

Principled Development of Workplace English Communication Part 3: An Integrated Design and Appraisal Framework

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:** An expanded skillset is needed to meet today's shifting workplace demands, which involve collaboration with geographically distributed multidisciplinary teams. As the nature of work changes due to increases in automation and the elevated need to work in multidisciplinary teams, enhanced visions of Workplace English Communication (WEC) are needed to communicate with diverse audiences and effectively use new technologies. Thus, WEC is ranked as one of the top five skills needed for employability. Even so, results of employer surveys report that incoming employees lack communication competency (National Association of Colleges and Employers [NACE], 2018). To address this issue, with a focus on WEC teaching and assessment, we describe a framework used to guide the design of WEC modules. We suggest that conceptual frameworks can be used to inform the design process of the module. In

this article, we illustrate one such conceptual framework: the Integrated Design and Appraisal Framework (IDAF). IDAF holds consequences of testing as one of its central elements to guide test design and development. It emphasizes categorically identifying and ecologically modeling variables impacting WEC in general and the writing context in particular. It emphasizes the need for developing clearly articulated construct models to underpin the assessment, as well as incorporating a foundational focus on fairness and social consequences into the design process and use of assessments.

- **Questions Addressed:** In reference to the IDAF, our article addresses questions in the following three areas: the nature and benefits of an integrated design and appraisal approach to test design, development, and evaluation; the application of IDAF to complex communication tasks captured in formative assessment scenario-based modules of WEC; and the paramount importance of considering fairness and social consequences in the design and use of assessments administered to diverse populations. Thus, this article elaborates on the use of the IDAF to inform the design of WEC modules by explicitly articulating the needs of the test takers, the anticipated uses of the modules, and the contexts in which the modules would be used. Our questions are designed to address increasing complexities associated with the design of complex constructs such as WEC. This article describes considerations for the development of integrated learning and assessment modules for WEC. We start by reviewing principled assessment design frameworks, which have been used to inform the development of complex tasks across disciplines or fields. Following a description of WEC in terms of domain analysis and design patterns, we illustrate the application of the IDAF to inform the design of the modules. We conclude by providing an overview of our research questions and of how our article addressed them. We also discuss lessons learned with respect to the design of the prototype and the delicate balance of engaging in a principled design process that supports goals that empower students of diverse backgrounds to learn WEC.
- **Conclusions:** This article illustrates the application of the IDAF to inform the design and development of WEC modules. 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 approaches.

Keywords: anticipatory design frameworks, digital modules, Integrated Design and Appraisal Framework (IDAF), Workplace English Communication (WEC), writing analytics

1.0 Background

There is a critical demand for the development and assessment of Workplace English Communication (WEC) across fields and institutional contexts (Oliveri & Tannenbaum, 2019). In an increasingly automated world, employees need to communicate with diverse audiences in complex ways as work activities change, new technologies emerge, and teams become more diverse (McKinsey Global Institute, 2018). These changes require increased proficiency in WEC, which involves effectively carrying out various communicative activities such as negotiating, writing requests or proposals, and conferring with clients or colleagues.

WEC—defined as a form of sophisticated discourse in which organizational and disciplinary norms for framing and communicating information are used for a variety of aims—is one of the top five most important skills needed for work. Nevertheless, results of employer surveys report that incoming employees lack communication competency (National Association of Colleges and Employers [NACE], 2018). Schriver (2012) suggests that employee under preparation in WEC can in part be explained by differences between classroom-based writing and writing in the workplace, with students taught the former in the academic context and often not the latter. Essential differences between the workplace and academic contexts are that as compared to the academic context, writing at work has higher stakes and direct consequences associated with the communication act. The workplace context requires recognizing that some of the communications by clients, supervisors, or colleagues will be directed to employees and will require follow-up actions. Moreover, Schriver indicates that “professional communication is actually situated in a much broader social practice; that is, the practice of creating cognitive structures and relational networks among people through shared content” (p. 277). She then notes that the quality of those communications is judged by how well the communications meet the needs of the audience (e.g., stakeholders). The writer is to recognize that the workplace community is influenced by the organization’s culture, power, and knowledge structure (Dias et al., 2013). Understanding differences between the workplace and academic contexts is important to the design of the WEC prototypes we describe later in this article because this understanding influences the construct definition (see Corrigan & Slomp, 2021, this issue) and design features of the modules with curricular and instructional applications.

We highlight the importance of focusing on population differences because 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). Low proficiency in communication impacts various aspects of business activities, such as difficulties communicating with external clients or vendors, which lead 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). A focus on improving

workplace readiness in communication is important for educational institutions to remain valuable avenues for formal training to prepare students for employment. In turn, the benefits of improved student preparation are increased workplace competence, greater productivity, and more opportunity for advancement and growth.

Improved preparation in communication is needed in academic settings to infuse instruction with the needed opportunities for students to demonstrate competency in WEC. Implications of improving preparation in WEC may include a variety of strategies such as providing students with opportunities to learn an array of communication forms that are connected to the workplace or 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 (Oliveri & McCulla, 2019; Oliveri & Tannenbaum, 2019).

We acknowledge that raising proficiency in WEC and designing WEC modules or training materials is neither an easy nor novel task. Previous advances have been made to teach WEC in e-learning and business communication classes provided by independent providers and on platforms such as Coursera and EdX, as well as through a few business communication games/simulations (e.g., CapSim; <https://www.capsim.com/assessments/>).

The approach we describe in the current article goes beyond these systems because it is based on a more elaborated conceptual and construct model that includes metacognitive questions that prompt students to reflect on their learning of WEC (see Corrigan & Slomp, 2021, this issue). We also use authentic-like scenario-based learning environments inspired by an earlier review of the Occupational Network database to identify the relevant WEC skills for today's workplace (Oliveri & McCulla, 2019).

Building on previous research, we discuss considerations for the design and development of integrated learning and assessment modules for WEC that focus on written communication in general and email in particular. The modules are informed by conceptual frameworks that define work-relevant aspects of communication and workplace use (Oliveri et al., 2017; Oliveri & McCulla, 2019), and the WEC construct elements elaborated in articles in this issue (see Corrigan & Slomp, 2021, this issue and Haigler, 2021, this issue).

Differences between the academic and workplace contexts present challenges for students to demonstrate the knowledge, skills, and attitudes needed to communicate at work (Hollis-Turner & Scholtz, 2010). Our goal was to provide opportunities for students to practice communication in contexts that held a verisimilitude with the workplace to support academic preparation of work-aligned communication ability (Mehta, 2014). Along these lines, in the modules, students play the role of a junior project manager in a U.S. company that specializes in the design of kitchens for corporations and private clients. Students work with a team of three (simulated) colleagues who specialize in design, budgeting, and construction processes. Students can also rely on the mentoring services of a long-term consultant and report to a senior manager for the

project. These characters are portrayed as 2D avatars, and each one has a specific role in the organization, such as a designer, project manager, and finance manager.

Throughout the modules, students engage in a variety of activities that help the company to bid on a lucrative contract for designing and showcasing a model kitchen with robotic technology and to negotiate with an external client throughout the development process upon securing the bid. Students write emails to colleagues, a mentor, and a supervisor and are provided with opportunities to draft and revise the emails. They are also provided with opportunities to explain their reasoning for how they structured their communications and why they selected particular linguistic patterns for communicating with different audiences (e.g., external clients, supervisor, or colleagues). We also included some traditional item formats such as multiple-choice and true/false, which allowed us to provide automated feedback to the students as they went through the tasks.

In what follows, we describe our design choices in developing the WEC modules. To begin, we review principled assessment design guiding frameworks that have been used to design tasks to assess complex constructs. We then present an overview of the WEC modules in terms of domain analysis and design patterns. With this overview established, we focus on one framework—the Integrated Design and Appraisal Framework (IDAF)—which we selected given our emphasis on consequential design decisions made throughout the lifecycle of assessments (e.g., Slomp, 2016; Slomp et al., 2014). Next, we illustrate our application of the IDAF to the design of the WEC modules. To conclude, we discuss lessons learned with respect to the design of the WEC prototype. Thus, by referring to the IDAF, we address questions in the following three areas: the nature and benefits of an integrated design and appraisal framework to test design, development, and evaluation; the application of IDAF to complex communication tasks captured in formative assessment scenario-based modules of WEC; and the paramount importance of considering fairness and social consequences in the design and use of assessments administered to diverse populations.

2.0 Guiding Frameworks

Assessment design frameworks guide assessment developers and researchers by helping to inform choices made when designing and developing assessments and associated instruments (Keehner et al., 2016). Nichols et al. (2016) and Ferrara et al. (2016) discuss the notion that design frameworks emphasize different aspects of the development or empirical investigation processes. It follows that framework selection processes may be carried out by aligning framework goals and the consequential aspects of the design, modeling, and analysis work that get foregrounded and backgrounded. Because the world of assessment is so diverse, spanning the assessment of multiple constructs and various purposes used for assessment, several different frameworks have been developed.

Table 1 provides an overview of principled design frameworks including focal goals of the frameworks and associated uses. Examples of design frameworks include the *cognitive design*

systems framework (Embretson 1998, 2020), which is grounded in theories of cognitive information processing and is most effectively operationalized for assessment tasks that have relatively narrow and clearly delineable cognitive operations and goals (e.g., spatial rotation tasks, mathematical computation tasks). In contrast, the *BEAR assessment system* (Wilson, 2013; Wilson & Sloane, 2000) is grounded in the learning sciences and provides a broader construct modeling approach; its focus is on modeling emerging reasoning and misconception pathways along one or multiple dimensions. Other frameworks such as *evidence-centered design* (ECD; Mislevy et al., 2003) were designed to be broadly applicable and agnostic to the particular learning and assessment solution at hand, focusing instead on broader underlying generative principles. Moreover, recognizing validity as an argument, Mislevy (2007) presents ECD as a tool for proactive assessment design that offers a design process focused on ensuring that the assessment, once designed and implemented, is able to sustain scoring, generalization (related to how well the test taker can perform from the observed sample of performance in some larger domain of tasks across occasions and conditions of observation), and extrapolation inferences (extending from observed scores to predict future performance in another context or performance domain; Kane, 2013).

