Communicating Assessment Information in the Context of a Workplace Formative Task

Diego Zapata-Rivera, *Educational Testing Service*

Jessica Andrews-Todd, *Educational Testing Service*

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

Structured Abstract

- **Background**: This article describes a theoretically grounded approach to scoring and communicating assessment information to teachers and students in the context of a prototype used to assess Workplace English Communication (WEC). WEC includes skills needed to succeed in a digitally evolving workplace that increasingly requires such skills to carry out workplace-relevant activities. We elaborate on issues relevant to the design and evaluation of score reports for formative purposes, the design of the WEC prototype including the construct, tasks and scoring approaches used to produce profiles of students based on their response behavior patterns (Andrews-Todd et al., 2018), and the process of designing prospective score reports for teachers and students in the WEC context.

- **Questions Addressed**: We address the following questions: What are key aspects to consider in the design of prospective score reports for teachers in the context of a learning and assessment tool? What scoring approaches can be used to produce the information included in the score reports? Which approaches can be used to design and evaluate prospective score reports with subject matter experts?

- **Conclusions**: To conclude, we discuss the potential of using innovative scoring approaches in assessments of complex constructs in digital environments. The goals are to support the development of score reports that teachers and students can use to inform current and future learning of WEC.

Keywords: prospective score reports, scoring approaches, Workplace English Communication, writing analytics

1.0 Background

Determining scoring and reporting approaches for complex constructs in digital environments can be challenging, as such constructs are often process oriented and multivariate in nature. Digital environments present advantages such as allowing for more available data to describe learners' capabilities (e.g., click streams, communication behaviors). Process data such as click streams and keystroke data can be used to uncover various strategies or approaches students use during the writing process (Guo et al., 2019). The availability of such data vis-à-vis the constructs targeted for assessment can shape the kinds of approaches desired and most appropriate for scoring and reporting, particularly when there is a need to provide score reports to multiple audiences (i.e., different stakeholder groups; for a fuller discussion of the targeted audiences, see the companion articles in this special issue, especially Oliveri, Mislevy, & Slomp, 2021 and Oliveri et al., 2021a, 2021b).

In this article, we describe approaches to reporting scores on complex constructs in digital environments. Among other things, the reporting of scores depends on the scoring approaches taken. We describe two scoring approaches that represent a simple and complex mode of scoring. First, given the simplicity often desired for stakeholders in score reports, one approach includes simple descriptives that provide frequencies corresponding to the extent to which learners demonstrate skills of interest. A second, more complex approach includes profiles which provide a way to characterize learners at a high level based on their response behavior patterns. Both approaches can provide the simplicity needed to convey progress and performance information about learners to stakeholders in score reports.

Following a framework for designing and evaluating reports facilitates the process of producing initial reports ("prospective score reports" as described later). By following the iterative evaluation process described in the framework, new versions of the reports will be developed and used to support users' decision-making. It is worth noting that this type of work usually requires the participation of an interdisciplinary group of people with expertise in assessment design, human-computer interaction, information visualization, psychometrics, and score report research.

Clear communication of assessment results to particular audiences is an important aspect of assessment validity (Tannenbaum, 2019). Formative tasks have the potential for providing

actionable information to various stakeholders (Andrade et al., 2019; Hopster-den Otter et al., 2017; Zapata-Rivera, 2011; Zapata-Rivera et al., 2018; Zapata-Rivera & Katz, 2014). Assessment information provided to stakeholders should support understanding of the reported information in accessible ways and minimize the misuse of assessment results. These goals are important not only to measurement specialists but also to the various experts involved in the design, development, and use of assessment results and those involved in interpreting assessment results to inform decision-making such as classroom instruction or placement decisions.

