outcomes, and (d) long-term outcomes for our stakeholders. We pay particular attention to the interconnected perspectives of fairness and consequences for diverse populations. The action mechanisms describe (or initially hypothesize) what underlying mechanisms are causing the intermediate effects.

As shown in Figure 2, the overall goals of our modules within a blended learning approach are intended to (a) better support student learning by providing more frequent and rapid feedback around WEC construct elements, (b) empower students to envision themselves as successful communicators in novel career pathways, (c) support teachers' evaluations of their instructional efficacy, and (d) encourage both teachers and students to incorporate more of these kinds of activities into the classroom or out-of-classroom learning experiences. The overall goals are advanced by a defined construct model used to inform tasks and provide structured feedback.

Specifically, Figure 2 lists, from left to right, the module (program) components, the hypothesized learning outcomes, and the intermediate and long-term outcomes. The second column lists hypothesized actions to use the materials and program elements listed in the first column. The third and fourth columns list the intermediate and long-term outcomes for the two primary stakeholders (i.e., students and teachers). In the following section, we walk through the overall logic of the ToA.

Figure 2

Theory of Action for WEC



4.2.1 Program Components

The first column in Figure 2 lists the program components, which include the following:

- *defined content domain* (Box A), which includes a task matrix developed from the review of the O*Net database and associated literature;
- *defined construct model* (Box B), which includes these variables: metacognition, critical discourse, discourse communities, rhetorical aims, genre, substantive knowledge, and communication task processes;
- *authentic scenarios* (Box C), which include a six-part story, supporting lessons for students and instructors, and resources for students and instructors to use to support student learning;
- *learning resources* (Box D), which provide examples and learning resources to support student learning to identify ways in which they can improve their learning;
- *formative feedback* (Box E), which is provided for all responses, either automatically or via an expert-derived scoring rubric that teachers and students can use;
- *automated writing evaluation tools* (Box F), which incorporate the use of automated evaluation writing tools (e.g., DocuScope; see Wetzel et al., 2021, this issue) to support student learning and the provision of (close to) real-time feedback on how students communicate;
- *summative (prospective) score reports* (Box G), which are designed to help students, teachers, and administrators learn about performance and engage metacognitively around the response patterns.

4.2.2 Hypothesized Action Mechanisms

The second column in Figure 2 lists the hypothesized action mechanisms, which include the following:

- *Authentic scenarios* (Box H). We hypothesize that the instructors will find the digitally administered modules to represent authentic workplace scenarios and that the modules will provide students with additional opportunities to learn WEC using additional contextualized learning materials in a blended learning environment. Instructors will either develop their own materials or use the materials accompanying the digitally administered modules and related metacognitive questions to support student learning of WEC.
- *Use of learning resources* (Box I). Students will use learning resources and examples to increase proficiency in their communication in diverse situations or contexts (e.g., co-op, internships).

- *Use of feedback to inform instruction* (Box J). Instructors will use feedback from their analysis of their students' responses on the use of language patterns for each genre to inform their instructional decisions such as the creation of more homogeneous ability groups or to focus on particular teaching, learning, and assessment of specific subskills.
- *Formative feedback and reflection questions* (Box K). We hypothesize that instructors and students will use the formative feedback provided in the modules and provided through the use of automated writing tools to gain increased awareness of how students are using language patterns by genre and to identify the extent to which those uses are (in)consistent with traditional uses of genre-related features. To the extent that such uses are inconsistent with the genre, teachers will develop additional opportunities for students to interact with the genres. Along these lines, they will also use metacognitive reflection questions to help students reflect on their uses of the language and provide students with opportunities to self-correct based on an enriched understanding of the examined genres, audiences, and context of the communicative activity.
- *Summative feedback* (Box L). Students and instructors will use summative assessment data and score reports that will provide information on students' use of language patterns per dimension to guide further (future) instruction and learning of students for work-related genres.

The starting point for design of the WEC prototype is (A) a well-defined construct model. The goal of our program is that students will be able to critically analyze communicative situations and effectively apply their insights to novel situations because the modules contain authentic scenarios (C) and formative feedback (E, F, G) for students and teachers, and provide students with opportunities to engage in metacognitive thinking to gain awareness of their writing approaches (K, L).