Along the same line, the design-for-assessment (DFA) framework (White et al., 2015) emphasizes categorical identification and ecological modeling of variables that impact the writing context. The DFA emphasizes the need for robust domain analysis and the associated construct models that underpin the assessment. DFA extends the ECD model by incorporating a foundational focus on fairness into the design process. IDAF is similar to DFA but it takes a broader view of social consequences than does DFA. Oliveri et al. (2015) and Oliveri and Lawless (2018) demonstrate the application of Kane's (2013) interpretive-use argument model in validating assessments that have been exported for use with diverse populations. Sireci (2015) describes the application of ECD in the design of the Partnership for Assessment of Readiness for College and Careers (PARCC) assessments. Moreover, Coppola et al. (2016) report on a study that applied the DFA model for program evaluation purposes.

Table 1

Examples of Principled Design Frameworks

Name	Focal points/Goals	Focal uses
Cognitive Design Systems (Embretson, 1998, 2020)	Develop predictive models for key factors affecting complexity of cognitive tasks in a construct-validation framework.	Information-processing tasks with clearly delineable cognitive models (e.g., visual/spatial rotation, arithmetic tasks).
Assessment Engineering (Luecht, 2013)	Develop a scalable architecture for the controlled creation of task families.	Large-scale assessments centered around focal areas in science, technology, and math disciplines with (semi) constrained tasks.
BEAR Assessment System (Wilson, 2013; Wilson & Sloane, 2000)	Develop models and scales for the evolving reasoning processing in domains using stage-based developmental theories and carefully designed activities targeting key milestones and misconceptions.	Large-scale formative assessment in classrooms and activities centered around developing reasoning embodied in the practices, concepts, and core ideas of the Next Generation Science Standards.
Design for Assessment (DFA; White et al., 2015; Coppola et al., 2016)	Collect evidence for writing program evaluation related to degree program and institutional accreditation.	Targets of evidence are identified as consequence, theorization, standpoint, research, documentation, accountability, sustainability, process, and communication.
Evidence-Centered Design (ECD; Mislevy et al., 2003; Mislevy, 2018)	Develop conceptual models and aligned computational architectures for design, implementation, scoring, and reporting based on evidentiary reasoning.	All kinds of assessments at different scales and with different purposes but mostly focused on a singular implementation time point.
Extended ECD for Learning and Assessment Systems (e-ECD; Arieli-Attali et al., 2019)	Extend the ECD framework to include longitudinal considerations for design, implementation, scoring, and reporting for learning and assessment systems.	Digitally delivered learning and assessment systems focused on providing actionable instructional guidance to support learning.
Integrated Design and Appraisal Framework (IDAF; Slomp, 2016; Slomp et al., 2014)	Collect carefully curated evidence about the consequences of educational learning and assessment systems.	Digitally delivered learning and assessment systems administered to diverse populations around which fairness considerations are of critical importance.

Although we could have used various frameworks listed in Table 1 to guide this work, of the frameworks listed in Table 1, we relied on ECD (Mislevy et al., 2003) and the expanded ECD framework (e-ECD; Arieli-Attali et al., 2019). The e-ECD is used 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 (see Oliveri,



Mislevy, & Slomp, 2021, this issue and Oliveri, Slomp, et al., 2021, this issue for an elaboration of the e-ECD framework applied to WEC modules). We also used the IDAF (Slomp, 2016). We selected these frameworks because they helped us break down key design decisions into separate parts to allow for more careful deliberation of the design choices while remaining attentive to consequences of our design choices when designing assessments for diverse populations. In using the e-ECD and IDAF frameworks as the intellectual guideposts directing our mindsets and practices, we do not distance ourselves from the abovementioned frameworks since most frameworks share common foundational principles, considerations, and goals.

In what follows, we describe design patterns from ECD and our application of the IDAF framework to the design of our WEC modules. Through its attention to consequential aspects of test use, the IDAF framework complements the sociocognitive framework (see Oliveri, Mislevy, & Slomp, 2021, this issue). Together, these frameworks draw attention to key considerations of situated-use design, measurement, and analysis models. The sociocognitive framework enriches our thinking about task and population considerations called for by the IDAF, enabling more principled design of assessments for diverse populations, especially with respect to reducing construct-irrelevant variance in assessments used with these populations. Similarly, the IDAF framework complements the e-ECD and ToA frameworks (see Oliveri, Slomp, et al., 2021, this issue). The e-ECD framework enables us to expand our thinking related to integrating teaching, learning, and assessment. The ToA enables us to think through the larger context in which the assessments will be used by explicitly identifying 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. These frameworks can be used in complementary fashion to identify potential threats or issues that may arise in the design, development, and use of assessments to inform score-based decisions.

3.0 Overview of WEC Prototype and Rationale for Development

To develop the WEC prototype, we built on earlier work conducted at Educational Testing Service that developed design thinking for scenario-based assessments that informed the present work (e.g., Cognitively Based Assessment of, for, and as Learning [CBAL]; Bennett, 2010; Deane et al., 2015; Sabatini et al., 2011). We also build on advances by recent large-scale initiatives such as (a) the Common Core State Standards for English language arts (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010), (b) the Gordon Commission (2013), and (c) the Partnership for 21st Century Skills (2008, 2015). These initiatives exemplify the development of new forms of assessment and constructs including expanding the literacy construct to include interactive communication, which is relevant for college and career readiness (Oliveri & Markle, 2017).

Building on earlier research, we made decisions to focus on learning and assessment modules that integrate teaching and assessment of WEC in general and one communicative activity in particular: email writing. A scenario-based approach to assessment design was used (Bennett,

2010; Deane et al., 2015), in which the literacy activities were contextualized within a professional business setting. The design goals were to provide students with opportunities to engage with authentic texts and forms of communication in emails used for different purposes (e.g., negotiation, writing requests). The design choices we made contrast with paper-and-pencil tests that do not use contextualized forms of teaching and assessing WEC. Other forms of assessments such as situational judgment tests (Cullen et al., 2006; Lievens et al., 2005) may not be able to capture the metacognitive processes underlying the learning of WEC in developmental ways for training purposes.

3.1 Domain Analysis

A central aspect of ECD is an analysis of the targeted domain, which can be informed by various sources such as consultation with subject matter experts, a review of the literature, or a combination of both. The foundation for our domain analysis included systematic reviews of the literature related to literacy (Quigley, 2018) and writing (Corrigan & Slomp, 2021, this issue; Leu et al., 2016). Construct models identified in the literature review were then analyzed with a view to the workplace communications context. Figure 1 captures a range of these activities organized around four modalities of literacy: reading, writing, speaking, and listening. While this language arts model represents a larger construct domain, in our modules we focused on email communications only.

Figure 1

Content Domain for WEC

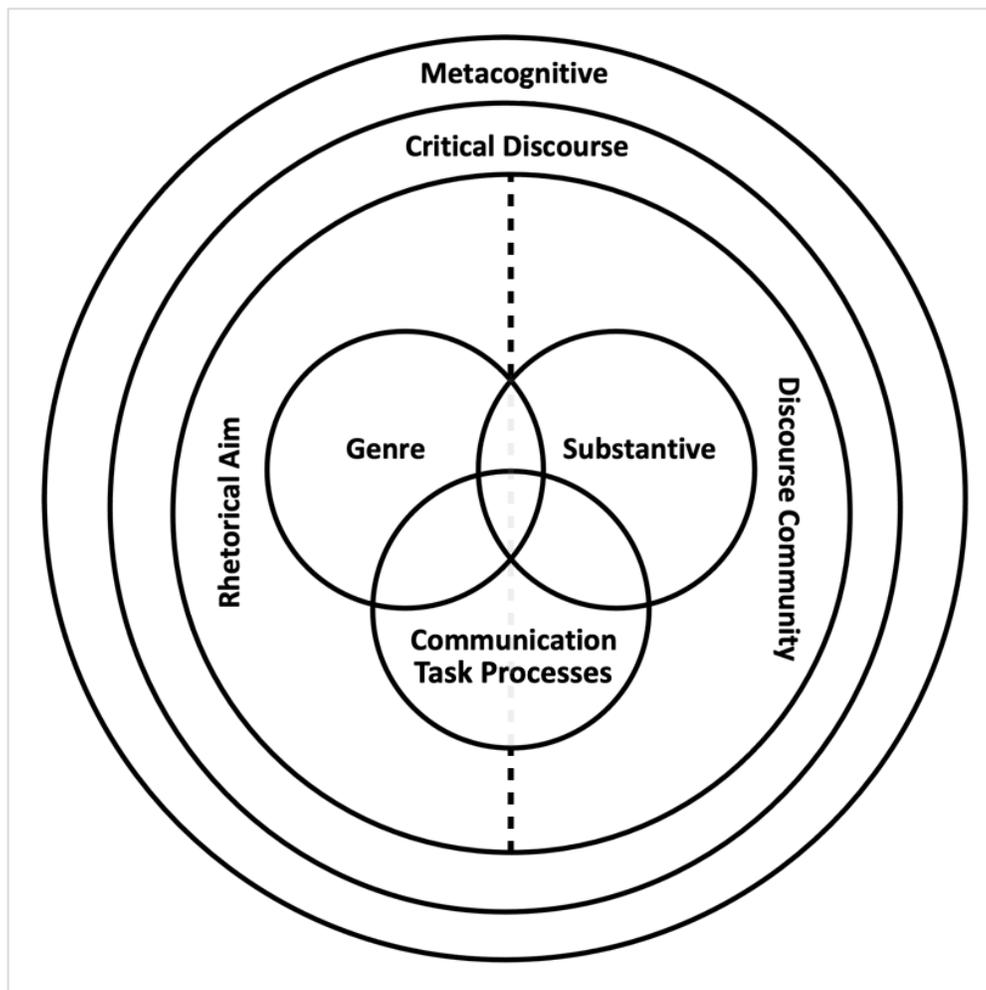


In designing assessments, it is important to move from the overarching, more expansive content model to the construct model and its operationalization. Our definition of the construct model of WEC is shown in figure 2. The figure captures our definition of WEC, which we suggest involves various forms of linguistic knowledge including rhetorical aim (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). WEC also involves having an understanding of communication task processes (related to the modality and forms of communications to use) and discourse processes (related to the type of discourse community one is communicating with, such as whether the communication is intended for a peer or colleague, supervisor, or a hierarchical or non-hierarchical institution, for instance). 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.