We discuss these issues in six sections. In this first section, we have provided background information related to the importance of clearly and coherently reporting scores to various stakeholders. Section 2.0 provides an overview of research on score reporting and frameworks used for designing and evaluating score reports, with a focus on prospective score reports. Section 3.0 describes design aspects of the Workplace English Communication (WEC) prototype, including the construct, tasks, and approaches used to score the tasks. Section 4.0 provides an overview of scoring approaches used to score items in prototypes of WEC skills; two strategies are described and illustrated. Section 5.0 provides a sample of prospective score reports that may be used in the context of the WEC prototype. The last section, the conclusion, provides an overview of the article and describes future directions for this research.

2.0 Research on Score Reporting

Research in the area of score reporting includes work such as the development of frameworks for designing and evaluating score reports (e.g., Hambleton & Zenisky, 2013; Zapata-Rivera et al., 2012). It also involves designing score reports for particular audiences (e.g., Hambleton & Slater, 1997; Zapata-Rivera & Katz, 2014), evaluating comprehension and preference aspects of representations used to communicate score report information (Zwick et al., 2014), and evaluating the effectiveness of ancillary materials (Goodman & Hambleton, 2004; Zapata-Rivera et al., 2016). Other research on score reports includes designing score reports for formative purposes (e.g., Brown et al., 2019; Zapata-Rivera, 2011; Zapata-Rivera et al., 2018).

The use of digital assessments enables us to gather both process and response data. These data are useful to inform teaching and learning, which can be facilitated through a richer understanding of students' learning processes or, in the case of WEC, the type of skills in which students need support at a finer-grain level.

Reporting systems should be designed to provide the intended audience with information that can be used to support decision-making. In order to provide audiences with the right type of information, the design of the assessment should be aligned with its purpose and the type of assessment results produced by it. Previous work on sharing insights of process data with teachers as a means of augmenting the information in the score reports showed that teachers appreciate information that provides additional context to the assessment results available in reports (Forsyth et al., 2017; Kannan et al., 2018; Zapata-Rivera et al., 2018).

Fortunately, there is guidance on the process of creating these reports. This process usually involves the contributions of an interdisciplinary group of experts who follow a user-based, iterative design and evaluation framework.

2.1 Frameworks for Designing and Evaluating Score Reports

Several frameworks for the design and evaluation of score reports have been proposed (e.g., Hambleton & Zenisky, 2013; Zapata-Rivera et al., 2012). These frameworks usually take into account the type of information that will be shared with the intended audience and apply design principles from areas such as information visualization, human-computer interaction, and cognitive science to produce report mockups that are evaluated in an iterative cycle through think-aloud studies, focus groups, or large-scale studies. These studies usually explore issues such as comprehension, usability, and appropriate use of assessment information.

User needs and expectations are considered from the beginning of the assessment design process. Zapata-Rivera et al. (2012) describe the use of prospective score reports (PSRs) as a way of gathering assessment information needs from the intended audience. This information is used to guide assessment design decisions.

2.2 Prospective Score Reports

Prospective score reports (PSRs) are shared with the assessment development team and used to identify potential misalignments between the desires of the audience and the type of assessment information available as part of the assessment. PSRs do not need to be functional reports. Simple sketches or mockups can suffice. PSRs may include assessment data from a student or data aggregated from a group of students. They may include a variety of information, including definitions of skills and subskills, performance levels, scores, task-level information, and information that explains the components of the PSR. Other phases of the framework include reconciling user needs with available assessment information and performing internal (with the development team) and external (with the intended audience) evaluations.

The next sections describe design aspects of the WEC prototype, including the construct, tasks, and approaches used to score the tasks. Subject matter experts (SMEs) who will be using this prototype with their students have participated in all phases of this project. They have also contributed to the design of the PSRs presented here.

3.0 The Workplace English Communication Modules

The modules discussed in this paper can be used within a blended-learning environment to teach and assess students' proficiency in WEC. The digital activities were designed to provide students with an opportunity to learn a range of communication tasks grounded in concrete communicative situations, such as writing emails to discuss proposals, schedule meetings, and solve work-related problems.

In the prototype modules, learners are provided with opportunities to solve problems occurring in a company setting. They are asked to collaborate with a supervisor, colleagues, and external clients to design and develop a proposal for a kitchen. The modules do not require specialized knowledge of kitchens or design; a kitchen was used to frame this scenario as most people have a mental notion of a kitchen (i.e., a place used to cook food).