4.2.3 Intermediate and Long-Term Outcomes

The third and fourth columns—Box M through Box V—list the hypothesized intermediate and long-term outcomes of the program. Examples include perceptions of the modules and the associated effects, teachers' reflections and actions with respect to the program, and the tracking of students' progress, each of which we elaborate on next.

4.2.3.1 Module Perceptions. Our expectation is that both teachers and students will perceive these modules as authentic (M) and that they will perceive the feedback provided by the modules and the associated automated writing evaluation tools as useful and valuable (N). This expectation is driven primarily by the use of authentic scenarios that target key WEC skills (C). This perception should lead to increased success in workplace situations when students (employees) are asked to engage in diverse workplace activities, which may lead to greater student agency, better problem-solving strategies, and increased understanding of ways to

address complex communication situations occurring at work (P, Q). These gains may be achieved through increased use of learning resources provided by the program or their instructors (D). Examples of how these relationships can be investigated include conducting interviews and surveys with both teachers and students, and tracking the enacted curriculum of teachers, to the degree that this is possible.

4.2.3.2 Teacher Reflections on Training and Instruction. Our expectation is that, through our modules, teachers will be able to more effectively and, perhaps more frequently, reflect on the efficacy of their instructional practices (O). This expectation is supported by the use of actionable summative score reports that effectively bring out distinctions in student performance on the targeted skills and subskills (G). This should then lead to improved student learning outcomes over time (T, U) and may lead teachers to change their instructional practice to incorporate a different mix of scenario-based learning and assessment activities in the classroom (R, S). Example ways in which these relationships can be investigated include conducting interviews and surveys with employers (V), as well as tracking student performance over time across modules or with independently-vetted external assessments, and tracking the enacted curriculum of teachers, to the degree that this is possible.

4.2.3.3 Student Tracking. Our expectation is that, through our modules, students will be able to track their progress more effectively and, perhaps, more frequently. This is supported through both actionable formative feedback (E, F) and summative score reports with at least partially dynamic dashboards (G).

These abilities should then lead to improved learning outcomes as well as success in workplace situations with respect to the genres students have engaged with in the classroom (T) as well as potentially novel genres they may not have yet encountered but that are relevant to carrying out their employment activities (U). Examples of how these relationships can be investigated include evaluating students' responses on the modules and their interactions with and use of automated writing tools (e.g., DocuScope). These relationships may be investigated by administering surveys to students, teachers, or employers regarding students' WEC proficiency, and by interviewing students about their workplace experiences using WEC and their career choices as well as empirically tracking these experiences, as possible.

4.3 Logic of the ToA Relationships between Elements, Action Mechanisms, and Outcomes

In Figure 2, we present the logic of our ToA, describing the envisioned links between elements, action mechanisms, and outcomes. In the narrative that follows, letters contained within parentheses correspond to boxes shown in Figure 2. Numbers contained in parentheses correspond to linking arrows across the boxes shown in Figure 2.

The starting point for our ToA is a well-defined construct model (A) and clearly articulated content domain that (1) guides the development of scenario-based tasks and resources for use in the classroom (C, D) and that (6) shapes the design of analytic strategies, formative and summative score reports (E, F, G). These scenarios are (2, 3) supplied to teachers (H) and

students (I) who use them to guide instruction and learning. Through interactions with the WEC modules, both instructors and students (4, 5) appreciate the authentic nature of the tasks and their value for promoting workplace situated learning.

The data presented in the formative and summative score reports and the writing analytics reports (E, F, G) will be presented (7) to teachers (J) to guide instruction. Our expectation is that (8) experience with feedback derived from the WEC modules and associated instructional materials will be deemed valuable by teachers and students (N).

Feedback derived from WEC modules (D, E, F) will also be used to (9) focus student learning (K, L). As teachers and students (10) work with the feedback derived from the WEC modules, they recognize their value in promoting competence and independence with workplace communication tasks (N).

Recognizing the value of the WEC modules and associated materials will (11) spur teachers to reflect upon and improve their instructional practices related to teaching workplace communication skills (O). Over time (12), teachers will gain confidence in utilizing a sociocognitive model of writing instruction and will seek out and lead professional education experiences to further support their teaching in this area (R, S).

