J of W Analytics

Research Article

Principled Development of Workplace English Communication Part 1: A Sociocognitive Framework

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

Robert J. Mislevy, *Educational Testing Service*

David Slomp, *University of Lethbridge*

Structured Abstract

- **Background**: This study advances a sociocognitive approach to modeling complex communication tasks. Using an integrative perspective of linguistic, cultural, and substantive (LCS) patterns, we provide a framework for understanding the nature and acquisition of people's adaptive capabilities in social/cognitive complex adaptive systems. We also illustrate the application of the framework to learning and assessment. As we will show, understanding the connection between measurement models and users' needs is important to increase assessments' educative usefulness.
- **Questions Addressed**: Our framework is designed to address questions regarding the following four areas: the nature of sociocognitive perspectives in educational measurement, the application of LCS patterns to complex communication tasks captured in an extended formative assessment of Workplace English Communication (WEC), the usefulness of psychometric models for instruction and assessment with such complex tasks, and considerations for measurement modeling.

- **Conclusions**: Our study concludes with reflections on the challenges of complex assessments such as WEC, the advantages of sociocognitive modeling for new assessment genres, and the roles of situated measurement models in meeting the challenges.

Keywords: anticipatory design frameworks; cognitive diagnosis models (CDM); communication tasks; deterministic inputs, noisy "and" gate (DINA) model; linguistic, cultural, and substantive (LCS) patterns; Q-matrices; sociocognitive models; sociocultural perspectives; Workplace English Communication (WEC); writing analytics

1.0 Background

Concurrent with changes in education characterized by increased uses of online learning, shifts in the nature of work, and the knowledge, skills, and attitudes necessary to carry out workplace activities, our assessment practices and theories also are shifting. In terms of the broad field of educational measurement and research specializations of writing analytics, such shifts include using unobtrusive evaluation while learners work through performance tasks, using formative assessments with feedback to students and teachers to aid learning, and assessing learners in ways that depart from more traditional (summative) forms of assessment. The Gordon Commission on the Future of Assessment in Education, whose charge was to generate recommendations that would improve pedagogical practice, educational measurement, and student achievement, described the challenges that lie ahead and speculated on developments that would be needed to meet them (Gordon & Heincke, 2013).

To deepen our conversation on complex assessments, writing analytics, and assessment, we begin by providing a background in which we discuss changes in assessment in relation to what, who, how, and when we assess. We also provide a brief overview of the Workplace English Communication (WEC) example. We then describe a sociocognitive perspective and discuss implications for integrated teaching and digitally distributed online assessments. We introduce the term linguistic, cultural, and substantive (LCS) patterns and look ahead to their pivotal role in learning and assessment. Turning to writing analytics, we demonstrate an application of LCS patterns to a WEC task and show how DocuScope, an analytic platform designed to statistically identify and visually display language patterns in student texts, can be used to build both learner awareness of rhetorical, lexical, and linguistic constructions for formative purposes and automated evaluation in the same conceptual framework for summative purposes.

To demonstrate the integrative force of cognition, assessment design, and analysis, we discuss designing psychometric models for assessment and instruction, with special attention to a representation form called a Q-matrix from cognitive diagnostic modeling (CDM) that can be used to connect measurement models, tasks, and students' performances. The relevance of psychometric approaches to 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—grounds this discussion. We continue with a more general discussion of considerations for measurement modeling and describe how modeling choices can be built from the foundation laid in the Q-matrix (or extensions thereof). We also address implications of such modeling for instruction and assessment. In our conclusion, we reflect on the usefulness of sociocognitive modeling, from its conceptual value to its technical implications. We consider future directions for this work. We pay particular attention to remaining challenges of complex assessments, advantages of sociocognitive modeling for new assessment genres, and the nature of the situated use of measurement models.

1.1. Overview of Changes in Assessment

This section discusses changes in assessment that are occurring in relation to what, who, how, and when we assess. We elaborate on what we mean with respect to each of these elements and the implications they hold for the design and development of assessments and their relation to learning in a domain.

1.1.1 Changes in the Constructs Assessed (What We Assess)

With respect to what we assess, Oliveri et al. (2020) discuss the importance of expanding the skillset assessed to better prepare learners for an evolving workplace. Considering forms of assessment beyond those used for selection, the authors suggest expanding the skills assessed with a focus on assessing more complex constructs, such as workplace communication and collaboration.

We aim to assess 21st century skills such as communication through integrated learning and assessment tools to provide learners with more immediate and useful feedback. WEC is one of the top five most important skill clusters needed for work, yet students are often under-prepared 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 lack competency in communication (National Association of Colleges and Employers [NACE], 2018).

1.1.2 Changes in Populations Assessed (Who We Assess)

Changes are also occurring in relation to the populations we are assessing in terms of learner preparation for the workplace, the types of workplaces learners will enter, and the types of teams they will work with in diverse employment settings. Such changes have implications for how to assess an increasingly diverse population to reduce potential sources of construct-irrelevant variance. Addressing this goal demands a view that encompasses concepts and research along multiple fronts, including cognitive and situative psychology (Mislevy, 2018; Weir, 2005). Locally, these insights have implicitly contributed to innovations in assessment research from

cognitive, sociocultural, and sociocognitive perspectives (Mislevy & Elliot, 2020; Oliveri et al., 2020).

In this section, we focus more specifically on a perspective we are labeling *sociocognitive*. A sociocognitive perspective is an integration across fields ordinarily described as cognitive and sociocultural, combining research results and insights into the nature and acquisition of people's capabilities in complex adaptive systems (see Section 2).[1] We focus on a sociocognitive perspective as it heightens our awareness to the abilities that are assessed, how they are learned, and how they are assessed. We elaborate on this perspective in later sections of this article and advocate for the use of a more explicit design framework for its consideration, with a joint emphasis on theoretical grounding and practical usefulness.

1.1.3 Changes in Frequency (When We Assess) and Mode (How We Assess)

Changes are also occurring with regard to how we assess and how often. We recognize the tension between (a) local, situated assessments that occur in classrooms and include discussions of individual students' performances and (b) assessments that are meant to provide broader measures of capabilities, such as formal assessments that are used to generalize over time, place, or specific performances. The former may occur more frequently in classrooms. The latter may occur less often, in standardized formats, for instance. We argue that identifying key linguistic, cultural, and substance (LCS) patterns, which encompass activities as well as knowledge in a learning area, provides a solid foundation for the design of both types of assessment (see Section 2). Formative and summative assessments can differ in their uses, degree of contextualization, and forms of feedback, yet in its own way, each is consistent with, and takes advantage of, a common understanding of the targeted capabilities (see Oliveri et al., "expanded Evidence-Centered Design and Theory of Action Frameworks," this issue).

1.1.4 Implications of Changes in Assessment

Changes in what, who, and how (often) we assess hold implications for the collection of evidence, its analysis, and its interpretation. Traditional forms of assessment have relied heavily on overall scores of some sort, where it is presumed that individuals' abilities and individual item scores are ordered and monotonically related. Alternative measurement methodologies, some of which have been developed but are not frequently used in large-scale assessments, hold promise to provide information that aligns better with teachers' needs for finer-grained feedback at the student-level and for students engaging with assessment tasks. This article is one of three articles that describe such changes and present anticipatory frameworks to advance the design, measurement, and development of innovative forms of assessment that integrate learning and

[1] We refer the interested reader to Ellis and Larsen-Freeman (2009), Gee (1992), and Sperber (1996) for general background and Chalhoub-Deville (2003), Mislevy (2018), Moss et al. (2008), and Weir (2005) for discussions focusing on assessment.

assessment in the service of teaching, learning, and advancing opportunity to learn for all learners.

1.1.5 Running Example

To illustrate the approaches we proposed in the previous section, we use a running example that focuses on WEC. From a fine-grained methodological perspective, the WEC work encompasses considerations for the following:

- the context in which the activities are used,
- the scaffolding around the tasks in the activity,
- the articulation and delivery of feedback,
- the scoring and reporting of response patterns,
- the design of professional development activities for instructors,
- the experiences learners bring to the activity,
- the suggestion of secondary learning opportunities for students to develop these skills, and
- a commitment to improving the assessment so that additional opportunities to learn may be identified and realized.

In terms of these perspectives, consider Figure 1 as part of the WEC family of tasks. The figure shows an organization chart (revealing hierarchical structures), collaborative work environments (revealing collaborative diversity), and a task (in which a call for proposals is defined). For more on this task, see Oliveri et al., "expanded Evidence-Centered Design and Theory of Action Frameworks," this issue.

Figure 1

Scenario-Based Framing for Workplace English Communication (WEC)

Note. Reprinted by permission of Educational Testing Service. All rights reserved.

The scenario of this extended task is collaboration that leads to a proposal for a design of a modular home kitchen (see Oliveri, Slomp, Elliot, et al., this issue). Designing such a task itself requires multidisciplinary collaborations to connect the skills taught and assessed with societal needs and increase students' motivation and engagement. To motivate learning, the types of evidence collected should connect closely to the targeted skills assessed. They should also be communicated to stakeholders (e.g., students, teachers, program administrators, or prospective employers) to guide improvement and support learning. We suggest that these considerations need to be framed within a sociocognitive perspective for assessment, as we elaborate in the next section.

2.0 A Sociocognitive Perspective

In this section, we describe a sociocognitive psychological perspective and its application in service of teaching, learning, and assessment. In particular, we highlight LCS patterns and their implications on the nature of people's capabilities and how they acquire and use them. Understanding this perspective and its associated LCS patterns, activities, and interactions in the domain targeted for assessment is important to ground the designs used to guide the development of assessments and the methods used to measure skills to conceptualize and operationalize fairness in the design process (see Oliveri et al., "An Integrated Design and Appraisal Framework," this issue).