Figure 2

Sociocognitive Construct Model of Writing Expertise (Corrigan & Slomp, 2021, this issue)



Understanding the central elements of the construct and its forms is important to proceed to the development of tasks and assessments. This process can be facilitated by using design patterns. Design patterns provide a space for documenting the necessary elements and features of tasks designed to measure the targeted constructs (e.g., WEC), the types of performances one seeks to assess, and the types of observables sought to elicit the associated evidence. The patterns help developers identify the task features and rubrics related to the assessment argument (Liu & Haertel, 2011; Mislevy & Haertel, 2006; Riconscente et al., 2005).

3.1.1. Design Patterns

The use of design patterns facilitates common language across multidisciplinary teams tasked with designing templates to implement the design of complex tasks for complex constructs at an early stage of development. Acknowledging the choices early on is important to identify the types of issues teams of developers, designers, (technology) engineers, and user interface, user experience specialists face when designing scenario-based tasks. For instance, subject-matter experts and researchers need to work together to identify the key skills and subskills targeted for assessment while also considering the needs of user interface, user experience specialists for interpreting construct-related elements. These diverse needs have implications for the kinds of design choices that will be involved in creating the scenarios, as well as the considerations engineers need to face to create tasks that are accessible across the various kinds of technologies that may be used in diverse classrooms in which the modules will be used.

Table 2 provides an example of design patterns applied to our WEC modules. Similar patterns have been used in the design and development of complex assessments of 21st century skills (for an example application of ECD-based design patterns to the development of assessments of collaborative problem-solving skills, see Oliveri et al., 2019). Table 2 shows that each module is defined by a few fixed contextual characteristics and then allows for variations of incidental features to create variants. For example, in an email where the student is asked to make a case to the senior manager for engaging in a project, one can use the design patterns and vary associated variable features to produce task variants. In such variants, one can vary the gender and ethnicity of the manager, the power relationships in the company itself (e.g., from a hierarchical structure to more horizontal/flat management structure), and the interactional style (e.g., from a formal micro-managerial style to a more relaxed empowering style), to name a few. Providing students with such variants would enable them to practice using various features of the focal aspects of language related to the focal abilities (e.g., rhetorical aim, critical discourse) to gain metacognitive awareness of how language features vary across activities and situations. (For elaborations of this concept, see Oliveri, Mislevy, & Slomp, 2021, this issue). As we elaborate in the next section, it is important to provide students with opportunities to engage with these variable aspects of language to improve workplace readiness and preparation with regards to WEC and the various genres, discourse communities, and communication task processes they will engage with at work.

Table 2

Higher-Level Design Pattern for WEC Module

Attribute	Description
Overview	<ul style="list-style-type: none"> ✓ Students engage in a scenario in which they act as junior project managers in a design firm whose primary goal is to successfully bid on a lucrative contract for designing and displaying a model kitchen. ✓ Students use email as their main mode of communication throughout the scenario.
Focal Skills	<p>Focus on a defined construct composed of:</p> <ul style="list-style-type: none"> ✓ Metacognition ✓ Critical discourse ✓ Rhetorical aims ✓ Ability to use appropriate email conventions including greetings, closings, subject lines, and structured messages ✓ Ability to compose emails for different purposes (e.g., share and request information, ask for help, make a case) and audiences ✓ Ability to apply metacognitive strategies to write linguistically and rhetorically appropriate emails ✓ Ability to apply knowledge of the writing process including planning, drafting, revising, and composing emails ✓ Ability to recognize differences in stylistic preferences of recipients (e.g., virtual co-workers, boss, mentor, and clients)
Students	<ul style="list-style-type: none"> ✓ Students enrolled in postsecondary professional and technical communication programs ✓ Students in high school before entering the workplace in technical careers
Additional skills	<ul style="list-style-type: none"> ✓ Facility with word-processing interfaces ✓ Awareness of the cultural norms and customs of U.S. workplaces ✓ Proficiency in English with respect to typing/keyboarding skill or facility with grammar, usage, mechanics
Potential Work Products	<ul style="list-style-type: none"> ✓ Emails written for different audiences and rhetorical purposes ✓ Answers to reflection questions about rationales for writing choices ✓ Responses from selected-response questions about intentions or email characteristics
Potential Observations	<ul style="list-style-type: none"> ✓ Linguistic markers that capture linguistic and rhetorical features of emails ✓ Responses to selected-response questions that provide indications of intentions around content or style ✓ Responses to selected-response questions that evaluate appropriateness of linguistic or rhetorical email features
Characteristic Features	<ul style="list-style-type: none"> ✓ Organizational structure includes a supervisor, consultant, and multiple team members who have different roles and responsibilities ✓ Scenario focuses on creating a bid for the design of a kitchen, negotiating with a client, and creating a final design ✓ Emails are the main mode of simulated communication
Variable Features	<ul style="list-style-type: none"> ✓ Composition of team in terms of gender, ethnicity, work experience ✓ Preferences of team members in terms of communication styles ✓ Organizational structure in terms of power and distance relationships ✓ Mixture of rhetorical goals across emails

4.0 The Integrated Design and Appraisal Framework

Beyond intricacies arising in the development of assessments of complex constructs due to the need for collaboration across cross-functional teams, other intricacies must be paid attention to given the diversity of the populations taking assessments. A focus on assessment design is important not only because of methodological and use-related implications but also because design choices have implications for the consequences of testing. The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) have increased the expectations for the design and appraisal of assessment programs. Historically absent in many validation studies has been attention to the consequences of test use (Lane, 2014; Zumbo & Chan, 2014). The *Standards*, however, make clear that concern for fairness must be addressed in the design, implementation, appraisal, and use of large-scale assessments. Borsboom et al. (2004) have argued that traditional validity evidence is incommensurate with evidence related to social consequences. This incompatibility between classes of evidence, they claim, is responsible for the paucity of research on the consequences of testing. This reduced attention to test-use consequences raises concerns for the field's turn toward fairness (Poe et al., 2018). The IDAF model was developed to promote fairness in writing assessment design (Slomp, 2016). The IDAF, through its generalizable questions and attention to consequences, is well suited to inform the design of prototypes and assessments like the WEC module while simultaneously supporting our inquiry on fairness in the development of digital assessments.

With this in mind, we focus on the IDAF model to address the following questions: What are the benefits of using the IDAF? What are its key steps? How can those steps be applied to the design of WEC modules?