During the activities, learners solve various problems that require them to carry out authentic communication tasks such as responding to emails. Student-centered goals for completing these activities include supporting the learning and experiencing of the type of language used in the workplace, including pragmatics, situational awareness, and negotiation turns used when interacting with others and solving problems. The objectives are to enable learners to extrapolate principles for problem-solving communication tasks that are novel to them. The activity designers' objectives are to assist learners in improving skill acquisition and obtaining more elaborate feedback on their skills.

Preparation to communicate effectively at work requires an understanding of the key job-related demands; communicative needs of the workplace; and explicit strategies for how to navigate communications across various audiences such as with one's supervisor, colleagues, and external clients. Successfully carrying out these activities requires expanding the construct of English language arts and writing to include an expanded skillset that occurs in the workplace (e.g., communication and collaboration skills), as discussed later in this article.

3.1 Constructs Measured

Table 1 lists the dimensions and subdimensions assessed in the prototype. These dimensions aim to capture critical aspects of communications occurring in the workplace. (See Corrigan & Slomp, 2021, this issue and Oliveri, Mislevy, & Slomp, 2021, this issue for an expanded description of each of these dimensions and subdimensions.)

In this article, we illustrate scoring and score reporting considerations for these dimensions with an application to scoring emails. Email is a form of workplace communication that is widely used and very important to carrying out workplace activities, yet students are underprepared with respect to this kind of communication (Oliveri & McCulla, 2019). Thus, the application of these dimensions to email writing are described next.

Table 1

Dimensions of the Targeted Construct

Level 1	Level 2	Level 3
Metacognition: Knowledge or awareness about the test takers' understanding of their responses and the ability to articulate their understanding **Language (Critical Discourse):** Critical discourse describes the need to understand the relationship between individuals working together, such as one's relationship with a supervisor, colleagues, or subordinates; that is, understanding one's role in the hierarchy of an organization.	**Rhetorical Aim:** Learners demonstrate awareness and understanding of how to structure text to achieve their goals and the choices made to deliver on their intention (e.g., word choice, diction, content, organization of the email, sequence of presentation of the information). **Discourse Community:** Learners understand socially accepted behaviors, language use, signs, and ways of thinking that identify oneself with a group or community.	**Genre Knowledge:** Learners understand the features of the form of communication and select the length of the message and the features of the genre accordingly. Learners demonstrate an understanding of the structure, organization, and substance of the type of communication. **Substantive Knowledge:** Learners are able to understand the content of the communication and the level of detail or context needed to convey the necessary information based on contextual cues and their understanding of the organization/context. **Communication Task Process:** Learners are able to use context information to understand, incorporate, and respond to information conveyed in the communication in ways that are appropriate to fulfilling the task.

3.2 Example: The Email Writing Task

In the email task, a component of the WEC module, test takers are presented with several emails from team members requesting information or asking for help with administrative activities. They are asked to respond to each email, considering issues such as the appropriate level of formality and language norms given the context and the target audience (Oliveri, Rupp, & Slomp, 2020).

3.2.1 Metacognition

With respect to the email writing task, metacognition involves consideration of information presented in the communications (e.g., emails), the ability to re-evaluate decisions made in earlier drafts of the emails, and flexibility to revise/change information. It also involves explaining one's reasoning and monitoring one's understanding of the choices made, such as the rhetorical goals used in email writing.

3.2.2 Language (Critical Discourse)

In email writing, language (critical discourse) involves creating texts or communications that show an awareness of social/historical relations and the distribution of power within an organization or company. For instance, in the prototypes, learners are asked to write emails that demonstrate their understanding of their role and responsibilities in prioritizing communication or deciding which level of formality to use in the communication (e.g., more or less formal depending on whether they are writing to their boss, external client, or colleague).