2.1 A Sketch of a Sociocognitive Perspective

The language and concepts of an emerging integration of individual, situative, and social perspectives on cognition, taken together, can be called a sociocognitive perspective. A sociocognitive perspective draws on research in fields including sociology, linguistics, cognitive psychology, learning sciences, and writing studies. The sociocognitive perspective concerns the complex adaptive systems in which individual persons develop and use capabilities to act and interact in the physical and social world (Gee, 1992; Sperber, 1996).

The *socio-* in sociocognitive highlights the knowledge and activity patterns that structure individuals' interactions with the world and with each other. These interactions occur as regularities among the myriad unique interactions occurring among people. These include the structures and ways of using language, knowledge representations, cultural models, and the patterns of activities of communities, personal interactions, classrooms, workplaces, and so on. We refer to these as linguistic, cultural, and substantive (LCS) patterns. (For more on LCS patterns, see Mislevy, 2018). Note that these "across persons" LCS patterns also encompass ways of thinking and acting, tools, and representations individuals use when engaging and working with each other.

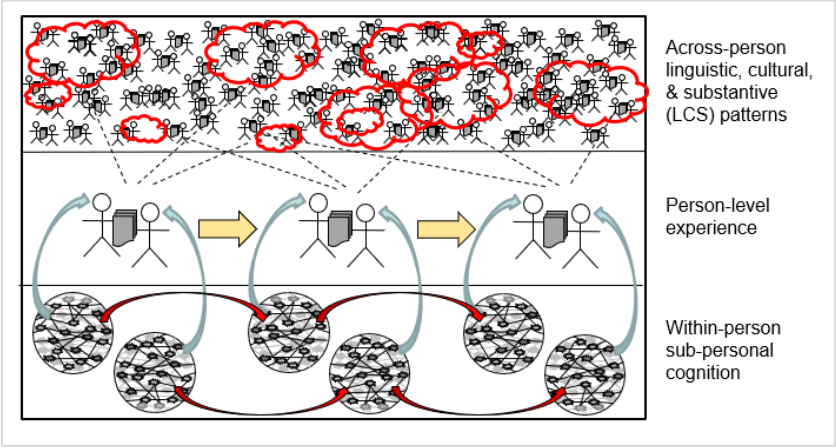
The *-cognitive* highlights within-person mental patterns occurring over an individual's experiences. These patterns are continually assembled, adapted, and revised to make meaning and guide action in each new situation. Richard Young (2009) uses the term "resources" to

describe individuals acquired cognitive patterns of knowledge, relationships, actions, feelings, and motives, together attuned to aspects of patterns across people's patterns as they have experienced them. We assemble resources in the moment to make our way through the world, extending, and adapting them in the process. Our experience of a given situation blends the particulars of that situation and resources that we have developed in previous experience.

Figure 2 depicts key ideas we will draw on in the sections that follow. The figure has three panels. The top panel shows across-person activities, which are regularities in the interactions of people in their overlapping identities and communities; such joint or community-level actions provide value or meaning to collectively conducted activities. The middle panel shows person-level activities, which represent the actions, events, and thoughts as we experience them as individuals. The person-level activities are mediated by extrapersonal (or across-person) activities shown in the top panel (Dennett, 1969). The bottom panel represents within-person (non-conscious) cognitive processes that give rise to an individual's actions occurring at the subconscious level. As people engage in activities and their meaning and significance is elevated through jointly developed activities, individuals become attuned to LCS patterns as well as the patterns' affordances and constraints, conditions of use, and how people use them to accomplish their goals and objectives. Along this line, individuals gain resources attuned to particular LCS patterns, which constitute an individual's capabilities to understand, create, and act in particular kinds of situations.

Figure 2

Across-Person, Person-Level, and Within-Person Levels in a Complex Sociocognitive Adaptive System



Note. Mislevy, 2018. Reprinted by permission of Educational Testing Service. All rights reserved.

Now that we have explained the sociocognitive system, let us also explain its complementary term, sociocultural perspective. We view the sociocultural and sociocognitive frameworks as complementary. Broad sociocultural knowledge and activity patterns help us understand the

unique ways that individuals structure cognitive patterns (Greeno, 2007). Social structures such as discourse communities enable individuals to acquire cognitive patterns that may, in specific circumstances, become resources that are tailored to the situation at hand.

Mediating the sociocultural and sociocognitive frameworks are LCS patterns. By nature, LCS patterns are not rigidly defined *things*, but regularities that arise, with variation, across individual people and activities. Neither should the LCS labels suggest crisp boundaries and strict categorizations. Moreover, although each person's resources are unique, as developed through unique histories of experience, it is the emergent regularities that enable us to talk and think and design assessments—in terms of *constructs.* Constructs concern certain kinds of resources that people might have, to do certain kinds of things, in certain kinds of situations (Mislevy, 2018). Recognizing the situated nature of language use, we build our thinking about designing and using assessments up from LCS patterns and resources, particularly as they connect to the WEC construct and the ways individuals expand learning. Table 1 gives LCS examples related to WEC construct elements (e.g., sharing technical information in various ways across discourse communities) that look ahead to our work.

Table 1

Linguistic, Cultural, and Substantive Patterns

Patterns	What we mean	General examples	Emails in WEC Tasks
Linguistic	Lexicon, discourse constructions, phonology, morphology, stance, grammar, and associated language patterns. Colloquially thought of as fixed and well-defined, but everyone's English is to some degree unique. Communication possible by regularities that emerge from communication activities. Associations among linguistic patterns, cultures, and discourse communities reveal situated language use.	*Broadly shared patterns*: Subject-verb-object sentence structure; construction of wh- questions. Regularities and local variations in pronunciation. In writing, expected word choice in a given discourse community. *Situated use patterns*: Shades of meaning among synonyms; different constructions across genres. In writing, word choice variation within a discourse community to produce a given aim.	Use of language forms in setting up formal project update meetings; sharing technical background information with colleagues before a project meeting. *Situated use patterns* (more variation): Setting up informal briefing meetings; sharing impressionistic, situation specific information with a close colleague before a briefing meeting.
Cultural	Frameworks of beliefs, knowledge, and activities, which are shared widely or locally, and shape actions and interactions. Cultural patterns, and the genres in which they are contained, are often constitutive.	*Broadly shared patterns*: What it means in a culture to be sick; graphs and how they are used; how to buy products in a store; the permission schema; Initiation-Reply-Evaluation sequence in classrooms. *Situated use patterns*: In writing, given and new genres as they are used in given communities of practice. These cultural patterns include schema selection, associated rhetorical moves, and value communication.	Structures of emails and variations of that genre based on aim and context, including subgenres of emails with their own regularities and variations. Variations across discourse communities as to frequency, tone, use vs. alternatives.
Substantive	Knowledge about what, who, how, why, and when activities and interactions exist, including "know what" and "know how."	*Broadly shared patterns*: Literature of quantum physics; background of interactions among workers in an office; values of the company one works for and needs and wants of clients. *Situated use patterns*: Language patterns, often specialized, including concepts, equations, and technical craft expertise. Substantive language patterns convey expertise and ability to complete successful transactions in a domain.	In email conversation about choices about designing a kitchen, for example, understanding refrigerator options, costs, function, and client needs. In terms of audience awareness, knowing what intended recipients know and don't know in order to choose what to convey and what can be assumed.

2.2 Importance of a Sociocognitive Approach to Teaching, Learning, and Assessment

Robert Stake (1991) captured an important implication of a sociocognitive perspective for teaching and learning: "The teacher sees education in terms of mastery of specific knowledge and sophistication in the performance of specific tasks, not in terms of literacy or the many psychological traits commonly defined by our tests" (p. 245). Table 2 lists the constructs we used in the email communication activities in the WEC prototype. It is organized as a hierarchy, with four higher-level constructs, which encompass finer-grain construct elements (finer skills) composing the higher-level constructs. A higher-level construct such as proficiency in "Substantive language alignment" comprises three finer-grained constructs: control of linguistic, vocabulary, and technical features of messages. These elements are described in general terms of controlling complexity, which involves making choices about words, constructions, and substantive content that are appropriate to situations, as seen through the lenses of LCS patterns implicated in the situation at hand. A student's performance reflects resources for implementing text features and recognizing situations to make such choices among them, as well as metacognitive resources to make choices that are appropriate to the particular context. Students develop all these resources concerning word choices, linguistic constructions, and email contexts—as well as metacognitive resources to activate them in context—only through experience, feedback, and reflection in particular contexts. In turn, assessment reports in terms of the higher-level construct can be useful for tracking progress across assignments or capabilities at the end of a course, but it is feedback at the finer-grain size on specific tasks—and noting how it relates to the higher-level conceptions—that informs experiences to develop those more general capabilities (see Zapata-Rivera et al., this issue).

From a sociocognitive perspective, building and using measurement models is a situated activity, crafted to suit the context, population, and purpose of a given assessment, in concert with the design of tasks, evaluation procedures, and reporting and uses of results. Different combinations of models, evaluation procedures, and reports are needed for different purposes, even as all are compatible with the view of the learning domain at issue. A common understanding of the nature of diverse LCS patterns, activities, and interactions in the domain can thus ground a variety of designs and analytic methods, each as suited to assessment contexts, populations, and purposes.