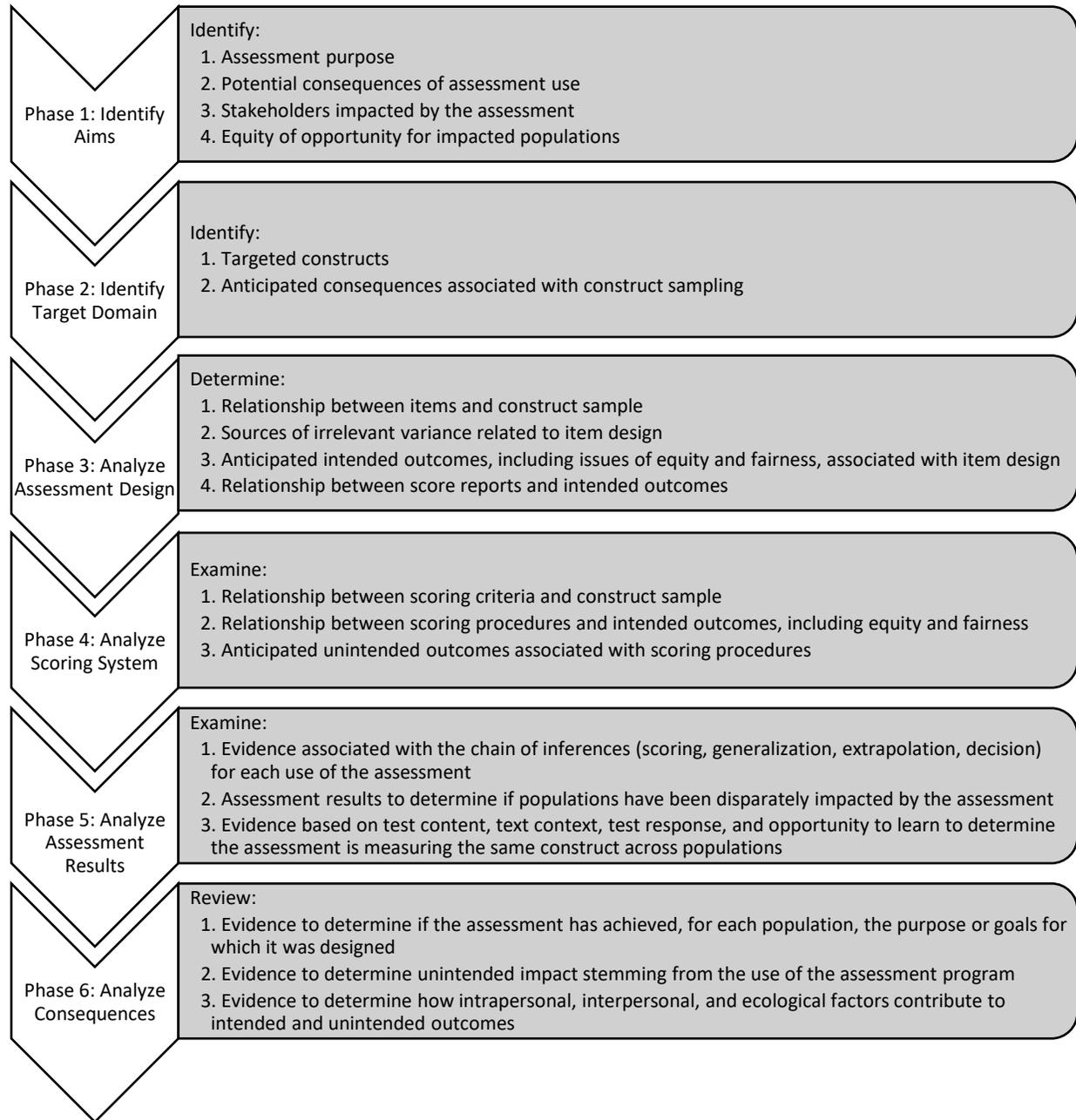
The use of the IDAF allows us to

- engage in an interdisciplinary development and evaluation approach,
- support meaningful and actionable feedback to support both learning and assessment,
- empower students to see themselves in educational careers and situations that may be unfamiliar to them,
- provide feedback and evaluation information that is acceptable to the diverse communities of students that we seek to engage with, and
- utilize digital technology not as the sole driver of instructional value but, rather, one key component of instruction in a blended classroom approach.

As shown in Figure 3, the IDAF model proceeds in six iterative and recursive phases.

Figure 3

Overview of Integrated Design and Appraisal Framework



The six phases are as follows:

1. **Identify aims.** Key issues related to the assessment program's design and use are identified. These include the information needs motivating the assessment; the audiences for that information; the inferences, decisions, and actions to be taken based on that information; the populations to be affected by the assessment; and the intended and unintended positive and negative consequences that could result from this assessment process.
2. **Identify target domain.** Analysis of Phase 1 information is used to identify elements foundational to the assessment's design. Particular attention is paid to understanding the constructs and content domains at the core of the assessment program.
3. **Analyze assessment design.** The assessment instruments are critically selected and examined with respect to how well they capture both the constructs of interest and the content domains identified in Phase 2. Particular attention is paid to issues of construct underrepresentation and construct-irrelevant variance.
4. **Analyze scoring systems.** The scoring system is critically examined. Scoring procedures are screened for how they influence construct representation and for how they impact populations of test takers.
5. **Analyze assessment results.** Inferences based on assessment results (e.g., based on pilot testing) are critically examined. The emphasis is on analyzing the strength of the scoring, generalization, and extrapolation inferences while also attending to issues of disparate impact with respect to populations of interest identified in Phase 1.
6. **Analyze consequences.** Assessment consequences are analyzed with a focus on understanding intended, unintended, positive, and negative outcomes of the assessment's implementation and use.

The IDAF model shares common features that are extensions or extrapolations from Kane's (2006, 2013) Interpretive Use Argument model of validation. The IDAF model advances a program of research throughout the design or validation processes that systematically integrates attention to social consequences—including fairness and impact on educational systems and the multiple stakeholder groups associated with those systems. Few field studies employing these integrated models to the assessment of literacy exist in the literature. The next section describes how the key steps of the IDAF can be applied to the design of WEC modules.

5. 0 Application of IDAF to WEC Modules

In what follows, we describe the IDAF model and how we used it to inform the design and development of the WEC prototype. This description follows the phases noted in Figure 3.

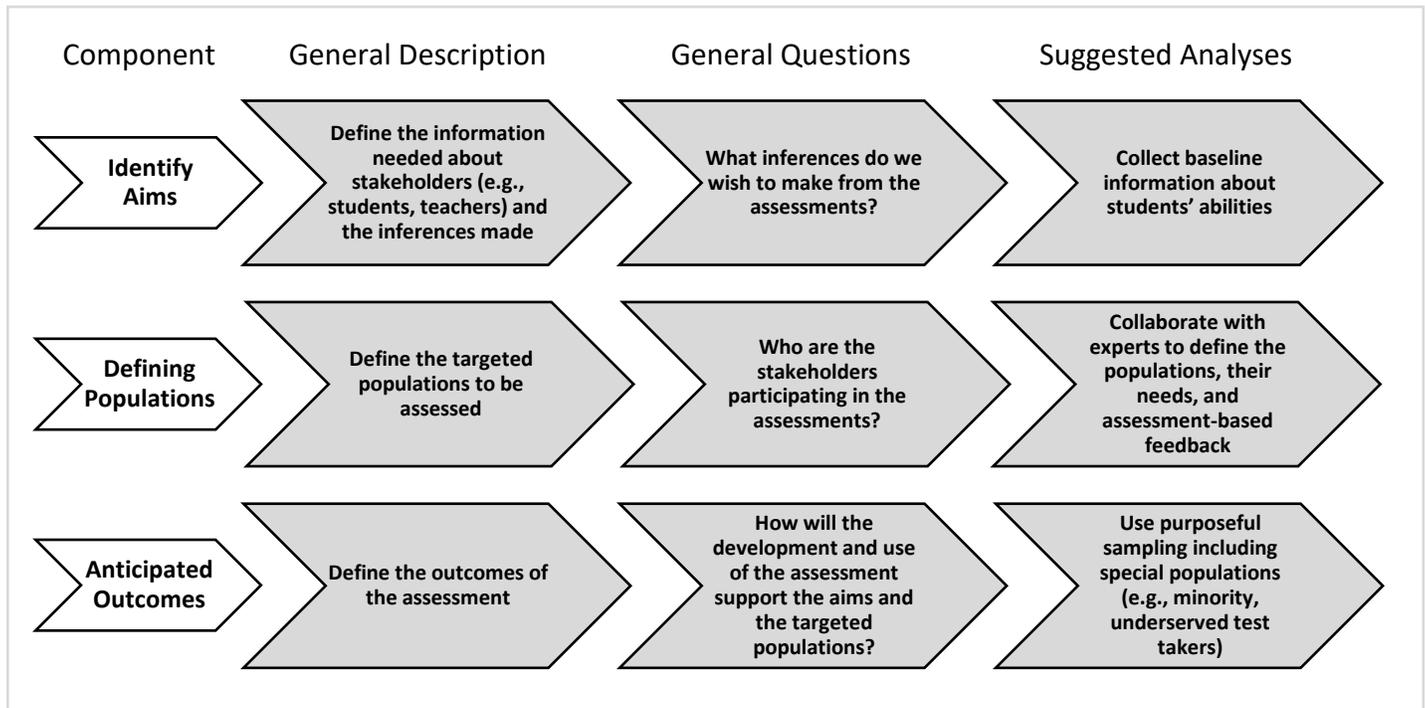
5.1 Phase 1: Identify Aims

The first phase is to define the aims, populations, and anticipated outcomes for the assessment. Figure 4 shows the key elements of this phase, the general questions, and the suggested analyses.

With respect to the WEC prototype, our design is motivated by employers and policy makers concerned about student preparation to independently take on complex communications tasks required of them when they enter the workforce. To address this problem, the WEC prototype is being designed as a learning platform. Therefore, the primary information it is designed to collect and report focuses on formative assessment data to guide teaching and learning.

Figure 4

Phase 1: Identify Aims



The WEC modules were designed to help address the following high-level questions: (a) Have students developed independent capacity to analyze, respond to, and successfully complete email communication tasks? (b) How effectively can students problem-solve and successfully complete novel workplace communications tasks? As mentioned, in this article, we focus on email writing; thus, in the modules, we included assessment tasks and embedded feedback to students that would help them evaluate their own email compositions based on traditionally used linguistic, cognitive, and substantive patterns often used in the workplace (see Oliveri, Mislevy, & Slomp, 2021, this issue, for an elaboration).

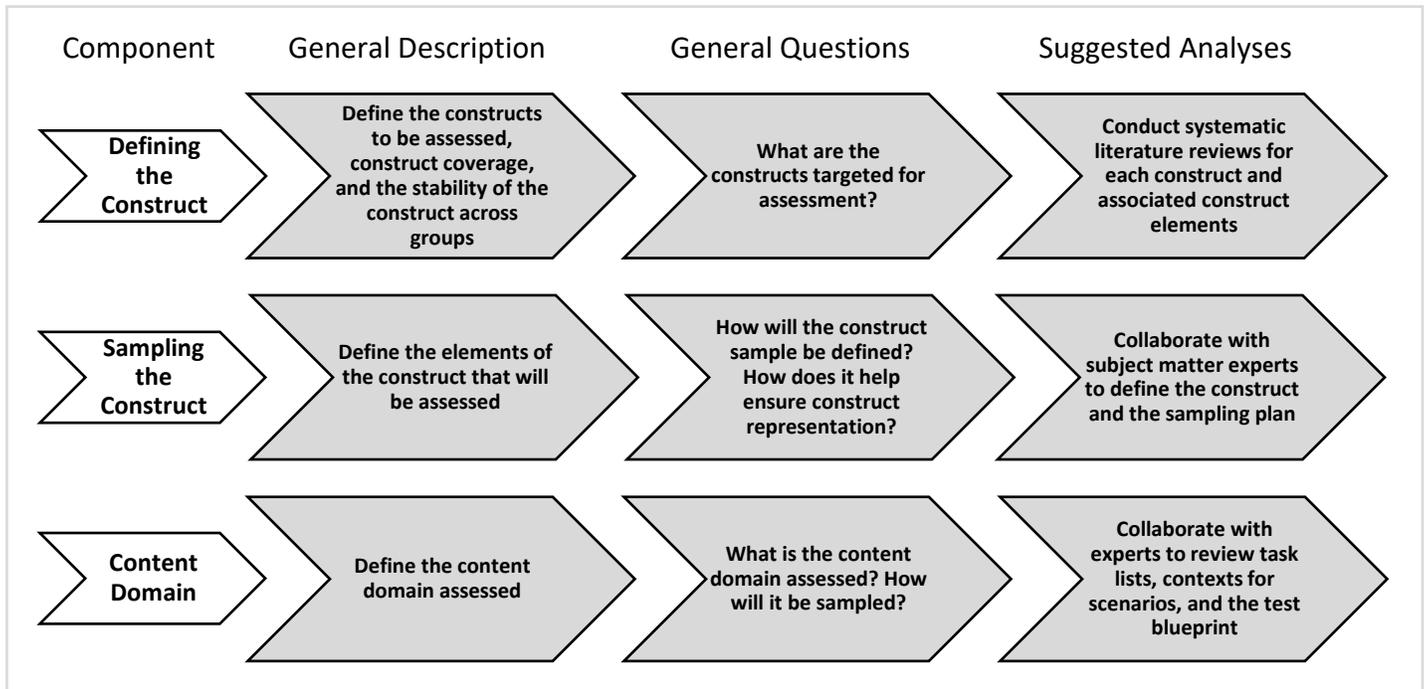
As noted in Table 2, the modules were designed to provide students enrolled in high school or postsecondary professional and technical communication programs with opportunities to learn