Students' positive experiences with the WEC modules will similarly (13) drive their learning and achievement in the area of workplace communication across genres in simulated workplace environments (P, Q). Given the WEC module's focus on developing transferable knowledge through the use of metacognitive reflection tasks and heuristic-based instruction, students will (14) gain confidence, competence, and resilience in workplace communication (T, U).

As teacher (15) and student (16) competence in the teaching and learning of workplace communication increase, employers will notice an improvement in student capacity and competence with workplace communication (V), the ultimate goal driving the development of our WEC modules.

As we noted above, an important strength of the ToA is that it lays bare the assumptions driving the design of an assessment program, along with the accompanying Interpretation and Use argument necessary to support the ethical use of that assessment program. Simultaneously, it presents a program of research needed to interrogate the intended and unintended outcomes from the use of the assessment program as each of the assumptions in the ToA need to be tested. With respect to the WEC prototype, the key assumptions that need to be tested include the following:

- Instructors and students will value the scenario-based learning and assessment.
- Instructors and students will value the formative, summative, and analytic reports.
- The ecological environment in which teachers are working enables and supports the kinds of professional autonomy required and the professional learning envisioned (Marynowski et al., 2019).
- Ecological, interpersonal, and intrapersonal factors that shape student development support student ability to acquire and transfer the workplace communication abilities envisioned in the construct model (Slomp, 2012).

Because teaching and learning is such a complex undertaking, this program of research is necessary to distinguish what factors beyond the design and implementation of the WEC prototype itself will help to achieve the envisioned goals of this program of work. Because we worked with the initial phases of the prototype, our goal was to clearly outline intended consequences while being mindful of unintended consequences (e.g., the design of modules with diverse populations in mind, the use of language that was accessible for high school students and beyond). These issues are further discussed in other articles in the special issue; for instance, see Oliveri et al. (2021, this issue) and Oliveri, Mislevy, and Slomp (2021, this issue).

5.0 Conclusion

In this article, we first reviewed the impact of Industry 4.0 on the need to expand the concept of communication, with particular attention to WEC. We also illustrated conceptual (i.e., e-ECD and ToA) frameworks and discussed their application to the design of the WEC modules. In our review, we paid particular attention to language models, formative assessments (or assessments for learning), and a scenario-based framing. We described the use of the frameworks for the design of digital modules within a blended learning approach. During the discussion, we noted key expected relationships and sketched out, at a very high level, mechanisms for collecting evidence to evaluate these relationships.

This article is one tile in the mosaic that this special issue of *The Journal of Writing Analytics* constructs. None of the articles in the special issue is by itself a traditional hypothesis-methods-findings research paper. Rather, together they address interwoven elements in our deliberation on a single overarching question—a question that is central, we believe, to the future of educational assessment. As we have discussed in this article, and our colleagues have further elaborated in the companion articles, there are revolutionary developments at once in several areas that are integral to learning and assessment. These include demands for higher-level skills, in more complex worlds of work; diversity in learner populations and life-long learning; advances in understanding of social, cognitive, and situative nature of learning and acting; and technologies for interactive digital environments for learning and assessment, for using massive data to provide personalized feedback, and for integrating instruction and assessment anytime and anywhere. Each advance by itself pushes beyond the boundaries of familiar forms, indeed conceptions, of instruction and assessment.

We are therefore led to an overarching question: *How can we develop coordinated systems for learning and assessment in this new age?*

The key, we believe, is identifying design frameworks that address capabilities, instruction, assessment, and learners, all coordinated to the end of maximizing every individual's opportunities to learn. Each article in the special issue highlights certain frameworks, discussing them more fully and illustrating them in the context of WEC—several with the prototype modules that this article and several others use, and others with closely related examples. Together, they constitute an attempt at an existence proof as our answer: Here, we believe, is an

(not the) approach that builds on recent research in assessment design, program evaluation, writing studies, and instructional design. The contribution is to explain the frameworks, describe the interconnections, and demonstrate their value by constructing an artifact that incorporates insights and methods across the multiple diverse areas that are implicated.

The three foci of the present article in our program of research—an overview of WEC, the e-ECD framework for assessment design in learning environments, and the ToA framework—are intended to be integrative. WEC offers fruitful opportunities as well as challenges, for it is both complex and highly valued, rich and multidimensional in its activities, a target of social and cognitive research, an area in which natural language and corpus analysis researchers have tackled, and inherently and increasingly diverse in its real-world manifestations.