A sociocognitive perspective allows assessment designers to augment traditional approaches to conceptualizing and operationalizing fairness by incorporating differences among characteristics of individual test-takers and subpopulations in design and modeling adaptations. As pointed out by Oliveri, Lawless, and Mislevy (2019),

> While it is true that differences may exist in individuals' behaviors in real-world contexts, it is also true that commonalities in skill-related behaviors exist across cultures and populations. Note that by culture, we do not solely mean culture as determined by nationality (e.g., US citizens, Canadians, or Japanese), but we also

> suggest culture in terms of the places where learning may take place, instructional practices, and content domains. (p. 272)

In the WEC context, culture pertains to workplace environments as well, which vary by industries and roles (Haigler, this issue; Knoch & Macqueen, 2019).

Table 2

Constructs and Dimensions for WEC Email Scenario Tasks

Constructs and Dimensions	
Metacognitive Reasoning (MR)	Provides explanation for reasoning
Connects information presented earlier in the story (CIES)	Reflects on information provided earlier in the story's narrative (behaviors: connects answers to earlier events presented in the narrative)
Reevaluates prior decisions (EPD)	Reevaluates prior decisions based on information presented earlier (behaviors: changes order of answers)
Monitors own thinking (MOT)	Explanation provides evidence of monitoring one's thinking
Evaluates one's self-appraisal (OSA)	Explanation provides evidence of evaluation of one's critical self-appraisal
Sets goals (GS)	Explanation provides evidence of goal-setting (i.e., actions one is trying to achieve)
Critical Discourse (CD)	Provides evidence of an understanding of the relationship between interactions (e.g., supervisor/supervisee)
Understands Relationship Dynamics (RD)	Shows an understanding of relationship dynamics between interactants (e.g., language and conventions used differ when communicating with clients, supervisors, and colleagues)
Genre/Rhetorical Intent (GRI)	Answer provides evidence of proficiency in selection of features to communicate in the relevant genre
Uses concise language for genre (C)	Writes concise messages as required by genre (e.g., emails)
Uses correct structure for genre (S)	Structures emails appropriately
Frames messages appropriately (F)	Frames messages appropriately
Uses detail to convey purpose (P)	Uses enough detail to convey purpose of the message
Uses detail to convey context (Cx)	Uses enough detail to convey context of the message
Substantive Language Alignment (SLA)	Language used is aligned with the audience with respect to vocabulary/technical complexity
Controls linguistic complexity (LC)	Uses appropriate language complexity for audience
Controls vocabulary complexity (V)	Uses appropriate vocabulary for audience
Controls technical complexity (TC)	Uses appropriate technical complexity for audience

Measurement models can mediate between descriptive information that is garnered from assessments and purposeful information that is reported to teachers. In these uses, the selection of a measurement model must be sensitive to stakeholders' needs in order to be meaningful and useful to inform instructional practices. As we will show, understanding this connection between measurement models and users' needs is important to increase assessments' educative usefulness. In the WEC program, for example, we suggest that learners may benefit from finer-grained feedback from contextualized purposive activities on construct-relevant subskills. We propose that providing finer-grained information can be facilitated by capturing, interpreting, and reporting performance as seen through the lens of LCS patterns.

2.3 People Develop Resources Related to LCS Patterns in Context

Thinking in terms of LCS patterns and how individuals develop resources for engaging in practices that build around them is helpful for designing and using assessments. In addition to widely shared patterns that are involved in a certain kind of practice (such as communicating with informative emails), others that are necessarily involved will be more specific to situations, cultures, or fields of study. Realistic variations, which learners can begin to recognize and enact through their stabilities and variations, may be authentically represented in scenario-based tasks, and resulting performances may be evaluated in context. This understanding becomes ever more important in the assessment of complex constructs, which are designed to elicit test-takers' performances in more contextualized ways.

At the heart of a sociocognitive perspective is the dynamic interplay of persons acting through within-person resources, in situations organized around across-person LCS patterns—LCS patterns of many kinds, at many layers, combined and recombined in new ways in situations. In activities such as email communication that involve writing words and syntax, domains and situations, conversation structures, and conventions may be tailored to roles and purposes. They may also enable individuals to coordinate their thoughts and actions. Such resources are unique to individuals, for every person's lived experience is unique, and similarities among individuals' resources depend largely on the similarities in the LCS patterns involved in situations they have experienced. It is these similarities, and resulting similarities in resources we develop, that make it possible for us to interact in recurring social situations.

Linguistic patterns, for example, span lexicon and syntax; there are typical uses of given words and structures, but every situated meaning depends on context, users' intentions and listeners' expectations, and the interpersonal functions they serve. Genres are examples of cultural patterns; they encompass, in substantiative ways, typical ways people structure writing, and they are intertwined with kinds of purposes and uses people have in recurring kinds of social situations. Finer-grained linguistic patterns as to word and grammar choices tend to be used in certain ways in different genres, such as an informative email as opposed to a cover letter for a job application.

Substantive patterns address knowledge structures and activity structures in the social and physical world, from repairing toasters to greeting friends to writing a proposal for a kitchen design. In every situation, we perceive, act, and interact by assembling resources attuned to LCS patterns of many kinds and at different levels and blending them with the situation at hand. We provide examples of the application of LCS patterns as they appear in workplace emails.

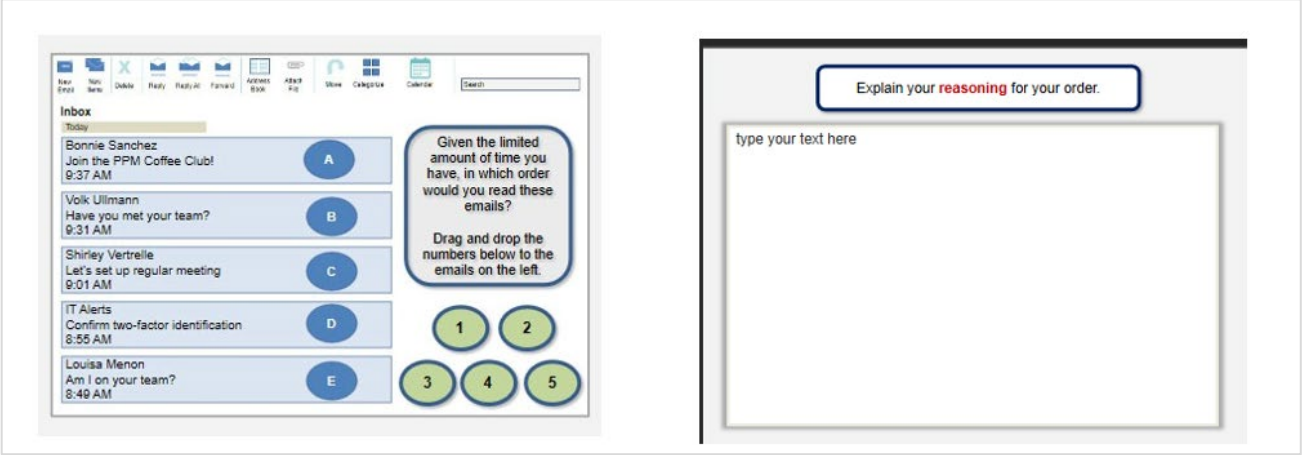
A notion with critical implication for education and assessment is that cognitive resources are initially bound closely to the conditions of learning. Regarding learning, people develop more general capabilities only through further experiences that reflect unique instances of the same LCS patterns, and they develop more effectively with support, sequencing, feedback for reflection, and higher-level language for recognizing kinds of situations and knowing the kinds of things one can do in them. As James Gee once put it, "Abstract representations of knowledge, if they exist at all, reside at the end of long chains of socially and technologically situated activity" (as cited in Hickey, 2011, p. 139).

This view had implications for the design of the WEC scenario-based tasks. We not only integrated key elements of relevant situations and activities into the tasks, but we included metacognitive questions to encourage students to think and reflect on the choices they were making, and that could be made, to communicate in the unbounded space of "situations like these." We also considered these elements in developing automated feedback such that learners did not receive information on correct/incorrect responses but rather were shown how to consider the context of their communication—the audience, the situation, and the genre in which they were communicating.

Figure 3 shows an example of the explicit metacognitive questions we included to support student learning of the relevant linguistic and substantive aspects of WEC. As shown, students are asked to prioritize emails and then to provide reasons for that prioritization. The metacognitive questions provided opportunities for students to consider their uses of the language and the demands of the situation asked in particular items.

Figure 3

An Example of a Metacognitive Task in WEC



Note. Reprinted by permission of Educational Testing Service. All rights reserved.

3.0 Writing Analytics: Application of LCS Patterns to WEC Modules

In the context of communication, we will elaborate on targeted capabilities that involve higher-level LCS patterns. Opportunities to learn and practice WEC—and associated metacognitive abilities in contextualized communicative situations—are useful for classroom practice, which can be labeled as performance-oriented and construct-oriented feedback.

For performance-oriented feedback, each task's design and tailored evaluation frame help structure learning activities and feedback with respect to specific tasks and performances. For construct-oriented feedback, patterns in performance are expressed in a more general evaluative framework. The framework is consistent with those for tasks that synthesize information across tasks and over time in terms of person variables. These may be simple aggregations that are amenable to classical test theory or expressed as variables in more complex models. Attention to the ways LCS patterns vary across particular classrooms, cultures, and programs or fields of study within professional technical communication subjects allows for adaptation of tasks, evaluation procedures, and psychometric models, all while maintaining the focus on targeted capabilities.

For the current project's focus on workforce communication, then, the ability to analyze LCS patterns is important for fairness, as individuals have their own signature/use of language, problem solving approach, and way of interacting with information. One-way-fits-all may not be the most culturally-sensitive approach to make sense of students' responses and leads to highly restrictive teaching and learning opportunities. It is also important for students' opportunity to learn, for when students are provided with highly restrictive approaches to learning, they may disengage and become uninterested (Moss et al., 2008).