WEC with a focus on traditional email writing patterns. They were also designed to provide instructors with feedback on students' writing patterns as elaborated in Zapata et al. (2021, this issue). As described in Haigler (2021, this issue), they were reviewed by subject matter experts that had experience working with students in high schools and community colleges.

The anticipated outcomes aspect of Phase 1 sets up a combined focus throughout the design process on issues of fairness and anticipated outcomes. The broadest aspect of fairness is the concern for opportunity to learn (OtL). OtL addresses not only the instructional milieu in which learning is expected to take place but also the ecological context in which that student is situated (Gee, 2008). Assessment programs designed to optimize OtL need to ensure students have access to strong, robust programs of instruction. Thus, a collaboration with subject matter experts with familiarity working with marginalized learners was sought to identify issues that may lead to construct-irrelevant variance for marginalized learners, such as language that was unnecessarily complex and the use of images or avatars that represented a diverse population of learners. Moreover, content that was deemed to lead to communication issues such as politeness when interacting with clients was included as part of the outcomes assessed by the modules, as we elaborate in Phase 2.

5.2 Phase 2: Identify Target Domain

The second phase is to define the construct, sample it, and define the content domain. Figure 5 shows the key elements of this phase, the general questions, and the suggested analysis. In relation to the WEC modules, to define the construct, we asked the following: What construct(s) are targeted in the assessment in ways that sample the construct associated with workplace writing and provide learners with opportunities to learn key construct-relevant elements such as essential aspects of email writing?

Figure 5
Phase 2: Identify Target Domain


The goal of Phase 2 is to ensure from the outset of the design of the WEC modules that there is a close affinity between the modules being designed and the workplace communications construct assessed. This alignment is important to support the WEC prototype goal of promoting student ability with respect to workplace communications. To achieve this goal, we included opportunities for students to work with different communication genres, rhetorical aims, and discourse contexts. Earlier sections in this article provided a general overview of the construct (see also Corrigan & Slomp, 2021, this issue, for a more focused discussion).

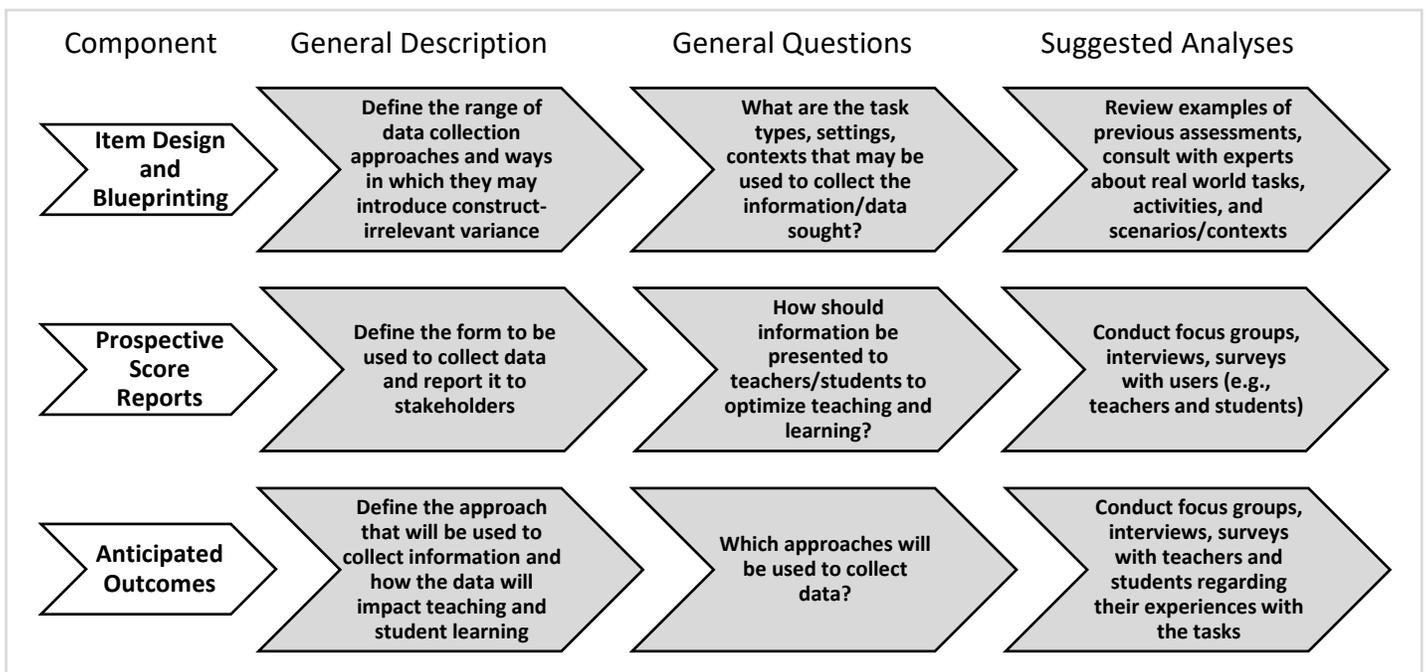
In this section, we discuss key construct elements represented in our WEC module, which provides a narrower construct representation than the one we articulated earlier. Our modules contain micro-scenarios that ask the student to complete a range of tasks that vary in terms of rhetorical goals and audience, along with a variety of metacognitive reflection questions. For instance, we included activities for students to reflect on key aspects of email writing. We also included opportunities to teach and assess metacognitive abilities to help learners interact with email subgenres and support their development of genre knowledge (Driscoll et al., 2020). To develop substantive knowledge, the WEC modules have been designed to minimize the participants' need to develop specialized content knowledge, having them focus instead on the manipulation of that content to achieve rhetorical aims.

5.3 Phase 3: Analyze Assessment Design

Consideration of Phase 3 involves designing the items, the blueprint, the prospective score reports, and the anticipated outcomes. Figure 6 shows the key elements of this phase, the general questions, and the suggested analyses. In relation to the WEC modules, relevant questions include the following: What are the task types that may be used to collect the information/data sought? How should information be presented to teachers/students to optimize teaching and learning? Which approaches will be used to collect data? These questions could be investigated by conducting reviews of sample tasks, or by conducting focus groups, interviews, and surveys with users (e.g., teachers and students) about the utility and relevance of the information presented in the prospective score reports as well as their experiences with the tasks (see Zapata-Rivera et al., 2021, this issue, for examples of these reports).

Figure 6

Phase 3: Analyze Assessment Design



Assessment design is an iterative process that must be guided by constant reflection on the (a) aims for which the assessment program is designed, (b) information needs it is designed to fulfil, (c) construct and content domains it needs to reflect, and (d) populations who will be impacted by the assessment.

The first set of questions in Phase 3 focuses on identifying the range of techniques for collecting student performance data. These questions are followed by an exploration of how each of these options introduce potential construct-irrelevant variance into the assessment data.

In the initial stage of design, it was decided that the WEC prototype would be a scenario-based assessment. The scenario-based approach enables the design team to approximate the types of tasks and contexts that the O*Net review demonstrated occurred more frequently, and that had significant impacts on workplace efficiency or were determined to be relevant and important to the workplace (Oliveri & McCulla, 2019). The initial phase of design focused on alignment with the content domain identified in Phase 2.

The initial set of WEC tasks was subjected to reviews by experts. Two focus groups were conducted. One focus group included experts that represented employers. A second focus group included academicians.

The participants evaluated the tasks for construct alignment and construct sampling. That review suggested revisions to the WEC design that would enable greater resonance with the construct identified in Phase 2. This included (a) accompanying the communications tasks with a sequence of metacognitive reflection tasks keyed to each construct element, (b) restructuring of tasks to position students to take a problem-solving stance toward their learning, (c) including a blend of direct instruction and experiential learning opportunities into the scenario, (d) utilizing selected and constructed-response items to provide formative assessment data, and (e) taking a more sophisticated approach to assessing discourse community knowledge and genre knowledge.