3.2.3 Discourse Community

In email writing, discourse community involves learners' understanding of social aspects of the community (e.g., accepted behaviors, language use, signs) and use of this knowledge to write email messages that clearly place them as members of the community (e.g., as an administrative assistant or a project manager).

3.2.4 Rhetorical Aim

In email writing, the use of rhetorical skills would demonstrate appropriate selection of the length of the message; the structure of the text; and linguistic moves that convey the use of appropriate word choice, organization of the text, and the sequence of the information presented in light of writers' understanding of who they are communicating with and for what purposes. This element goes beyond discourse community because the former implies an understanding of the relationships involved in the communication whereas the latter implies the ability to communicate the messages at the appropriate level.

3.2.5 Genre Knowledge

With respect to the email writing task, learners should demonstrate understanding of the genre used in the communication (e.g., the nuances of appropriate use of the greeting and subject line, sender, CC, BCC, time stamp, and closing) and their communicative purpose and audience.

3.2.6 Substantive Knowledge

Substantive knowledge in the email writing task involves opportunities for learners to infer the substance and importance of an email from cues (e.g., the subject line and the sender) and inform their decisions regarding how to prioritize information based on such cues.

3.2.7 Communication Task Process

In the context of the email writing task, communication task processes involve opportunities for learners to review communications, select sequences of activities, and justify these sequences based on particular criteria. They are able to review information presented in earlier screens (or earlier correspondence) and connect it to the current problem or issue they need to address.

3.3. Description of WEC Module Items

The module used in this study to discuss scoring and score reporting contains 46 items. It requires approximately 90 minutes to complete on average (although it is not a timed test). It was developed in consultation with experts including professors of professional and technical communication, professors of assessment of writing, measurement specialists, and assessment developers with expertise in the development of innovative assessments such as the Programme for International Student Assessment (PISA) collaborative problem solving. The module was administered online.

Item types in the assessment vary, including selected-response, drag and drop, constructed-response, and open-constructed response (student provides an open-ended response, usually involving higher-order cognitive activity) items. More information on this module can be found in Oliveri, Rupp, and Slomp (2020).

Metacognition was assessed using constructed-response items. The items asked learners to describe their rationale for their selection of particular answers or their decisions with regards to revising items or using particular language features in their correspondence with different avatars representing different roles in the company (e.g., external clients, colleagues, or a supervisor).

A combination of items (i.e., constructed-response and selected-response) was used to assess various aspects of language including rhetorical aim and genre knowledge. Learners were asked to interact with continuous text addressing different types of speech acts (narration, exposition, argumentation, and other types of prose), non-continuous text (graphs, forms, and lists), or mixed (containing both continuous and non-continuous) text.

4.0 Overview of Scoring Approaches

This section provides an overview of approaches used to score items used in WEC prototypes. Two strategies are discussed and illustrated: one that uses simple descriptions of behaviors and another one that is based on score profiling.

4.1 Strategy 1: Simple Descriptions of Learner Behaviors

One strategy for providing information to students or instructors about learner behaviors is to provide simple descriptions of behaviors. These descriptions can come in the form of frequencies or counts related to the extent to which learners engage in targeted behaviors or proportions that take into account the learner's total for all targeted behaviors or behaviors for a population of interest (e.g., other students in a classroom). In carrying out this strategy for a process-oriented competency like WEC, one could engage in qualitative coding with transcriptions of learner communications and task log files to identify each instance where a learner demonstrates a targeted behavior from the construct model in Table 1. With the frequencies approach, where each learner would have a count for how often they engaged in behaviors along targeted dimensions, this would result in a set of frequencies for each learner for each behavior of interest.

This is a time-consuming process that requires training and calibrating raters. However, there are strategies for improving the process, including coding a representative sample of responses and making use of automated scoring approcaches (e.g., Latent Semantic Analysis) to automatically score responses and identify cases that require human scoring (e.g., when scoring agreement with a human is low; Bauer & Zapata-Rivera, 2020).