The great value of the ToA framework is that it concisely brings out the underlying who, what, why, how, and what-can-go-wrong arguments for a system of instruction and assessment. Note that our walk through the ToA for the WEC modules pulls in, at a higher level, all of the recent developments mentioned above, but more importantly, relates and coordinates them in our rationale for the socially contextualized use.

The fact that the ToA is a contextual rather than a technical framework is important for three reasons. First, it is accessible across all the users and all the specialists whose needs and knowledge need to come together in the complex artifact that is envisioned. Second, sketching out a ToA begins early in the process, long before the assessment is given to a single student. Designing a complex artifact with constraints coming from different directions is always a method of successive approximations, as designers work through tradeoffs across the technical areas that are involved, the needs of stakeholders, and the resources and constraints that cut across seemingly unrelated facets of the enterprise. Note in particular the attention to designing in flexible ways for design options across diverse learners and for generating new modules in the same framework. The ToA is a framework for everyone involved, enabling collaboration around these issues in shared representations and language. Third, the ToA provides a starting point for specialists to communicate contextual factors related to assessment use, assessment purposes, and the mechanisms and actions needed for assessment use.

It is the last of these ToA roles that the e-ECD assessment design framework takes up, with regard specifically to assessment design. Its structures are extensions of the ECD assessment design framework to the learning framework—a framework which integrates the analysis of the WEC domain with assessment and learning modules' contents and forms, structures and activities, automated analyses and measurement models, and feedback and reporting as it is needed by different users and in different modes of use. Here, the contextual factors expressed in the ToA are now blueprinted in the e-ECD framework. Note, for example, how the higher-level ToA requirement of flexibility for diverse learners and contexts is translated into more technical specifications for adaptability in task design, evaluation and feedback routines, and measurement models. Fairness in this sense is designed in from the start rather than after-the-fact modifications of a single artifact. Focusing on fairness and opportunity to learn throughout the

entire design and implementation process helps ensure that assessment data supports all learners while simultaneously working to identify and mitigate unintended negative outcomes stemming from the development of an assessment program. In this way, an Integrated Design and Assessment Framework (IDAF) design process integrates well with the ToA framework: ToA makes explicit the intended outcomes of an assessment program, along with the action mechanisms and design features hypothesized to achieve those outcomes, while IDAF provides an integrated process for interrogating those assumptions (Slomp & Elliot, in press). These topics are further explored in Oliveri, Slomp, et al. (2021, this issue) and Oliveri, Mislevy, and Slomp. (2021, this issue)

The ToA and e-ECD frameworks are connected to the IDAF framework, which considers consequences as central to the design, implementation, and validation process for such systems. Oversimplifying somewhat, we can view the different frameworks as nested in terms of the scope of questions they address. Specifically, the e-ECD framework is most critically concerned with the core learning and assessment machinery. The IDAF framework suffuses the entire work process within this framework with deeper considerations around validity, fairness, and consequences. Accordingly, the ToA framework zooms out even further and considers systemic effects around implementation claims.

Beyond elaborating these frameworks and applying them to other prototypes and assessments of 21st century skills, we hold that future research is needed on forms of assessment that (a) facilitate better connections between assessment and learning, (b) support decision-making across levels of the education system, and (c) support creation of meaningful strategies for monitoring outcomes and establishing accountability. These approaches have the potential to better serve classroom learning, provide data on valued learning outcomes, and facilitate long-term positive effects in classroom and school teaching practices in terms of the KSAs required for students to thrive in the fourth industrial revolution.

Author Biographies

María Elena Oliveri is Associate Director of Psychometric Consulting and Research Associate Professor at Buros Testing Center, University of Nebraska – Lincoln.

David Slomp is Associate Professor of Education at the University of Lethbridge.

André A. Rupp is founder of Mindful Measurement and Leader of Instructional Data at Success Academy Charter Schools.

Robert Mislevy holds the Frederic M. Lord Chair in Measurement and Statistics at Educational Testing Service, and is Emeritus Professor of Measurement, Statistics, and Evaluation at the University of Maryland.