3.1 Uses of LCS Patterns

We highlight the use of LCS patterns because they shape how we approach communicative situations, how we understand them, how we create them, and how we interact within them. 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, varying LCS patterns across subgroups within these broad contexts affect understandings and expectations within communicative situations, which, more broadly construed, are common across groups. Further, individuals' unique experiences shape their access to both broader and subgroup LCS patterns and the expectations and understandings that accompany them.

Students will develop some broader resources with respect to certain LCS patterns, for example, as they experience a series of situations involving the targeted patterns, each in its own unique way. Ideally, in each case, they get feedback on how each experience is an example of something describable at a higher level, and reflect on the details of their performance through the lens of the higher-level description. This kind of feedback is very useful for learning. It enables extending and generalizing resources further to increase the chances of activating them in future situations in which they could be useful.

LCS patterns allow us to go beyond coding for the presence and absence of correct or incorrect responses and allow us to track students' uses of those patterns. Then, levels of performance are noted based on LCS-pattern uses within and across work products with examples of when they were or were not used within the situation (genre) in question.

The following example, shown in Figures 4 and 5, illustrates an application of the LCS-pattern analysis to an email response, which can be useful for classroom practice through the use of a rubric. Of particular interest here is the DocuScope Classroom platform described by Helberg et al. (2018), which is an implemented example of this kind of automated analysis and instructional use. The original DocuScope is a text analysis and visualization platform developed at Carnegie Mellon University to statistically identify patterns and rhetorical strategies in texts (Ishizaki & Kaufer, 2012). It can code documents using some 22,000 semantic categories and over 58 million uniquely classified words and phrases. Using DocuScope Classroom, a student's writing document—say a business email of a certain kind for a certain purpose—can be parsed and tagged for the student in genre-relevant categories. For instruction, the individually-tagged document is displayed, and the distributions across categories are compared with the distribution of category use in documents of the same subgenre. The goal is reflective practice: "to help students see their writing decisions by making students aware of the genre-specific linguistic and rhetorical features they use" (Helberg et al., 2018, p. 40). The value of DocuScope is its ability to allow researchers to build dictionaries of key language patterns—words, phrases, constructions, syntactic structures—and establish categories of each of the patterns as they are commonly used in some targeted area—say emails to schedule a meeting or cover letters for a job application.

This example demonstrates extension of the directions of LCS patterns in a domain and context that differs from (expands) those investigated in Mislevy (2018).

Figures 4 and 5 illustrate a student report assignment and corresponding self-reflection assignment. Note that "moves" are tagged portions of code that reflect LCS patterns, and the report shows how the presence of such moves in the student's work compare with other documents in the genre. This is not an evaluation but a prompt for awareness of the rhetorical tools used to effect writing goals in the genre (LCS patterns within LCS patterns). One important lesson for novices is to build awareness of how rhetorical, lexical, and linguistic constructions are used differently in different genres and situations.

Figure 4

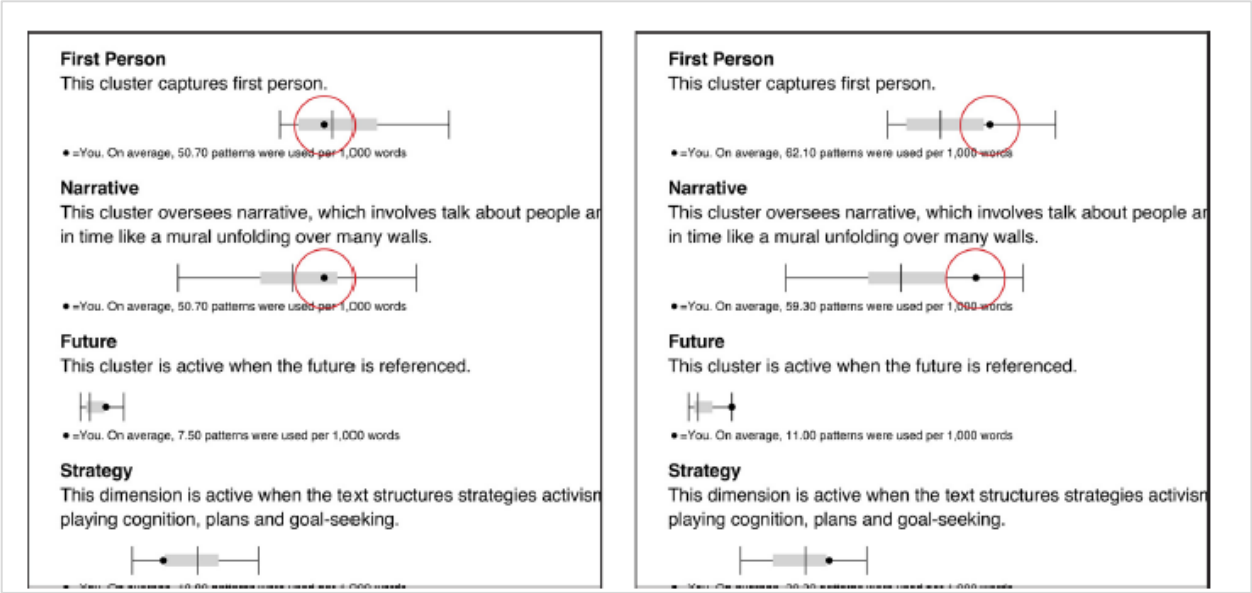
DocuScope Self-Reflection Assignment

Genre "Move"	Codes Used	Example from Your Text
A. Positioning self in relation to potential employer	"First Person," "Future," "Interactivity," "Confidence High"	***I am*** [First Person] ***confident*** [Confidence High] ***that I*** [First Person] ***will be*** [Future] ***an asset*** [Confidence High] ***to your*** [Interactivity] ***company.***
B. Show how takeaways and results from past experiences relate to future job	"First Person," "Future," "Positive Values," "Strategy"	***The technical expertise I*** [First Person] ***gained in hands on*** [Strategy] ***classes as well as the skills*** [Positive Values] ***I*** [First Person] ***gained in collaborating will allow*** [Future] ***me*** [First Person] ***to make a positive impact as an intern at your company.***

Note. From Helberg et al. (2018). Reprinted by permission of Elsevier. All rights reserved.

Figure 5

DocuScope Report



Note. From Helberg et al. (2018). Reprinted by permission of Elsevier. All rights reserved.

LCS patterns are all about persons acting in situations; thereby, looking at an email through the LCS lens as a capture of such actions involves being informed by persons' LCS-related resources, in the situation, for the purpose. It also encourages us to evaluate a response not simply as if it were a thing, but rather to evaluate a series of situated actions: the awareness of aims, options, and conventions to be integrated with the particulars of the people, context, and subject matter of the email. We want to examine how the action was (or was not) seemingly guided by standards of effectiveness in such situations in terms of LCS patterns people draw on to understand and act in such situations.

In some applications, we may be able to capture and scrutinize the actual series of actions. In others, we may have drafts and final products, which illuminate some higher-level thinking; traditionally, we have only final products. Even with final products, we can talk about them as resulting from target knowledge and metacognition, and we can reflect on the choices and rationales that produced them.

As instructors and raters, we understand this intuitively when we talk in shorthand about *scoring a response*. The use of LCS patterns connects more deeply to the psychology and sociology of situated activity, hence learning. Moreover, we need to better understand and explicate the reasoning if we are to automate some construct-valid approximation of it.