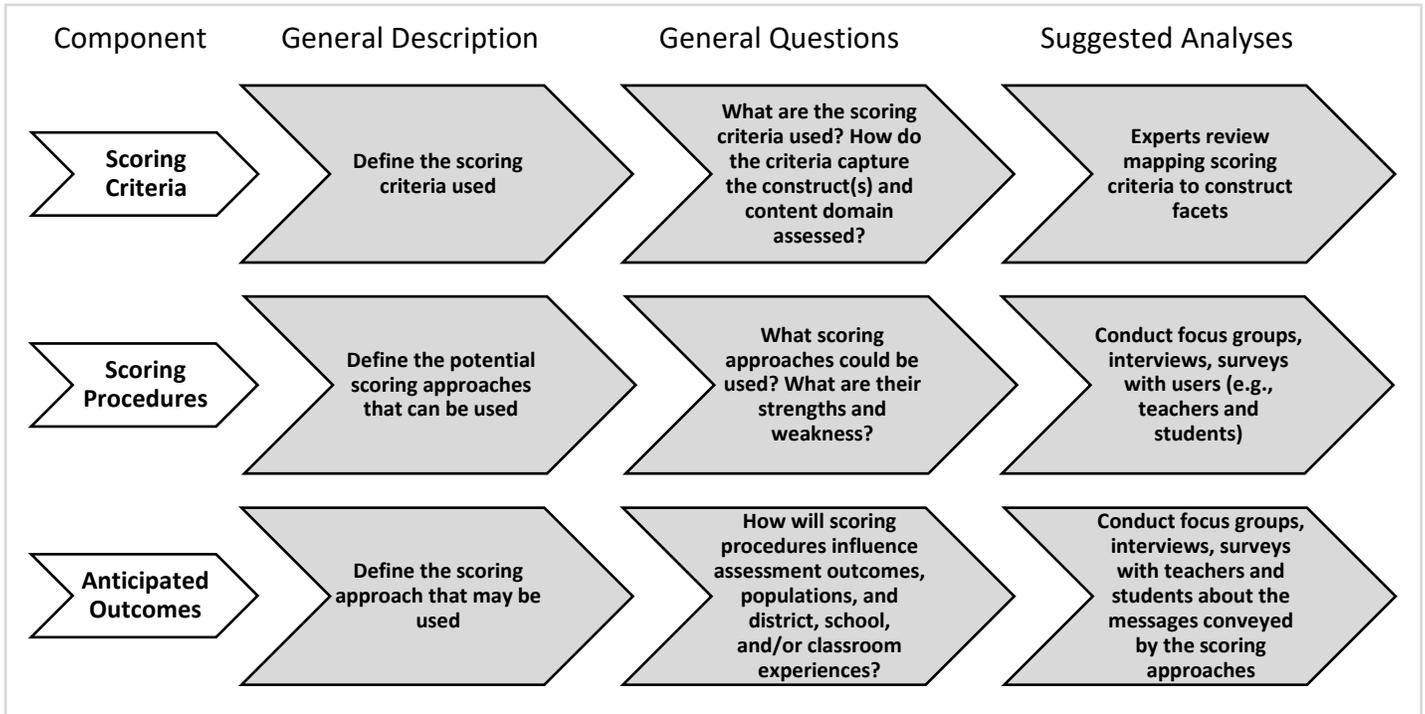
In response to this feedback, the second iteration of the WEC prototype dismantled the singular scenario into a series of episodic modules targeted toward different construct elements. These redesigned modules incorporate a blend of instructional components and formative assessment tasks.

5.4 Phase 4: Analyze Scoring System

Consideration of Phase 4 involves designing the scoring criteria, scoring procedures, and the anticipated outcomes. Figure 7 shows the key elements of this phase, the general questions, and the suggested analysis. In relation to the WEC module, relevant questions include the following: What are the scoring criteria used? How do the criteria capture the construct(s) and content domain assessed? What scoring approaches could be used? What are their strengths and weaknesses? How will scoring procedures influence assessment outcomes; populations; and district, school, and/or classroom experiences? These questions could be investigated by conducting focus groups, interviews, or surveys with users (e.g., teachers and students) about the scoring criteria, scoring procedures, and the messages potentially conveyed by the selected scoring mechanisms. A sample of 250 responses to the WEC tasks were collected through Amazon Mechanical Turk, and rubrics were developed to score these sample responses by construct domains and test-taker variables. (See Oliveri, Mislevy, & Slomp, 2021, this issue for an illustration of Q-matrices that demonstrate the relationships between items and construct domains.)

Figure 7

Phase 4: Analyze Scoring Systems



We briefly discuss considerations made in selecting our scoring choices. First, we note that there are differences across populations and groups with regard to the use of language and its patterns across different workplace settings, industries, and business settings. When designing our modules, the automated feedback, and the scoring criteria, we purposefully chose to not provide information about whether a response was correct or incorrect. Instead, we created opportunities for learners to think through what the appropriate answers were for the context, communicative situation, and discourse community presented to the students in the modules. These choices have implications for how feedback is phrased and for how pieces of evidence are identified and aggregated more generally (e.g., Mislevy, 2018; Mislevy & Oliveri, 2019).

Importantly, we wanted to move away from the idea of “normative scoring” as the starting point for our evaluation towards the idea of mutually informative feedback. The notion of scoring, whether it is on a dichotomous (e.g., 0-1) or polytomous (e.g., 0, 1, or 2) scale, implies that there is a normative “ground truth” that allows a human rater or an automated system to make an ordered judgment about a response. This implies that the underlying machinery is robust and essentially unassailable, which is only defensible in simplistic, extreme cases (e.g., scoring inflammatory language as inappropriate or scoring one-word responses to a complex email as insufficient).

In our context, however, as we expect differences in how different populations might respond, we used a “learning mode” and provide no correct or incorrect feedback, but instead

provide suggestions for the type of statement that could have been used in a context as the one presented in the modules. More importantly, however, we wanted to use our modules as data-collection devices that helped us understand how students from different backgrounds and with different experiences communicated in a workplace environment, such as the type of conventions and language they used when composing emails and how they justified their choices. Consequently, as the IDAF framework encourages, we did not use value-laden scoring rules during early tryouts and instead used these tryouts to have students share their perspectives, experiences, and rationales with the development team through reflection questions. (See Oliveri, Mislevy, & Slomp, 2021, this issue for an elaboration of the types of linguistic, cognitive, and substantive patterns that may arise when administering an assessment to socioculturally diverse populations in relation to framing and conveying responses in socioculturally diverse ways depending on demographic variables and/or the type of organization one may work in.)

5.5 Phase 5: Analyze Assessment Results

Consideration of Phase 5 involves analyzing the Interpretation and Use Argument. Figure 8 shows the key elements of this phase, the general questions, and the suggested analysis. Questions in this phase include the following: Within the population being assessed, have the populations who might be disparately impacted by the assessment been identified? Can differences in performance be attributed to actual differences in ability in relation to the construct being measured?

At this phase, an evaluation of the assumptions related to scoring, generalization, extrapolation, and decision inferences is made (Kane, 2013). Specifically, for each of these types of inferences, the following questions are asked:

Scoring Inference: Underlying the scoring inference is the assumption that each student's scores accurately reflect their performance on the workplace communications construct being measured. With this assumption in mind, we ask the following questions: Are differences in scores primarily due to differences in the constructs assessed or are they due to potential sources of construct-irrelevant differences? The scoring inference could be investigated by examining variability in scores across students as well as the potential sources of the score-based differences.

Generalization Inference: Underlying the generalization inference is the assumption that students' performance on the modules and their interactions with the modules represent what a student would be expected to obtain over an extended set of tasks completed across assessment sessions. With this assumption in mind, we ask whether the results across occasions yield similar results (evidence of performance). The generalization inference could be investigated by examining students' performance across tasks or different aspects of email composition.

Extrapolation Inference: Underlying the extrapolation inference is the assumption that the results of students' responses and interactions with the modules represent how a student would be expected to perform in non-testing (i.e., real-life) contexts. With this assumption in mind, we

ask the following: Do the students' performances in simulated (scenario-based) situations predict performance in non-testing situations? The extrapolation inference could be examined by analyzing students' performance occurring in settings outside of the classrooms, such as when they enter their first year of employment.

Decision Inference: Underlying the decision inference assumption is whether the modules contribute to providing information or informing decisions about student learning, which is the main goal for using the modules. With this assumption in mind, we ask whether those decisions are justifiable, equitable, and based on sufficient and useful evidence.