References

Aguinis, H., & Kraiger, K. (2009). Benefits of training and development for individuals and teams, organizations, and society. *Annual Review of Psychology*, 60, 451-474. https://doi.org/10.1146/annurev.psych.60.110707.163505

Alberta Education. (2021). *Student Learning Assessments.* https://www.alberta.ca/student-learning-assessments.aspx#:~:text=Student%20Learning%20Assessments%20(SLAs)%20are%20digital%20tests%20administered,SLAs%20contain%204%20components:%20Literacy%20%E2%80%93%20Digital%20questions

Almond, R. G., Mislevy, R. J., Steinberg, L., Yan, D., & Williamson, D. (2015). *Bayesian networks in educational assessment*. Springer.

Andrade, H. L., Bennett, R. E., & Cizek, G. J. (Eds.). (2019). *Handbook of formative assessment in the disciplines*. Routledge.

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*, 1-17. https://doi.org/10.3389/fpsyg.2019.00853

Arthur, W., Jr., Bennett, W., Jr., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology*, *88*, 234-245. https://doi.org/10.1037/0021-9010.88.2.234

Beaufort, A. (1999). *Writing in the real world: Making the transition from school to work*. Teachers College Press.

Beaufort, A. (2008). Writing in the professions. In C. Bazerman (Ed.), *Handbook of research on writing: History, society, school, individual, text* (pp. 221-235). L. Erlbaum Associates.

Bennett, R. E. (2010). Cognitively Based Assessment of, for, and as Learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research and Perspectives*, *8*(2-3), 70-91. https://doi.org/10.1080/15366367.2010.508686

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Policy and Practice*, *18*(1), 5-25. https://doi.org/10.1080/0969594X.2010.513678

Bennett, R. E. (2014). Preparing for the future: What educational assessment must do. *Teachers College Record*, *116*(11). http://www.tcrecord.org/Content.asp?ContentID=17623

Bennett, R. E. (2016). *Opt out: An examination of issues* (Research Report RR-16-13). Educational Testing Service. https://doi.org/10.1002/ets2.12101

Bhatia, V. K., & Bremner, S. (2012). English for business communication. *Language Teaching*, *45*(4), 410-445. https://doi.org/10.1017/S0261444812000171

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7-74.

Casner-Lotto, J., & Barrington, L. (2006). *Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants into the 21st century workforce*. Conference Board, Partnership for 21st Century Skills, Corporate Voices for Working Families, and Society for Human Resource Management. https://files.eric.ed.gov/fulltext/ED519465.pdf

Chappuis, S., Chappuis, J., & Stiggins, R. (2009). The quest for quality. *Educational Leadership*, *67*(3), 14-19. http://www.siprep.org/uploaded/ProfessionalDevelopment/CDRP/QualityAsmnt.pdf

College Board. (2004). *Writing: A ticket to work…or a ticket out: A survey of business leaders* (Report of the National Commission on Writing for America's Families, Schools, and Colleges).https://archive.nwp.org/cs/public/download/nwp_file/21479/writing-ticket-to-work.pdf?x-r=pcfile_d

Corrigan, J. A., & Slomp, D. H. (2019, May). *Articulating a bioecological model for writing development* [Paper presentation]. The Canadian Society for the Study of Education XLVII Annual Conference, Vancouver, BC, Canada.

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*.

Deane, P., Sabatini, J., Feng, G., Sparks, J., Song, Y., Fowles, M., O'Reilly, T., Jueds, K., Krovetz, R., & Foley, C. (2015). *Key practices in the English language arts (ELA): Linking learning theory, assessment, and instruction* (ETS Research Report No. RR-15-17). Educational Testing Service. https://doi.org/10.1002/ets2.12063

de Freitas, S., & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education*, *46*(3), 249-264. https://doi.org/10.1016/j.compedu.2005.11.007

Delise, L. A., Gorman, C. A., Brooks, A. M., Rentsch, J. R., & Steele-Johnson, D. (2010). The effects of team training on team outcomes: A meta-analysis. *Performance Improvement Quarterly*, *22*, 53-80. https://doi.org/10.1002/piq.20068

Dell'Aquila, E., Marocco, D., Ponticorvo, M., Di Ferdinando, A., Schembri, M., & Miglino, O. (2015). *Educational games for soft-skills training in digital environments: New perspectives*. Springer.