3.2 An Illustration from WEC

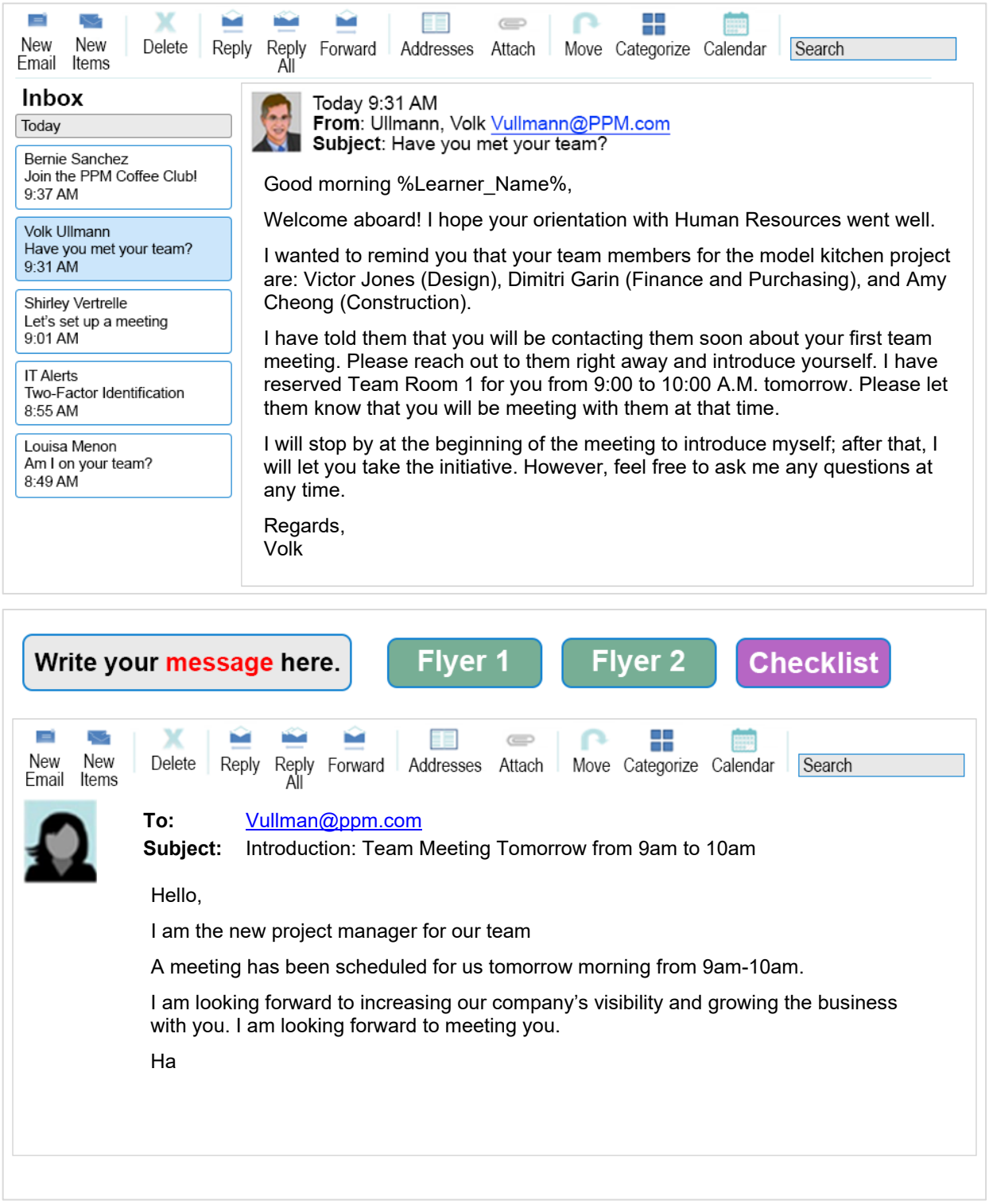
With the foregoing discussion in mind, let us discuss an example that relates to business emails as a mid-level LCS pattern, with connections to conditions of use and standards of what works, what is to be communicated, and how to accomplish it in context. The *how to accomplish it* aspect of the issue draws on finer-grained sized LCS patterns. The patterns overlap with their uses in other higher-level genres (such as social emails) and involve considerations for how to use linguistic structures, tone, content, and so on. At this level, one thinks about what is useful in task/performance-specific feedback for learning purposes but can be collapsed into coarser categories or be input for observable variables in psychometric models. The goal is to summarize information across performances or over time. In Section 4, we will also see how the LCS lens informs the content and form of the information in assessment design representation forms such as Q-matrices and extensions of them.

Figure 6 is an email that a student constructed in response to the new manager's need to invite the team members to a kick-off meeting. Note that an "invitation" is not quite the same when it comes from your manager or from your friend. We want students to become aware of the social implications of each different audience and become adept at using alternative resources for each type of audience.

In terms of genre, we can classify the text as an email and the subgenre of scheduling. If we parse the text for its LCS patterns, we can identify key verb patterns ("has been scheduled") under our linguistic framework. We could also train our language-based algorithms to look for specific information or numbers within the linguistic structure.

Figure 6

Sample Email Task and Student Performance



Note. Reprinted by permission of Educational Testing Service. All rights reserved.

With respect to evaluating substantive patterns —the "know what" and "know how" aspects— we may wish to look at the third paragraph in the student's email. The sentence "I am looking forward to increasing our company's visibility…" reveals the use of an advanced form of know-how. The use of the linguistic and cognitive patterns establish competency at the level of an advanced writer—in this case, a writer who seeks to project the present ("looking forward") and align that future event with the visibility and growth of the company (both values associated with mutual profit). Only a sophisticated writer could make the move shown in the third sentence. Thus, to one of the editors of this special issue who is a rhetorician, it indicates a substantive rhetorical move.

We can consider (at least) four kinds of evaluations of a performance of this kind, all grounded in analysis through a sociocognitive perspective and the applicable LCS patterns. The evaluations can be distinguishable by their nature, grain size, potential usefulness, and potential as input to psychometric modeling. These evaluations include:

1. A task- and performance-specific qualitative analysis like the one in the preceding paragraph. It can be supplemented by the evaluation of specific features that make explicit additional lexical, grammatical, and inclusion/exclusion choices. These choices, in the genre and by virtue of the assembly, made the response effective. This evaluation is at the appropriate grain size and focus for at least two audiences. These are: an expert's analysis in a research study and the conversations to be contemplated between the student and a peer or teacher, both of which may have differing styles and formality levels.

2. The use of automated tagging followed by categorization of specific lexical, grammatical, and rhetorical features employed in the email. The DocuScope Classroom example illustrated this type of tagging with distributions of codings of an individual's use of email features as compared to a larger corpus of email codings. This example shows writers how their email composition choices and use of LCS patterns statistically compare to how others use LCS patterns in this type of situation.

3. A human assignment of an overall score according to a rubric, yielding a score of "advanced" or "4 points" in a 4-point rubric. This scale is most like the scoring employed in large-scale tests. This approach may not be particularly useful for feedback for learning. However, it may be based on an expert's analysis of LCS pattern usage, either implicitly, or, as in this project, explicitly.

4. An automated overall score, trained on expert raters and predicted by the feature codings in #2 above and supplemented by the natural language processing (NLP) features used in standard automated scoring programs. Of particular interest would be the degree to which the LCS features of #2, which are instructionally useful, factor into the prediction equation. In this way, we would see ways in which automated scoring could capture some of the information of expert observations of #1.

The LCS-pattern framing gives us a deeper understanding of elements of the interplay between the many-layered social practices and personal cognitive resources in the person-acting-within-situation. It not only helps us understand the practices and how people develop proficiencies (Lave & Wenger, 1991), but allows for deeper understanding of learning and assessment. Consequently, it allows us to move beyond familiar language and practices and the provision of more descriptive feedback about language uses to increase assessment's educative usefulness.

4.0 Integrating Cognition, Assessment Design, and Analysis

In the previous section of this article, we provided an overview of how LCS patterns can ground formative assessment and feedback to support learning. In this section, we discuss how the same LCS-pattern starting point can be used to ground measurement models in terms of constructs that are consistent with the formative uses.

Measurement models in education consider features of students' specific performances as evidence for inferences about higher-level constructs. The models also allow the inclusion of variables representing aspects of capabilities or tendencies of what students know or can do in more general terms than descriptions limited to specific, realized assessment performances. These higher-level constructs may be cast very broadly, such as the historical measures of, say, verbal ability. They also may be more focused, such as the extent to which students are able to recognize communicative goals for business emails and marshal linguistic and cultural knowledge to achieve such goals. The use of LCS patterns to obtain more specific information about performance thus applies to measurement models, first in evaluating performances, then as capabilities or propensities as they might apply to future actions, past learning, or prospective learning opportunities.

We begin this section of our article with a brief overview of cognitive diagnosis modeling in its most basic form, relate it to a sociocognitive perspective, and discuss extensions that build around sociocognitive analyses. A sociocognitive perspective makes us aware of the nature of LCS patterns and activities in some targeted domain, and what people need to learn to do in what kinds of situations, and the kinds of capabilities learners develop. The elements of tasks, performances, and potential constructs would support building a variety of assessments that would all be grounded on the same sociocognitive analysis of a targeted domain, but could suit different purposes, populations, and contexts.

The idea is not to think in terms of designing an assessment and a model, but rather designing an assessment-design space and a conformable model-design space. In other words, we want to understand the domain, the involved capabilities, and situations of use in a way that allows us to design a range of tasks and assessments that can be tuned to different needs and populations, and evaluation procedures and specific measurement models tuned to the populations and purposes of interest.

The basic setup of cognitive diagnosis modeling (CDM) shows how these essential, interconnected aspects of activity can be coordinated explicitly in a measurement model in simple assessments. We then suggest and illustrate directions in which to extend the CDM paradigm to more complex assessments. The section that follows then discusses how the ideas connect with measurement modeling.

4.1 Basic Elements of Cognitive Diagnosis Models and Q-Matrices

The binary skills or DINA (deterministic-input, noisy "and" gate) model (Rupp et al., 2010) illustrates the central idea in its simplest form. We provide a non-mathematical description and present an example to illustrate the model. (Figure 7 provides the equations for the interested reader.)

Figure 7

The Mathematical Form of a Binary Skills, or DINA, Cognitive Diagnostic Model

In the basic binary skills model, the probability distribution of X_{ij}, the 0/1 response of Person i to Item j, depends on a vector of 0/1 unobserved variables for persons' skills, $\phi_i = (\phi_{i1},\ldots,\phi_{iK})$, and a vector of 0/1 known variables for item features, $q_j = (q_{j1},\ldots,q_{jK})$. Here ϕ_{ik} indicates whether Person i has Skill k and q_{jk} indicates whether Item j requires Skill k. The matrix of q_js for all items is denoted Q. Define the indicator $\xi_{ij} = \prod_k \phi_{ik}^{q_{jk}}$, which boils down to 1 if Person i has the skills Item j requires and 0 if not. The link function is

$$\text{Prob}\left(X_{ij} = 1 \mid \phi_i, q_j, \pi_{j0}, \pi_{j1}\right) = \left(1-\pi_{j1}\right)^{\xi_{ij}} \pi_{j0}^{\left(1-\xi_{ij}\right)},$$

where $(1-\pi_{j1})$ is the "true-positive" probability of a correct response to Item j when a person has the skills it requires, and π_{j0} is the "false-positive" probability of responding correctly when he does not. Conditional independence is assumed over persons and items. Denote by $F(\phi)$ the distribution of ϕ in a group of persons.