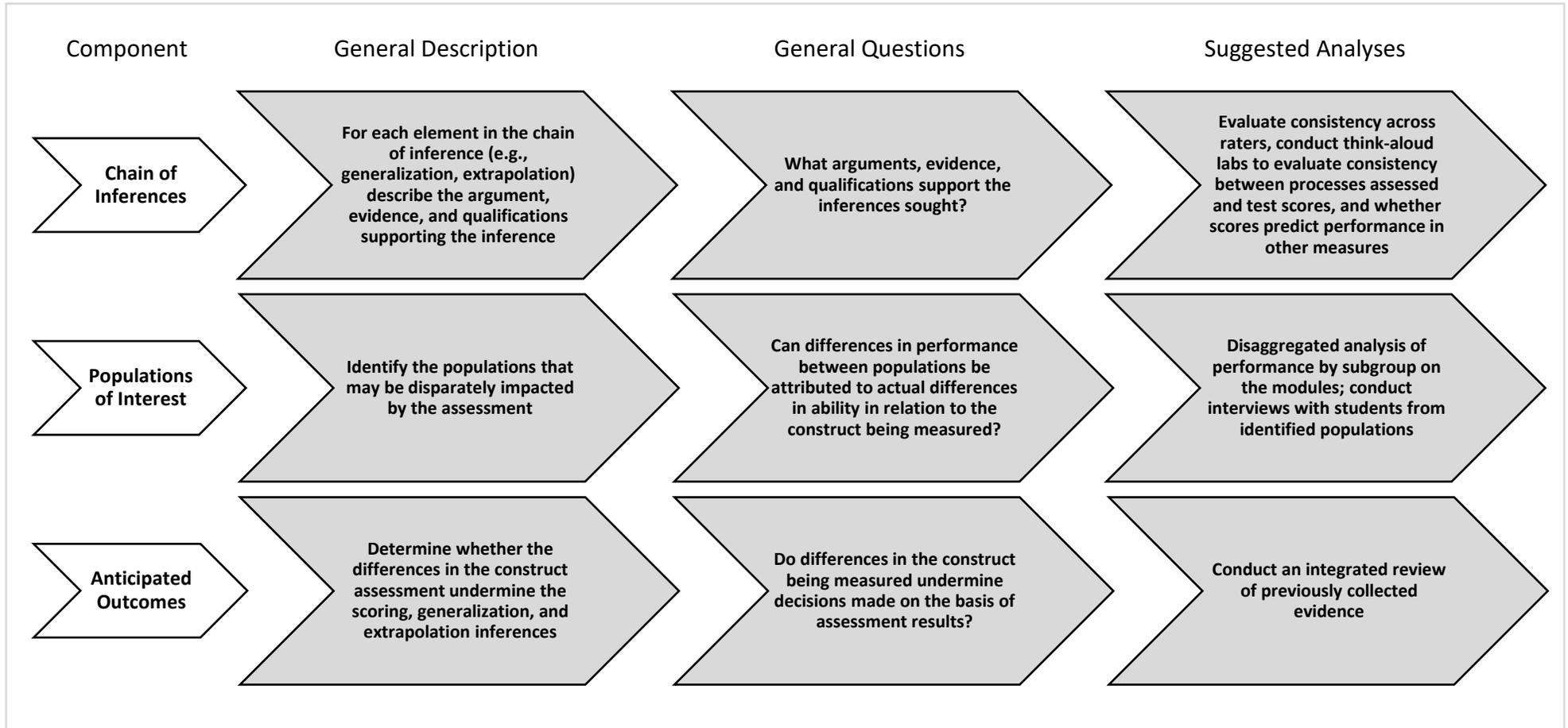
Evaluating these inferences is important to

- identify potential students who might be disparately impacted by the assessment,
- determine whether differences in performance between students may be attributed to actual differences in ability in relation to the constructs measured,
- identify cases where differences in students' performances have been identified, and
- determine whether the evidence based on content, response processes, and internal structure indicates that the modules are measuring the same construct across the students.

These questions are relevant in analyzing the Interpretation and Use Argument relevant to the uses of the modules in current and future-looking stages. It may involve conducting think-aloud data analysis from different groups of students to examine the alignment between cognitive processes and scores. Additional studies may involve comparing scores from the modules against scores on other similar measures or on variations of the same modules administered across time. Moreover, other studies may involve examining evidence that assessment scores predict performance in non-testing situations (e.g., in real-life settings).

Figure 8

Phase 5: Analyze Assessment Results



5.6 Phase 6: Analyze Consequences

Consideration of Phase 6 involves evaluating intended and unintended outcomes. Figure 9 shows the key elements of this phase, the general questions, and the suggested analysis. Questions in this phase include the following:

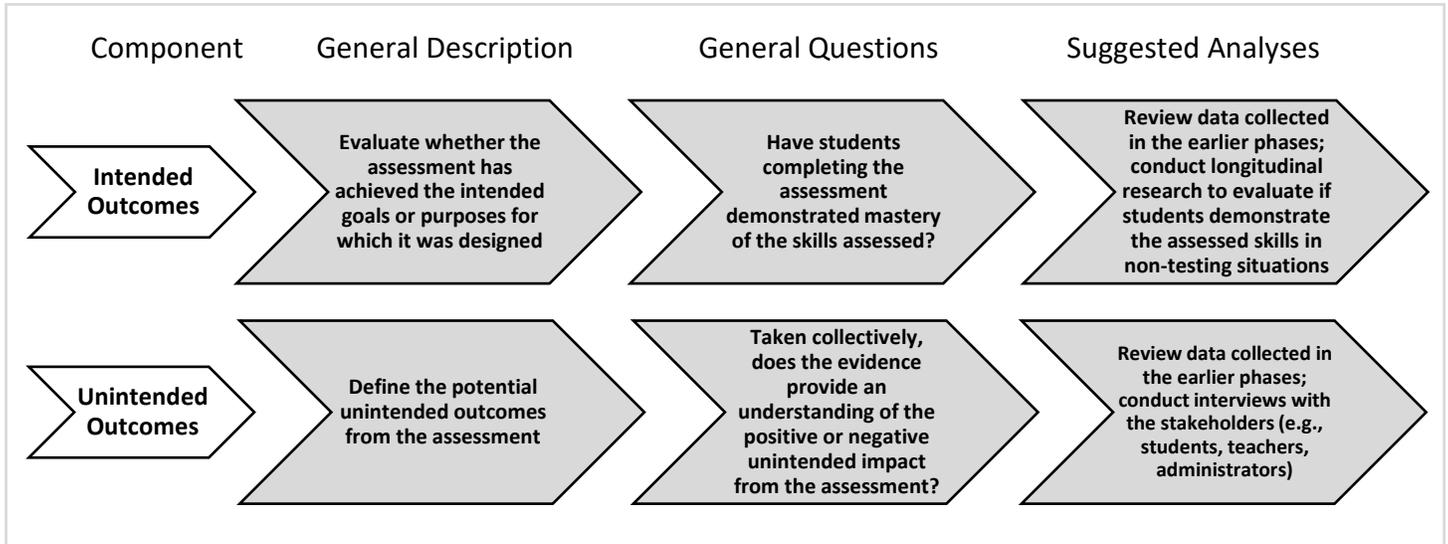
- **Intended Outcomes:** Taken collectively, does the evidence gathered indicate that the assessment has achieved the purpose or goals for which it was designed?
- **Unintended Outcomes:** Taken collectively, does the evidence provide an understanding of unintended impact, whether positive, negative, or unknown? For each population affected by the modules? For the classroom serving those populations?

At this phase, an evaluation of the intended and unintended outcomes is conducted. With respect to the WEC modules, we aim to ask the following (in future investigations):

- **Intended Outcomes:** Have students who successfully completed the modules been able to demonstrate increased learning of the cognitive (metacognitive), linguistic, or substantive knowledge and skills needed to independently analyze and respond to an array of situated communications tasks they will face in their future workplace?
- **Unintended Outcomes:** Have the use of the modules resulted in unintended learning or has their use resulted in intended positive consequences on instruction and learning of the targeted skills?

Figure 9

Phase 6: Analyze Consequences



It is still too early in the design of the WEC prototype to address these questions directly. However, as this special issue of *The Journal of Writing Analytics* demonstrates, placing questions such as whether the assessment as designed leads to the desired (intended) outcomes and whether the developer has identified and mitigated unintended outcomes in the foreground from the outset of the design process has potential benefits. For instance, these questions may enable the design team to focus on ensuring the utility and value of this assessment for students and teachers in every design conversation in which we engaged. Thus, this explicit focus on unintended and intended outcomes sharpens decision-making about item design, sampling plans, scoring systems, and score reporting, and narrows the focus from what is possible to what is necessary.

The procedures described in this article represent only a small snapshot of the overall process of the design, implementation, and evaluation cycles for our project. The framework we described helped us inform the design decisions made throughout the design and development phases. We acknowledge that the choice of any particular framework or set of standards or guidelines does not result in a singular recommendation for how innovation work can proceed. Instead, interdisciplinary teams have to translate the fundamental values, principles, and recommendations into work processes, artifacts, and mindsets that support a continually critical self-evaluation (see Rupp et al., 2020). In other words, work is always work-in-progress.

6.0 Conclusion

Rapid developments in technology, diversity of student populations, needs for higher-level skills, cognitive and social research on learning, and demands for better integration of assessment with learning all present educators with both challenges and opportunities. Assessment in particular sorely needs to adapt, not simply in more effective ways of carrying out traditional roles in familiar niches, serving familiar roles in existing social structures, for these too are changing. Any assessment is a nexus of multiway communication among individuals, within and across institutions, that shapes their lives and activities. As institutions, lives, and activities change, so must assessments. As illustrated in the WEC example discussed across this article and the companion articles in this special issue, taking advantage of our increased knowledge and tools to address our challenges requires insights across the domains involved, taking new forms and methods in new environments. Traditional, self-contained instructional design and assessment development practices are not adequate.

Recent developments in learning sciences and assessment offer support to meet these challenges. Principled frameworks for assessment design explicitly bring considerations of construct definition, principled use of technology, and connection with social and cognitive underpinnings into an integrated design process; issues of validity, reliability, and fairness are likewise integrated into the process, all in the service of meeting an assessment's intended purposes. There is no prescriptive approach to designing assessments; there are instead frameworks that organize thinking, communication, and collaboration across the specialists whose work needs to come together to produce effectual artifacts, within effectual systems. We collaborated in this project to not only integrate insights from our disciplinary team, but to illustrate using one exemplary prototype, how the ideas play out in practice, and not only with the artifacts and processes we created, but to make explicit the frameworks and methods by which we produced them.

Turning to lessons learned from the use of the IDAF and the challenges we anticipate facing as we move forward with the design of modules of complex constructs, one of the key lessons we learned was to identify the types of representations needed by each of the members of the interdisciplinary team we worked with and which types of activities each member was able to contribute. That is, we identified team members' roles and expertise to carry out this complex set of activities. One strategy that helped us work together was the use of common artifacts and representations. Examples presented in this article include the use of design patterns (see Table 3) and the use of a common framework (e.g., see the IDAF model, Figures 3 to 9).