We illustrate this approach using data from a prototype used to assess collaborative problem solving (CPS) skills in which students completed a three-person collaborative simulation-based task on electronics concepts. This task was used in college-level engineering and electronics courses given curriculum needs for supporting students' communication and collaboration skills in preparation for the workforce. In the task, each student worked on a separate computer that ran a simulation of an electronics circuit board. All three circuit boards were connected to form a series circuit, and students were tasked with communicating via text chat and performing a range of actions (e.g., calculations, measurements, resistor changes) to reach a specified goal voltage value across each circuit. In one form of scoring, raters analyzed task log files by labeling each row of log data as evidence of one of nine CPS skills from a CPS ontology, which provided a construct and evidence model for CPS (Andrews-Todd & Kerr, 2019). The CPS ontology included a social dimension corresponding to collaboration behaviors (i.e., maintaining communication, sharing information, establishing shared understanding, and negotiating) and a cognitive dimension corresponding to problem-solving or taskwork behaviors (i.e., exploring and understanding, representing and formulating, planning, executing, and monitoring).

This strategy of analysis results in a corresponding frequency for each of the nine CPS skills for each individual student, which can provide basic descriptive information related to the extent

to which each learner is demonstrating each behavior of interest. These behaviors can be visualized using bar charts which can graphically display multivariate data. In such charts, each bar corresponds to a CPS skill, and the height of the bar corresponds to the magnitude of the variable. An example in Figure 1 shows the frequencies for each CPS skill for Learner 1 and compares that to the class average. This visualization can help bring to light patterns in individuals' behaviors. For example, in Figure 1, we can see that Learner 1 tends to demonstrate more cognitive CPS behaviors and fewer social CPS behaviors relative to the class on average.

Figure 1

Visualization of Frequency Strategy in the Context of Collaborative Problem Solving
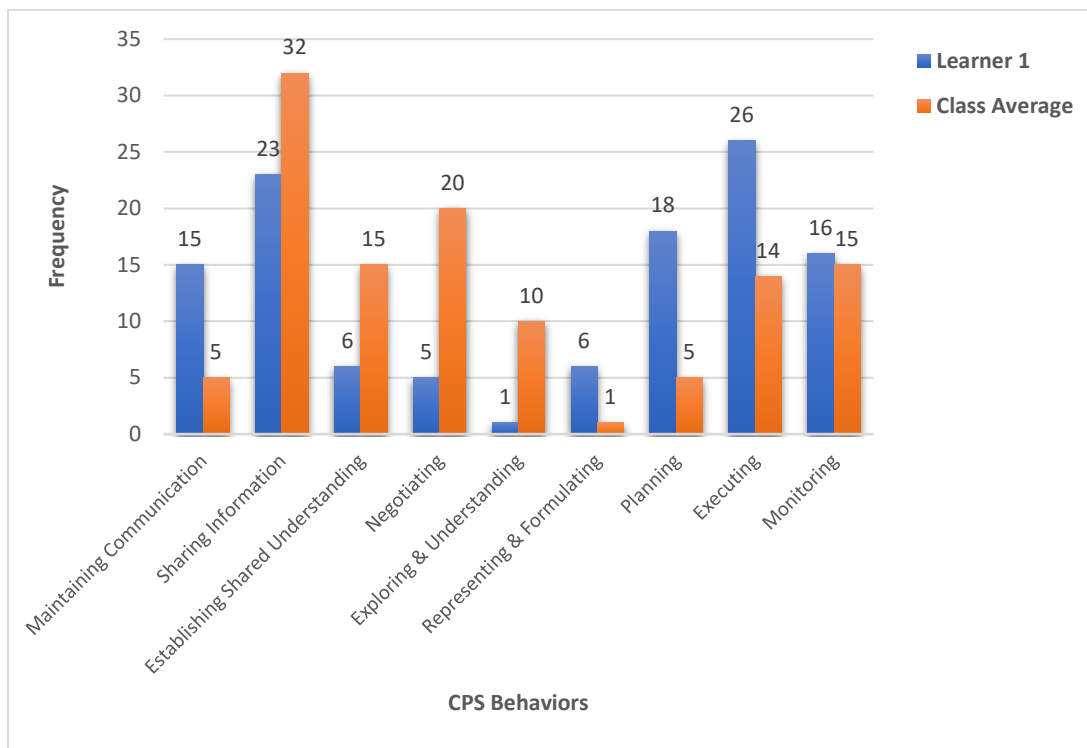
4.2 Strategy 2: Score Profiling