Dias, P., Freedman, A., Medway, P., & Paré, A. (2013). *Worlds apart: Acting and writing in academic and workplace contexts*. Routledge.

Ding, S., Luo, F., Wang W., & Xiong J. (2016). Dichotomous and polytomous Q matrix theory. In L. van der Ark, D. Bolt, W.C. Wang, J. Douglas, & M. Wiberg (Eds.), *Quantitative psychology research* (pp. 277-289). Springer. https://doi.org/10.1007/978-3-319-38759-8_21

Dziuban, C., Graham, C. R., Moskal, P. D., Norbert, A., & Sicilia, N. (2018). Blended learning: The new normal and emerging technologies. *International Journal of Educational Technology in Higher Education*, *15*(3). https://doi.org/10.1186/s41239-017-0087-5

Flower, L. (2017). Reflection: What can cognitive rhetoric offer us? In P. Portanova, J. M. Rifenburg, & D. Roen (Eds.), *Contemporary perspectives on cognition and writin*g (pp. 331-345). WAC Clearinghouse.

Flower, L., & Hayes, J. R. (1980). The cognition of discovery: Defining a rhetorical problem. *College Composition and Communication*, *31*(1), 21-32. https://doi.org/10.2307/356630

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*(4), 365-387. https://doi.org/10.2307/356600

Foster, M. (1972). An introduction to the theory and practice of action research in work organizations. *Human Relations*, *25*(6), 529-556. https://doi.org/10.1177%2F001872677202500605

Gallup & Lumina. (2014). *The 2013 Lumina study of the American public's opinion on higher education and U.S. business leaders poll on higher education: What America needs to know about higher education redesign*. Gallup. https://www.luminafoundation.org/resources/what-america-needs-to-know-about-higher-education-redesign

Gordon Commission. (2013). *To assess, to teach, to learn: A vision for the future of assessment* (Technical Report). https://www.ets.org/Media/Research/pdf/technical_report_executive_summary.pdf

Gorin, J. S., & Mislevy, R. (2013, Sept. 23-24). *Inherent measurement challenges in the next generation science standards for both formative and summative assessment* [Paper Presentation]. Invitational Research Symposium on Science Assessment, Washington, DC, United States. https://www.ets.org/Media/Research/pdf/gorin-mislevy.pdf

Hart Research Associates. (2015). *Falling short? College learning and career success: Selected findings from online surveys of employers and college students conducted on behalf of the Association of American Colleges & Universities*. https://www.aacu.org/sites/default/files/files/LEAP/2015employerstudentsurvey.pdf

Hauck, M. C., Wolf, M. K., & Mislevy, R. (2013). *Creating a next generation system of K-12 English learner (EL) language proficiency assessments*. Educational Testing Service. https://www.ets.org/s/research/pdf/24473_K12_EL_Paper.pdf

Hundleby, M., & Allen, J. (Eds.). (2017). *Assessment in technical and professional communication*. Routledge.

Kirsch, I., Braun, H., Yamamoto, K., & Sum, A. (2007). *America's perfect storm: Three forces changing our nation's future*. Educational Testing Service. https://www.ets.org/Media/Education_Topics/pdf/AmericasPerfectStorm.pdf

Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*. FT Press.

LeFebvre, M., & Clark, H. (2013). *The condition of work readiness in the United States*. ACT. https://www.act.org/content/dam/act/unsecured/documents/ConditionWorkReadiness.pdf

Leijten, M., Van Waes, L., Schriver, K., & Hayes, J. R. (2014). Writing in the workplace: Constructing documents using multiple digital sources. *Journal of Writing Research*, *5*(3), 285-337. https://doi.org/10.17239/jowr-2014.05.03.3

Leu, D. J., Kulikowich, J., Sedransk, N., & Coiro, J. (2009). *Assessing online reading comprehension: The ORCA Project*. www.orca.uconn.edu/ orca-project/project-overview/

Leu, D. J., Slomp, D., Zawilinski, L., & Corrigan, J. A. (2016). Writing research through a new literacies lens. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 41-53). Guilford.