As much as it may be tempting to believe that the collection of seemingly "objective" data-based artifacts such as writing samples and response patterns is the central goal of the design process, it is important to highlight that the design process also involved a series of iterations that were informed by data but were also informed by the values of the team that engaged in the process. These values were, of course, informed by supporting theories from the literature but, nevertheless, designing instrumentation required making a myriad of design decisions that

affected all subsequent narratives that one crafted around the instrumentation. For example, focusing on the WEC construct, embracing a metacognitive model of genre and discourse, choosing a digital delivery mode, contextualizing the tasks in a design company scenario, and opting to provide guiding feedback rather than merely evaluative scoring were fundamental choices that could be questioned. Making different critical choices could lead one down the path of developing a fundamentally different ecosystem of instruments and resources that would align better with alternate educational models, if that vision were the motivation. In other words, the design of an assessment—especially a complex one—is not a self-contained enterprise, but one shaped by one’s view of the system, values, and educational perspective of the environment in which it will function.

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

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Arieli-Attali, M., Ward, S., Thomas, J., Deonovic, B., von Davier, A. A. (2019). The expanded evidence-centered design (e-ECD) framework for learning and assessment systems: A framework for incorporating learning goals and processes within assessment design. *Frontiers in Psychology, 10*. <https://doi.org/10.3389/fpsyg.2019.00853>
- 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>
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*, 1061-1071.
- Coppola, N. W., Elliot, N., & Newsham, F. (2016). Programmatic research in technical communication: An interpretive framework for writing program assessment. *Programmatic Perspectives, 8*, 5-45.
- Corrigan, J. A., & Slomp, D. H. (2021). Articulating a sociocognitive construct of writing expertise for the digital age. *The Journal of Writing Analytics, 5*.

- Cullen, M. J., Sackett, P. R., & Lievens, F. (2006). Threats to the operational use of situational judgment tests in the college admission process. *International Journal of Selection and Assessment*, 14(2), 142-155.
- 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>
- Dias, P., Freedman, A., Medway, P., & Paré, A. (2013). *Worlds apart: Acting and writing in academic and workplace contexts*. Routledge.
- Driscoll, D. L., Paszek, J., Gorzelsky, G., Hayes, C. L., & Jones, E. (2020). Genre knowledge and writing development: Results from the writing transfer project. *Written Communication*, 37(1), 69-103.
- Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods*, 3, 300-396.
- Embretson, S. E. (2020, April). *Test development with cognitive design systems: A few examples* [Paper presentation]. Annual meeting of the National Council for Measurement in Education (NCME), San Francisco, CA, United States.
- Ferrara, S., Lai, E., Reilly, A., & Nichols, P. D. (2016). Principled approaches to assessment design, development, and implementation. In A. A. Rupp & J. P. Leighton (Eds.), *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 41-74). Wiley.
- Gee, J. P. (2008). A sociocultural perspective on opportunity to learn. In P. Moss, D. C. Pullin, & J. P. Gee (Eds.), *Assessment, equity, and opportunity to learn* (pp. 76-108). Cambridge University Press.
- 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
- Haigler, K. (2021). The role of occupational research in complex assessments: Aligning educational practices with workplace requirements. *The Journal of Writing Analytics*, 5.
- Hollis-Turner, S., & Scholtz, D. (2010). Business writing in academic and workplace contexts. *Southern African Linguistics and Applied Language Studies*, 28(3), 239-245.
- Kane, M. T. (2006). Validation. In R.L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). American Council on Education; Praeger.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Keehner, M., Gorin, J. S., Feng, G., & Katz, I. (2016). Developing and validating cognitive models in assessment. In A. A. Rupp & J. P. Leighton (Eds.), *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 75-101). Wiley.
- 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
- Lane, S. (2014). Validity evidence based on testing consequences. *Psicothema*, 26(1), 127-135.
- Leu, D. J., Slomp, D., Zawilinski, L., & Corrigan, J. A. (2016). Writing research through a new literacies lens. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 41-56). New York.

- Lievens, F., Buyse, T., & Sackett, P. R. (2005). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology*, *58*(4), 981-1007.
<https://doi.org/10.1111/j.1744-6570.2005.00713.x>
- Liu, M., & Haertel, G. (2011). *Design patterns: A tool to support assessment task authoring* (Large-Scale Assessment Technical Report 11). SRI International.
- Luecht, R. M. (2013). Assessment engineering task model maps, task models and templates as a new way to develop and implement test specification. *Journal of Applied Testing Technology*, *14*.
<http://www.jattjournal.com/index.php/atp/article/view/45254>
- McKinsey Global Institute. (2018). *Skill shift: Automation and the future of the workforce*.
<https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce#>
- 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
- Mislevy, R. J. (2007). Validity by design. *Educational Researcher*, *36*(8), 463-469.
- Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement*. Routledge.
- Mislevy, R. J., & Haertel, G. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, *25*(4), 6-20.
<https://doi.org/10.1111/j.17453992.2006.00075.x>
- Mislevy, R. J., & Oliveri, M. E. (2019). Sociocognitive assessment for diverse populations (Digital ITEMS Module 09). *Educational Measurement: Issues and Practice*, *38*(4), 110-111.
<https://ncme.elevate.commpartners.com>
- Mislevy, R. J., Steinberg, L. S., & Almond, R. A. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, *1*, 3-67.
- National Association of Colleges and Employers (NACE). (2018). *Job outlook 2019*.
- National Governors Association Center for Best Practices. (2010). *Council of Chief State School Officers Common Core State Standards for English Language Arts*.
- Neeley, T. (2012). Global business speaks English. *Harvard Business Review*.
<https://hbr.org/2012/05/global-business-speaks-english>
- Nichols, P. D., Kobrin, J. L., Lai, E., & Koepfler, J. (2016). The role of theories of learning and cognition in assessment design and development. In A. A. Rupp & J. P. Leighton (Eds.), *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 15-40). Wiley.
- 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. (2018). *Analyzing the validity of large-scale assessments administered globally* (ETS Research Report No. 18-35). 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*. <https://doi.org/10.1080/15305058.2018.1543308>
- 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., & Young, J. W. (2015). *A validity framework for the use and development of exported assessments* (ETS Office of Professional Standards Series). Educational Testing Service.
- Oliveri, M. E., & Markle, R. (2017). *Expanding skills in higher education* (ETS Research Report No. 17-09). Educational Testing Service.
- Oliveri, M. E., & McCulla, L. (2019). *Using the Occupational Network Database to assess and improve English language communication for the workplace* (ETS Research Report No. RR-19-28). Educational Testing Service. <https://doi.org/10.1002/ets2.12265>
- 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 Analytics*, 5.
- Oliveri, M. E., Slomp, D. H., Rupp, A. A., & Mislevy, R. J. (2021). Principled development of Workplace English Communication Part 2: Expanded evidence-centered design and theory of action frameworks. *The Journal of Writing Analytics*, 5.
- Oliveri, M. E., & Tannenbaum, R. J. (2019). Are we teaching and assessing the English skills needed to succeed in the global workplace? In V. Hammler Kenon & S. Vasant Palsole (Eds.), *The Wiley handbook of global workplace learning* (pp. 343-354). Wiley.
- Partnership for 21st Century Learning (P21). (2015). *Framework for 21st century learning definitions*. http://static.battelleforkids.org/documents/p21/P21_Framework_DefinitionsBFFK.pdf
- 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
- 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>
- Poe, M., Inoue, A. B., & Elliot, N. (2018). *Writing assessment, social justice, and the advancement of opportunity*. The WAC Clearinghouse; University Press of Colorado.
- Quigley, B. (2018). *Youthful bookworms: Students' experiences with critical literacies in the context of the Ontario secondary school literacy test* [Doctoral dissertation, Université d'Ottawa/University of Ottawa]. <https://ruor.uottawa.ca/handle/10393/37573>
- Riconscente, M. M., Mislevy, R. J., & Hamel, L. (2005). *An introduction to PADI task templates*. PADI Technical Report (Report No. 3). SRI.
- Rupp, A. A., Foltz, P. W., & Yan, D. (2020). Theory into practice: Reflections on the handbook. In D. Yan, A. A. Rupp, & P. W. Foltz (Eds.), *The handbook of automated scoring: Theory into practice* (pp. 478-489). CRC Press.
- Sabatini, J. P., Bennett, R. E., & Deane, P. (2011). *Four years of cognitively based assessment of, for, and as learning (CBAL): Learning about through-course assessment (TCA)* [Paper Presentation]. Invitational Research Symposium on Through Course Summative Assessment, Atlanta, GA, United States. www.k12center.org/rsc/pdf/TCSA_Symposium_Final_Paper_Bennett_Kane_Bridgeman.pdf
- Schrifer, K. (2012). What we know about expertise in professional communication. In V. W. Berninger (Ed.), *Past, present, and future contributions of cognitive writing research to cognitive psychology* (pp. 275-312). Psychology Press.

- Sireci, S. G. (2015). A Theory of Action for test validation. In H. Jiao & R.W. Lissitz (Eds.), *The next generation of testing: Common Core Standards, Smarter-Balanced, PARCC, and the nationwide testing movement* (pp. 253-272). Information Age.
- Slomp, D. H. (2016). An integrated design and appraisal framework for ethical writing assessment. *The Journal of Writing Assessment*, 9(1). <http://journalofwritingassessment.org/article.php?article=94>
- Slomp, D. H., Corrigan, J. A., & Sugimoto, T. (2014). A framework for using consequential validity evidence in evaluating large-scale writing assessments: A Canadian study. *Research in the Teaching of English*, 48(3), 276-302.
- White, E. M., Elliot, N., & Peckham, I. (2015). *Very like a whale: The assessment of writing programs*. University Press of Colorado.
- Wilson, M. (2013). Using the concept of a measurement system to characterize measurement models used in psychometrics. *Measurement*, 46, 3766-3774.
- Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education*, 13(2), 181-120.
- 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.
- Zumbo, B. D., & Chan, E. K. H. (Eds.). (2014). *Validity and validation in social, behavioral, and health sciences*. Springer.