When there are multiple groups, Q_g is the task-feature matrix for Group g; different sets of skills may be required in different groups. $F_g(\phi)$ is the distribution of ϕ in Group g, and $(1-\pi_{gj1})$ and π_{gj0} are the true- and false-positive probabilities for Item j in Group g.

Note. From Mislevy (2018). Reprinted with permission from Educational Testing Service. All rights reserved.

Consider a domain of wrong/right, or 0/1, test items that are all built around three solution rules, say A, B, and C. We might propose that a student either does or does not know how to apply each of the rules. Each item either does or does not require each of the rules; and students are likely to get an item right when it requires only the rules they have mastered and wrong if they have not mastered a rule.

Table 3 is the so-called Q-matrix for a five-item test. The rows are for items and the columns are for skills; the cell entry is 1 if the item requires the rule and 0 if it does not. Item 1 requires Rules A and B, Item 2 requires Rules B and C, and Item 3 only requires Rule A. A student who can apply Rules A and B but not C would probably get Items 1 and 3 right but miss Item 2.

Table 3

An Example of a Q-Matrix

	Solution rules		
Items	**A**	**B**	**C**
1	1	1	0
2	0	1	1
3	1	0	0
4	1	0	1
5	1	0	0

Although the items, the responses, and the psychology in this example are all simple, what is important from a sociocognitive perspective is that this basic Q-matrix depicts essential interrelationships among features of task situations, expected performance, and capabilities of persons. The columns indicate salient characteristics signaling LCS patterns that can be built around the situations being assessed. In this example, they are the demands for using rules for solving the items.

We will see that the Q-matrix form can be extended for more complicated situations and LCS patterning. It is assumed in applications of the DINA model that correctness is the only feature of item performance that matters, so the Q-matrix does not address the evaluation of performances. In the DINA model, there is a one-to-one match between rules that items can require and skills that students can apply. The term "attributes" is used to talk about both, with attributes of items referring to features of the items that require corresponding skills to solve. After looking at an example from WEC, we will discuss ways these ideas and representations can be extended to more complex assessments (see, for instance, the description of data cubes in Oliveri et al., "expanded Evidence-Centered Design and Theory of Action Frameworks," this issue).

4.2 An Illustration from WEC

The selection of a particular set of attributes (e.g., in WEC, the selection of skills and subskills as shown in a Q-matrix) reflects our analysis of the domain and the elements identified empirically and theoretically to be relevant to the communication needs in the workplace (i.e., see Corrigan & Slomp, this issue; Oliveri & McCulla, 2019). Table 4 provides a simplified Q-matrix as it was applied to the WEC modules. Developing the modules allows us to think about the interplay

between locally grounded, yet theory-based, assessment for learning and reporting through construct-based measurement models grounded in the same theory. These two kinds of assessment—assessment for learning and assessment for reporting—are often thought of as quite distinct, perhaps even incompatible. We are working in this project to flesh out the articulation, a goal reflected in, for instance, the influential volumes *Knowing What Students Know* (Pellegrino et al., 2001) and *How People Learn II: Learners, Contexts, and Cultures* (National Academies of Sciences, Engineering, and Medicine, 2018; see Oliveri et al., "expanded Evidence-Centered Design and Theory of Action Frameworks," this issue for an elaboration of going beyond binaries dividing formative and summative assessments and our proposal to instead focus on the commonalities between these two types of assessments).

Rather than providing simply 0/1 item scores or overall task ratings, WEC views the response process through the lens of LCS patterns within work products, which holds implications for scoring responses and reporting the scores to students. This means creating an LCS-based score report that is situated within the genre and dependent on context. As illustrated with the DocuScope Classroom example, local, fine-grained reports describe individual- and group-level uses of patterns that are genre-dependent—that is, based on the genres or work products with which students work. Higher-level constructs are grounded on this and additional information such as natural language processing text features, which can be used for reporting across tasks or over time. These modeling processes can be facilitated through the use of Q-matrices to indicate which clusters of LCS patterns are involved in the tasks. For more complex tasks such as those in WEC, information beyond the basic Q-matrix is needed.

Table 4

Simplified Q-Matrix Showing Constructs and Dimensions for WEC Email Scenario Tasks

	Metacognitive reasoning					Critical discourse	Genre/Rhetorical intent					Substantive language alignment		
Item	**CIES**	**EPD**	**MOT**	**OSA**	**GS**	**RD**	**C**	**S**	**F**	**P**	**Cx**	**LC**	**V**	**TC**
1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	1	1	1	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0	0	0	0	0	0
5	0	1	1	1	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	1	1	1	0	0	1	0	0	0
7	0	0	0	0	0	1	1	1	0	0	0	0	0	1
8	0	0	0	0	0	0	0	0	1	1	0	1	1	0
9	0	0	0	0	0	0	0	0	1	1	0	1	1	1
10	0	0	0	0	0	0	0	0	0	0	0	1	1	0

Note. This table shows which attributes are implicated in a particular item. Constructs and subsidiary dimension labels are detailed in Table 2. There are 10 items and 4 attributes in this example, which is a simplified version of one used in the actual assessment. An entry of 1 indicates that the attribute is implicated in the item, whereas an entry of 0 indicates that it is not.

4.3 Extending the Ideas Behind the Q-Matrix

In this section, we point out directions along which the defining feature of a Q-matrix—charting the relationships among task features, performance feature, and persons' capabilities— can be extended to more complex assessments. It is easiest to describe them one at a time as enhancements of a basic Q-matrix as in Table 3 as might be implemented in a spreadsheet. An actual implementation involving multiple extensions, however, would be better served by a structured database designed from the start to accommodate all of the fundamental relationships and accompanying information (e.g., Riconscente et al., 2005). Various kinds of relationships can be built in, along with views that make it easy to create, modify, manipulate, and present information as needed by various users and processes:

- *Nesting within column structures.* The columns of the basic Q-matrix encode salient features of items. In DINA models, the features are always used in both task design and measurement modeling, but we can relax these constraints. As discussed above, construct-relevant situation features can be hierarchical, and LCS-based specifics of finer-grained features that constructs encompass can depend on genre, purpose, or other contextual features. For a given design domain, we can determine the form of a column structure which can be expanded, collapsed, or augmented as needed to characterize tasks and to inform local reports or measurement models for different assessments within its scope. Consider, for example, the genre of email construction. A higher-level basic column can indicate whether or not email construction is an attribute in a task. If no emails are to appear in tasks for a given assessment, subconstructs are not relevant, and their columns can be collapsed. If emails are relevant in the assessment, finer-grained construct-based columns such as those in Table 4 can be useful for task design and perhaps rubric construction, but are not needed for measurement modeling as in a large-scale assessment. If only selected subgenres of emails are to be addressed, only LCS analyses needed for crafting rubrics and tasks within these subgenres need be detailed out.
- *Columns for socioculturally-relevant features of tasks with potential implications for design or modeling.* While the LCS patterns that are the focus of learning may be relevant for all students in an intended use, other non-focal LCS patterns may present options for aligning task features to students' backgrounds. Examples of the non-focal LCS patterns may include vocabulary, background materials on the focal topic, and supporting instructional materials. Both evaluation procedures (e.g., country-specific rubrics for informational emails) and interaction terms in measurement models (e.g., group-specific task parameters) can be employed to increase fairness and to focus on focal LCS patterns and capabilities (Mislevy, 2018). One potentially can conceive an infinite set of features. Thus, a Q-matrix starting with core columns related to common features can be augmented with

columns that are expected to be relevant with the population at issue or are discovered in data analyses.

- *Nesting within row structures, 1: Task dependencies.* One can also have rows nested within tasks that indicate subparts or phases of tasks. Examples are included in the scenarios in WEC. Such dependencies can be tracked in measurement models for so-called testlet effects (Wainer et al., 2007).
- *Nesting within row structures, 2: Work products.* The basic Q-matrix presumes a simple one-to-one relationship between items and work products: a single response unique to that item. For more complex tasks such as investigations or scenarios that require successive design adaptations, there can be multiple work products within tasks or connecting multiple tasks. Examples would be a vector of responses to a cluster of items, a written constructed response to a prompt, chat messages among students, the trace of steps in solving an extended problem, and a metacognitive reflection as in the WEC modules. One can also have subsidiary rows in a basic spreadsheet indicating the associated tasks, along with analogous Q-matrix cells indicating competency and/or LCS pattern clusters being addressed.
- *Nesting within row structures, 3: Evaluated performance features.* The basic Q-matrix presumes a simple one-to-one relationship between work products and evaluated features: a single evaluated response variable unique to that work product (and associated item[s]). For assessments in which multiple features are to be evaluated from one or more work products, one can also have subsidiary row observables. A subsidiary row would indicate from which work product(s) it is derived and include a link to descriptions of the nature of the feature(s) and associated evaluation procedures. It would also include potential interactions with sociocultural student-background variables, such as linguistic background. The nature of the features can include dichotomous, polytomous, categorical, count, or response time, categories as needed by the measurement model. Observables could be designated a priori or could be discovered in data mining after trials, as through machine learning or cognitive interviews. In other words, this would be an expandable set.

5.0 Measurement Modeling Considerations

A predominant view of assessment is that a test is a collection of tasks, often test items, usually self-contained, to respond to in a designated set of circumstances. Scores for each task indicate the success of each answer, such as right or wrong or an essay score determined by a rubric. A test score combines the task scores, usually simply as a total or average. Scores are taken to be a measurement of students' capabilities.

Alternative conceptualizations emerge in light of rapid developments that have occurred in areas central to assessment. These developments have led to new understandings of cognitive and social underpinnings of human learning and activity, and technical advances for interactive environments and big data analytics have opened up new opportunities to enrich our understanding of how humans learn as they engage in digital activities. As this article and the companion articles show, the WEC program of work draws on many of these advancements.