Martin, A. (2004). The "katakana effect" and teaching English in Japan. *English Today*, *77*(20), 50-55. https://doi.org/10.1017/S0266078404001087

Marynowski, R., Mombourquette, C., & Slomp, D. (2019). Using highly effective student assessment practices as the impetus for school change: A bioecological case study. *Leadership and Policy in Schools*, *18*(1), 117-137.

Mehta, J. (2014). Deeper learning has a race problem. *Education Week*. http://blogs.edweek.org/edweek/learning_deeply/2014/06/deeper_learning_has_a_race_problem.html

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-104). American Council on Education; Macmillan.

Michael, D., & Chen, S. L. (2005). *Proof of learning: Assessment in serious games*. http://www.gamasutra.com/view/feature/130843/proof_of_learning_assessment_in_.php

Michael, D. R., & Chen, S. L. (2005). *Serious games: Games that educate, train, and inform*. Course Technology.

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

Mislevy, R., Almond, R. G., & Lucas, 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., & Oliveri, M. E. (2019). Sociocognitive assessment for diverse populations (Digital ITEMS Module 09). *Educational Measurement: Issues & Practices*, *38*(4), 110-111. https://ncme.elevate.commpartners.com

Morell, J. A. (2005). Why are there unintended consequences of program action, and what are the implications for doing evaluation? *American Journal of Evaluation*, *26*, 444-463. https://doi.org/10.1177/1098214005281354

Moss, P. A., Pullin, D. C., Gee, J. P., Haertel, E. H., & Young, L. J. (Eds.). (2008). *Assessment, equity, and opportunity to learn*. Cambridge University Press.

National Association of Colleges and Employers (NACE). (2018). *Job outlook 2019*.

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. The National Academies Press. https://www.nap.edu/read/13398/chapter/1

Neeley, T. (2012). Global business speaks English. *Harvard Business Review*. https://hbr.org/2012/05/global-business-speaks-english

Ojanperä, M. (2014). *Effects of using English in business communication in Japanese-based multinational corporations* [Master's thesis, University of Oulu]. http://jultika.oulu.fi/files/nbnfioulu-201402131106.pdf

Oliveri, M. E., Lawless, R. R., Molloy, H. (2017). *A review of collaborative problem solving* (ETS Research Report No. RR-17-06; GRE Board Research Report No. GREB-17-03). Educational Testing Service.

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., & McCulla, L. (2019). *Using the Occupational Network Database to assess and improve English language communication for the workplace* (Research Report No. RR-19-28). Educational Testing Service.

Oliveri, M. E., Mislevy, R., & Elliot, N. (2020). After admissions: What comes next in higher education? 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. (2021). Principled development of workplace English communication Part 3: An integrated design and appraisal framework. *The Journal of Writing Analytics*, *5*.

Oliveri, M. E., & Tannenbaum, R. J. (2019). Are we teaching and assessing the relevant English skills for success in the international workplace? In V. H. Kenon & S. V. Palsole (Eds.), *The Wiley handbook of global workplace learning* (pp. 343-354). Wiley. https://doi.org/10.1002/9781119227793.ch19

Oliveri, M. E., & Wendler, C. W. (2017, April). *Enhancing the validity argument of assessments: Identifying, understanding, and mitigating unintended consequences of test use* [Professional development workshop]. Annual meeting of the American Educational Research Association, San Antonio, TX, United States.

O'Reilly, T., & Sabatini, J. (2013). *Reading for understanding: How performance moderators and scenarios impact assessment design* (ETS Research Report RR-13-31). Educational Testing Service. http://dx.doi.org/10.1002/j.2333-8504.2013.tb02338.x

Ortiz, L. A., & MacDermott, C. (2017). Beyond the business communication course: A historical perspective of the where, why, and how of soft skills development and job readiness for business graduates. *The IUP Journal of Soft Skills*, *XI*(2), 7-24. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3215098

Partnership for 21st Century Skills. (2008). *21st century skills, education, & competitiveness: A resource and policy guide*. http://www.p21.org/storage/documents/21st_century_skills_education_and_competitiveness_guide.pdf

Patton, M. Q. (1978). *Utilization-focused evaluation*. Sage Publications.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. National Academy Press.