But there have been advances in assessment and educational measurement as well, some of which we have already mentioned. Other advances have been developing in the measurement and psychometric literature for decades beyond the horizon of public awareness. A fundamental advance is viewing assessment not simply as measurement but in terms of evidentiary argument, situated in social contexts, shaped by purposes, and centered on students' developing capabilities for valued activities (Mislevy, 2018).

We have alluded to some of the measurement modeling techniques that have become available to support reasoning with new forms of assessment. Using such models effectively, in context, for particular purposes, in light of what is known about students, is more than a matter of using more complicated models. Effective use is a matter of developing assessments from a deeper understanding of targeted capabilities and how students develop them. Concurrently, it involves crafting assessment and learning situations that are grounded in these understandings and tailored to context and use. Only when these elements are in place, are we in a position to jointly assemble measurement modeling procedures that fulfill our aims. These may involve using more complicated models in some cases, but sometimes also familiar ones. These considerations have implications as we elaborate on seven aspects of a sociocognitive paradigm at the confluence of measurement modeling and design choices. We describe each of these seven aspects next. And where possible, we illustrate these implications by referring to the WEC modules.

5.1 When is Measurement Modeling Useful?

We suggest that the utility of measurement modeling can be augmented when scores are used for feedback purposes, particularly when we distinguish between task- and construct-level information for communicating assessment results. In a framework grounded on a sociocognitive analysis of some targeted domain and capabilities, such as WEC, reports can be cast at both the task and construct level. Both types of reports can be cast in terms of capabilities that concern students' resources for recognizing and acting with LCS patterns that shape the domain. They can also be set in terms of metacognitive skills for marshalling such resources in unique situations.

Task-level feedback concerns a student's performance on a given task, describing, critiquing, comparing to other exemplars, or reflecting on choices and suggesting alternative moves. These performances can be particularized in terms of the LCS patterns in the targeted domain. This kind of reflection is particularly beneficial to learning. It provides information about the realized

performance and its evaluation and helps learners become attuned to and gain facility in using particular LCS patterns. Measurement modeling such as generalizability analyses and comparisons across cultural groups are useful to study the characteristics of such task-level analyses. The benefits are better understanding students' performances on particular tasks. As we saw earlier in Figure 5, DocuScope Classroom provides reports of particularized performances. These are useful because of the fine-grain level of feedback they provide to students to gain awareness of their use of LCS patterns.

The value of task-level reporting is that performance in distinct tasks can be mapped into a common, more detailed, LCS-based framework, which in turn relates to instruction and reflection. As examples, this performance mapping can be done using ratings on rubrics linked to reporting variables and employed as directed in Q-matrices or extensions of them. Then, the matrices can be used to categorize the feedback into reporting on various aspects of performance that reflect different profiles, styles, or approaches used in email composition (see Zapata-Rivera et al., this issue).

In contrast, construct-level reports synthesize information from performance across multiple tasks or tasks over time during the course of learning, in terms of students' capabilities as defined at some higher level of abstraction. For instance, coarser-level feedback may be obtained by controlling for lexical features in email communications, for example, and the aggregated information may not clearly reveal differences in students' uses of specific resources or patterns. As a result, the construct-level reports may provide broader tendencies in performance as cast at a higher level—information that is less detailed when it comes to individuals' immediate learning but more useful for broader purposes such as curriculum planning, progress monitoring, or program evaluation.

5.2 What Kinds of Models are Available?

Earlier in the article, we used a CDM as a prototypical measurement model that can formally represent the relationships among sociocognitive theory, task design, and aspects of students' capabilities. Researchers have developed a wide variety of models incorporating and extending these core ideas. These models vary in relation to the types of student, task, and observable variables they can employ and whether they can capture longitudinal and/or hierarchical data structures. Examples of such models include extensions of CDMs (Rupp et al., 2010), similarly structured item response theory (IRT) models such as explanatory IRT models (De Boeck & Wilson, 2004), exploratory structural equations models (Asparouhov & Muthén, 2009), Bayesian inference networks (Almond et al., 2015), and hidden Markov models (Vermunt et al., 1999).

5.3 Can Measurement Models Account for Diverse Student Backgrounds?

Among the available families of models mentioned above are ones which can model differential patterns of performance among students with differences in known background variables. For example, multiple-group or IRT mixture models can be used to discover either

theoretically-predicted (using known groups) or empirically-discovered (using latent groups) differences due to students' use of differential background knowledge, problem-solving approaches, or strategy usage (von Davier & Carstensen, 2007). Person-level analyses, about which we will say more shortly, are also measurement-modeling tools to detect atypical patterns of performance beyond whatever is built into a model.

A cautionary note, however, is that one should not expect that more complicated measurement models alone can handle more complex assessment situations. Instead, developing more complex assessments requires richer understanding of a domain and learning principles associated with mastering the targeted domain. Moreover, it involves aligning the learning and assessment situations and the inclusion of principled design of tasks that include opportunities to track students' performance and learning in the domain. Constructing models suited to populations, contexts, and purposes builds on such a foundation. Sometimes a model will be constructed on some of these more complicated components. At other times, a suitable model may simply involve summing item scores.

Returning to our discussion of LCS patterns, we note that one needs to consider potential group differences that relate to learning and assessment situations early in the design process (Mislevy, 2018). In designing tasks, for example, certain LCS patterns that are targets of learning should, of course, pose relevant challenges in tasks. Others, however, that vary among students can be adapted so they do not present irrelevant sources of difficulty.

In reference to the WEC modules, if the kitchen appliances and configurations used in the tasks are unfamiliar to a group of students, the particulars of the task can be adapted accordingly without changing the focus of the email communication features. Similarly, it may be the case that evaluation procedures in the same version of a task need to be varied in different cultures. For example, Campbell (1998) found that expectations differed in Japanese and American companies as to the amount of background information that was critical but already known to recipients that was expected to be shared in a business memorandum.

5.4 What Student Variables Should Be in a Model?

Continuing the topic of tailoring a model to a population, context, and purpose, we return to our discussion of Q-matrices and elaborate on expandable/collapsible/augmentable Q-matrix columns. While the basic DINA model has a one-to-one match between features of tasks and aspects of student capabilities, the columns of a Q-matrix can be thought of as suggesting a space for student variables in a measurement model.

An extreme case occurs when a model needs only one student variable, such as the overall proficiency variable θ in a univariate IRT model, to serve the assessment's purpose. This variable indicates a propensity to get right answers or write effective responses for tasks in a domain, such as a broadly-cast *ability to write effective emails*. The Q-matrix could still guide task assembly, and score reports could still provide students, teachers, and educators with

information for overall comparisons among groups or summary-level trends in performance over time.

Finer-grained reports could be based on finer-grained constructs, as suggested in Table 2. However, as Sinharay et al. (2011) note, one needs to carefully consider the value of the information provided for various inferences at the more detailed scores given. One may wish to be selective with respect to the aspects of the construct that provide finer-grained information on particular construct elements while high-level results are reported for other construct elements. Other constructs that feature prominently in the domain need not be included in a measurement model. Such informational limits apply *even if they are critical to performance*, as when the user knows that the students being assessed are already proficient in that critical skill.

5.5 How Does Sample Size Affect Modeling?

The measurement models developed over the past few decades offer ways of accommodating many kinds of complicated cognitive and social factors that are important in learning and add to our understanding of students' performance. One can fall into the trap of assuming that *since we can model anything, we can model everything*. Additional parameters and structures generally require more data. The previous paragraphs mentioned the seductive attraction of more finely-grained skill reports, which, despite their compelling face validity, can turn out to have even less information about a targeted proficiency than an overall score. Here, assessment design and measurement tools can be of great help, as when we use a Q-matrix to check how well constructs are represented in tasks, or when we examine standard errors of student variables or generalizability over tasks in construct-based reports.

5.6 How Can Person-Fit Residual Analysis Add Value?

Even with careful assessment design, there can be too many attributes in a Q-matrix or too few students in a sample to fit as comprehensive a model as one might like. One measurement-modeling strategy that can be employed in this situation is person-fit residuals. We can describe it most easily with a unidimensional IRT model for 0/1 items, then note the extension to more complicated models.

In a unidimensional IRT model, each student's overall proficiency in a domain of items is represented by the single person parameter θ. Each item is represented by a parameter, in the simplest model, the item difficulty β. A sociocognitive perspective tells us there are actually many LCS patterns involved in comprehending, operating on, and responding to even the simplest item, so no single item parameter can provide a complete description of their varying mixtures of demands. Similarly, individuals have their own unique constellation of resources, and they vary as to what is hard and what is easy, what they know and what they do not know, and the ways they interact with each of the items. Nevertheless, the IRT model gives a probability for a correct response to each item j as a function of person i's θ and item j's β, say $p(\theta_i,\beta_j)$. A person-fit residual is basically a difference between the predicted response from the

model and the actual response. When this model fits data well, the probability is usually a pretty good description of patterns in the data, for example, with high predicted probabilities of a correct response when people have high estimated θs and items have low βs. One can then examine residuals with respect to every column of the Q-matrix. One can see if there are patterns, student by student, of unexpectedly large residuals. Examples are students doing much worse on the subset of items that requires knowledge of kitchen appliances, say, or in items that require expected forms of address, when compared to their performance on the entire module (Mead, 1976). Systematic tendencies of certain features—any particular one, or any particular combination of them—that render items unexpectedly difficult for some group of students are revealed, and, if they are very surprising, for individual students.

The same approach applies when more complicated models are used. We calculate person residuals for every observable variable and check for unusual response patterns for groups and individuals related to attributes that are not in the model, given the attributes that are in the model.

5.7 Can Learning Be Incorporated into Models?

The sociocognitive perspective on learning described in the opening of the article noted the interplay between task-specific reports for reflection and across-time based on constructs. It is worth mentioning two approaches that emphasize these uses.

In domains for which learning can be approximated by learning progressions and finer-grained constructs, one can apply the tenets of the learning progression to the definition of student variables and task features, and rubrics or scoring procedures to define observable variables. The BEAR Assessment System (Wilson & Scalise, 2006) has been developing measurement models in concert with classroom practices along these lines.

The expanded Evidence-Centered Design (e-ECD) for learning and assessment systems is a framework for incorporating learning goals and processes within assessment design (Arieli-Attali et al., 2019; see also Oliveri et al., "expanded Evidence-Centered Design and Theory of Action Frameworks," this issue). In addition to the assessment-design structures of evidence-centered design, e-ECD integrates a model for learning and measuring progress over time. The authors have applied the framework using a hidden Markov model to characterize change over time, which includes parameters for modeling the effects of instruction.

6.0 Conclusion

From a sociocognitive perspective, building and using measurement models is a situated activity, crafted to suit the context, population, and purpose of a given assessment, in concert with the design of tasks, evaluation procedures, and reporting and usages of results. As this article demonstrates—and this special issue reveals—an application of a sociocognitive perspective involves different combinations of models, evaluation procedures, and stakeholder reports.

Put straightforwardly, the instantiation of Figure 2 into a formative assessment such as WEC is an act of extraordinary complexity. Within WEC, the targets of metacognitive reasoning, critical discourse, genre/rhetorical intent, and substantive language alignment shown in Table 2 are themselves far from traditional written communication targets of essay writing such as organization, critical thinking, and knowledge of conventions. Moreover, use of a Q-matrix, such as those shown in Tables 3 and 4, further complicates the application of sociocognitive perspective in terms of target knowledge. It is no small challenge to design an assessment in which tasks are designed to elicit appropriate vocabulary for an audience in a given setting—and then to relate that vocabulary use demonstration to an understanding of relationship dynamics. Is it indeed true that the student who writes "I am looking forward to increasing our company's visibility" in the task shown in Figure 6 can also demonstrate the ability to navigate the hierarchical email-ordering scenario shown in Figure 3? Do these domains move together? Separately? Is one prerequisite to the other? Do these abilities differ by subgroup and, if so, what do we do with that information?

Such are the complex challenges of applying a sociocognitive perspective to scenario-based assessments such as WEC. Addressing these challenges requires new ways to think about assessment design and information use. We close by identifying two ways forward to meet these challenges.

First, multidisciplinary collaboration is necessary in applications of sociocognitive perspectives to complex scenario-based tasks. Psychometricians, cognitive scientists, task developers, and subject matter experts must, from the beginning, work in non-hierarchical ways using ECD and e-ECD frameworks, as well as theory of action perspectives (see Oliveri et al., "expanded Evidence-Centered Design and Theory of Action Frameworks," this issue). This kind of work will also require new project management strategies, including cost analyses, marketing strategies, and sustainability plans to ensure project success.

Second, in cases where applications of sociocognitive perspective are used for formative assessment, it is best to acknowledge that new pedagogies will have to be developed to ensure project success. A moment's consideration of the screenshots of linguistic patterns from DocuScope shown in Figures 4 and 5 lead one to realize how little we know about pedagogies involving using real-time corpus analysis with students. What exactly does an instructor do to help a student realize that text structure activities in a given assignment do not align with those in a larger corpus? How, exactly, does one go about even explaining text structure as understood through corpus techniques? This kind of work will require new ways of thinking about instruction and assessment in digital environments, especially when the assessment advertises itself as self-contained and providing the necessary pedagogy to students.

We opened our article with an observation: Amidst a time of change, the way we assess is concurrently shifting. We close by wondering if seismic shifts can result in a new way forward for education as sociocognitive perspectives continue to develop—and to be applied to complex scenario-based tasks.

Author Biographies

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

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.

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

References

Almond, R.G., Mislevy, R.J., Steinberg, L.S., Williamson, D.M., & Yan, D. (2015). *Bayesian networks in educational assessment.* New York: Springer-Verlag.

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 to incorporating learning goals and processes within assessment design. *Frontiers in Psychology*, *10*, 853.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equations Modeling*, *16*, 397-438.

Campbell, C. P. (1998). Rhetorical ethos: A bridge between high-context and low-context cultures? In S. Niemeier, C. P. Campbell, & R. Dirven (Eds.), *The cultural context in business communication* (pp. 31-47). John Benjamin.

Casner-Lotto, J., & Barrington, L. (2006). *Are they really ready for 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

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

Corrigan, J. A., & Slomp, D. H. (2021). Articulating a sociocognitive construct of writing expertise for the digital age. *The Journal of Writing Analytics*, *5*. De Boeck, P., & Wilson, M. (Eds.). (2004). *Explanatory item response models: A generalized linear and nonlinear approach.* Springer-Verlag.

Dennett, D. (1969). *Content and consciousness.* Routledge.

Ellis, N., & Larsen-Freeman, D. (Eds.). (2009). *Language as a complex adaptive system.* Wiley.

Embretson, S. E. (Ed.). (1985). *Test design: Developments in psychology and psychometrics.* Academic Press.

Gee, J. P. (1992). *The social mind: Language, ideology, and social practice.* Bergin & Garvey.

Gordon, E. W., & Heincke, P. (Eds.). (2013). *To assess, to teach, to learn: A vision for the future of assessment*. Gordon Commission on the Future of Assessment in Education.

Greeno, J. G. (2007). Toward the development of intellective character. In E. W. Gordon & B. L. Bridglall (Eds.), *Affirmative development: Cultivating academic ability* (pp. 17-47). Roman & Littlefield.

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

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

Helberg, A., Poznahovska, M., Ishizaki, S., Kaufer, D., Werner, N., & Wetzel, D. (2018). Teaching textual awareness with DocuScope: Using corpus-driven tools and reflection to support students' written decision-making. *Assessing Writing*, *38*, 40-45.

Hickey, D. T. (2011). Participation by design. In D. M. McInerney, R. A. Walker, & G. A. D. Liem (Eds.), *Sociocultural theories of learning and motivation: Looking back, looking forward* (Vol. 10, pp. 137-161). Information Age Publishing.

Ishizaki, S., & Kaufer, D. (2012). Computer-aided rhetorical analysis. In P. M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing: Identification, investigation and Resolution* (pp. 276-296). IGI-Global.

Knoch, U., & Macqueen, S. (2019). *Assessing English for professional purposes*. Routledge.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge University Press.

Mead, R. J. (1976). *Assessment of fit of data to the Rasch model through analysis of residuals* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global database. (UMI No. 302785354).

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

Mislevy, R. J., & Elliot, N. (2020). Ethics, psychometrics, and writing assessment: A conceptual model. In J. Duffy & L. Agnew (Eds.), *After Plato: Rhetoric, ethics, and the teaching of writing* (143-162). Utah State University Press.

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

National Academies of Sciences, Engineering, and Medicine (2018). *How people learn II: Learners, contexts, and cultures*. The National Academies Press.

National Association of Colleges and Employers (NACE). (2018). *Job outlook 2019*. https://www.odu.edu/content/dam/odu/offices/cmc/docs/nace/2019-nace-job-outlook-survey.pdf

Oliveri, M. E., Lawless, 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.

Oliveri, M. E., & McCulla, L. (2019). Using the occupational network database to assess and improve English language communication for the workplace. *ETS Research Report Series*, *2019*(1), 1-17.

Oliveri, M. E., Mislevy. R. M., & Elliot, N. (2020). After admissions: What comes next in higher education? In M. E Oliveri and C. Wendler (Eds.), *Higher educational admission practices: An international perspective* (pp. 347-375). Oxford University Press.

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

Oliveri, M. E., Slomp, D. H., Rupp, A. A., & Mislevy, R. J. (2021a). Principled development of workplace English communication Part 2: Expanded evidence-centered design and theory of action frameworks. *The Journal of Writing Analytics*, *5*.

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

Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment.* National Research Council's Committee on the Foundations of Assessment. National Academy Press.

Riconscente, M., Mislevy, R. J., & Hamel, L. (2005). *An introduction to PADI task templates* (PADI Technical Report 3). SRI International. https://padi.sri.com/downloads/TR3_Templates.pdf

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

Sinharay, S., Puhan, G., & Haberman, S. J. (2011). An NCME instructional module on subscores. *Educational Measurement: Issues and Practice*, *30*(3), 29-40.

Sperber, D. (1996). *Explaining culture: A naturalistic approach.* Blackwell.

Stake, R. E. (1991). The teacher, standardized testing, and prospects of revolution. *The Phi Delta Kappan*, *73*(3), 243-247.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20,* 345-354.

Vermunt, J. K., Langeheine, R., & Böckenholt, U. (1999). Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, *24*, 179-207.

von Davier, M., & Carstensen, C. H. (Eds.). (2007). *Multivariate and mixture distribution Rasch models.* Springer.

Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications.* Cambridge University Press.

Weir, C. J. (2005). *Language testing and validation.* Palgrave McMillan.

Wilson, M., & Scalise, K. (2006). Assessment to improve learning in higher education: The BEAR Assessment System. *Higher Education*, *52*, 635-663.

Young, R. F. (2009). *Discursive practice in language learning and teaching.* Wiley-Blackwell.

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*.