Piekkari, R. (2006). Language effects in multinational corporations: A review from an international human resource management perspective. In G. K. Stahl & I. Björkman (Eds.), *Handbook of research in international human resource management* (pp. 536-550). Edward Elgar. https://doi.org/10.4337/9781845428235.00038

Popham, W. J. (2008). *Transformative assessment*. Association for Supervision and Curriculum Development.

Rupp, A. A., Templin, J., & Henson, R. (2010). *Diagnostic measurement: Frameworks, methods, and applications*. Guilford Press.

Salas, E., Diaz Granados, D., Klein, C., Burke, C. S., Stagl, K. C., Goodwin, G. F., & Halpin, S. M. (2008). Does team training improve team performance? A meta-analysis. *Human Factors: Journal of the Human Factors and Ergonomics Society*, *50*, 903-933. https://doi.org/10.1518/001872008X375009

Schmidgall, J., Oliveri, M. E., Duke, T., & Carter Grissom, E. (2019). Justifying the construct definition for a new language proficiency assessment: The redesigned TOEIC Bridge® tests—Framework paper (Research Report No. RR-19-30). Educational Testing Service. https://doi.org/10.1002/ets2.12267

Schwab, K. (2017). *The fourth industrial revolution*. Currency.

Shepard, L., Hammerness, K., Darling-Hammond, L., & Rust, F. (2005). Assessment. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 275-326). Jossey-Bass.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153-189. https://doi.org/10.3102/0034654307313795

Shute, V. J., Hansen, E. G., & Almond, R. G. (2008). You can't fatten a hog by weighing it—or can you? Evaluating an assessment for learning system called ACED. *International Journal of Artificial Intelligence and Education*, *18*(4), 289-316. http://myweb.fsu.edu/vshute/pdf/shute%202008_a.pdf

Slomp, D. H. (2012). Challenges in assessing the development of writing ability: Theories, constructs and methods. *Assessing Writing*, *17*(2), 81-91.

Slomp, D., Corrigan, J., & Sugimoto, T. (2014). A framework for using consequential validity evidence in evaluating large-scale writing assessments. *Research in the Teaching of English*, *48*(3), 276-302. https://www.jstor.org/stable/24398680

Slomp, D., & Elliot, N. (in press). What's your theory of action? Making good trouble with literacy assessment. *Journal of Adolescent & Adult Literacy*.

Slomp, D. H., Oliveri, M. E., & Elliot. N. (2021) Afterword: Meeting the challenges of Workplace English Communication in the 21st century. *The Journal of Writing Analytics*, *5*.

Stiggins, R. J. (1999). Assessment, student confidence, and school success. *The Phi Delta Kappan*, *81*(3), 191-198.

Von Davier, A. A., Wong, P. C., Polyak, S., & Yudelson, M. (2019). *The argument for a "data cube" for large-scale psychometric data*. Frontiers in Education. https://doi.org/10.3389/feduc.2019.00071

von Davier, M. (2005). *A general diagnostic model applied to language testing data* (Research Report No. RR-05-16). Educational Testing Service.

von Davier, M. (2008). The mixture general diagnostic model. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1-370). Information Age Publishing.

von Davier, M., & Lee, Y.-S. (Eds.). (2019). *Handbook of diagnostic classification models: Models and model extensions, applications, software packages*. Springer.

von Davier, M., & Yamamoto, K. (2004). Partially observed mixtures of IRT models: An extension of the generalized partial credit model. *Applied Psychological Measurement*, *28*(6), 389-406.

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*.

Woo, Y., & Reeves, T. C. (2007). Meaningful interaction in web-based learning: A social constructivist interpretation. *The Internet and Higher Education*, *10*(1), 15-25. https://doi.org/10.1016/j.iheduc.2006.10.005

Wylie, E. C. (2017). Winsight™ Assessment System: Preliminary theory of action. *ETS Research Report Series, 2017*(1), 1-17. https://doi.org/10.1002/ets2.12155

Wylie, C., & Lyon, C. (2017). *Supporting teacher assessment literacy: A proposed sequence of learning* (Teachers College Record No. 22193). https://www.ets.org/s/winsight/pdf/balanced-assessment-article-tcr-ada.pdf

Yan, D., Rupp, A., & Foltz. P. W. (2020). *Handbook of automated scoring: Theory into practice*. Taylor and Francis.

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*.