Another strategy for providing relevant information to students and instructors is a profile approach which provides a way to characterize at a high level how learners are behaving (Andrews-Todd & Forsyth, 2020). An often-used method for developing profiles, particularly with big and complex data, is cluster analysis. In cluster analysis, individuals or observations are grouped in such a way that those similar to each other, but different from others, are identified and put into a group or cluster (Romesburg, 2004). This strategy builds on the frequency strategy described in the prior sub-section (see Strategy 1). In continuing with the CPS example from the prior sub-section, we could similarly conduct qualitative coding with the task log files to

generate frequencies for each CPS behavior for each individual in the sample. An exploratory clustering method can then be conducted using the resulting CPS skill frequencies. The final number of clusters can be determined based on how each resulting cluster aligns with interpretations from relevant theory and existing literature. Such literature can be used to name and characterize each cluster according to relevant behaviors existing in each cluster.

In carrying out this profiling strategy on prior CPS work with the electronics task, a four-cluster solution emerged as most appropriate with respect to interpretation related to literature in collaboration and psychological research. The four clusters were named Social Loafers, Active Collaborators, Group Organizers, and Chatty Doers based on key behavioral patterns associated with each cluster. Social Loafers included individuals who displayed below average demonstration of all targeted CPS skills. Active Collaborators were just the opposite and displayed above average demonstration of most CPS skills. Group Organizers included individuals who displayed CPS behaviors associated with establishing and maintaining organization of the problem and the group. Chatty Doers included individuals who displayed a high level of maintaining communication CPS behaviors which were content irrelevant and social in nature, but also a high level of executing behaviors (i.e., actions and communication used in the service of carrying out a plan) relative to other individuals (Andrews-Todd et al., 2018; see Forsyth et al., 2020 for additional information on how clustering approaches have been used to derive student profiles in the context of CPS).

In the context of the WEC prototype, cluster analysis can be similarly applied to carry out the profile strategy. Human or machine-driven approaches can be applied on response (e.g., constructed-response items) and process data (e.g., click streams) to generate frequencies for each targeted behavior (e.g., metacognition, genre knowledge, communication task process). Specifically, those frequencies can correspond to the number of times a learner correctly displayed each WEC skill when given the opportunity. Cluster analysis can create profiles that characterize different types of learners in terms of their strengths and weaknesses with particular WEC skills. For example, a profile may emerge in which learners are weak in areas associated with the metacognitive dimension, but strong in areas associated with the language (critical discourse) and substantive dimensions. An alternative profile may include learners who are strong in all three dimensions. These sorts of profiles can create meaningful groups that individuals interacting with the WEC module can be placed in depending on how they perform. These profiles can be included, with corresponding descriptions, in score reports that can be used by students and instructors to receive feedback about progress and performance.

5.0 Sample Prospective Score Reports in the Context of the WEC Prototype

Although different types of PSRs can be designed and used to document assessment information needs of relevant stakeholders (e.g., teachers, administrators, and students), in this article we focus on PSRs for two types of audiences: teachers and students. Reports for formative purposes

are designed to provide teachers and other stakeholders with ongoing feedback about student learning.

Information provided in the reports is used by teachers to inform instruction and students to guide their learning (Brown et al., 2019; Zapata-Rivera, 2011; Zapata-Rivera et al., 2018). Zapata-Rivera et al. (2012) show the use of formative hypotheses to provide teachers with information about possible next steps based on students' performance on a test. They also provide teachers with information about appropriate and inappropriate uses of assessment information. This is an important feature aimed at supporting valid use of assessment information (American Educational Research Association et al., 2014).
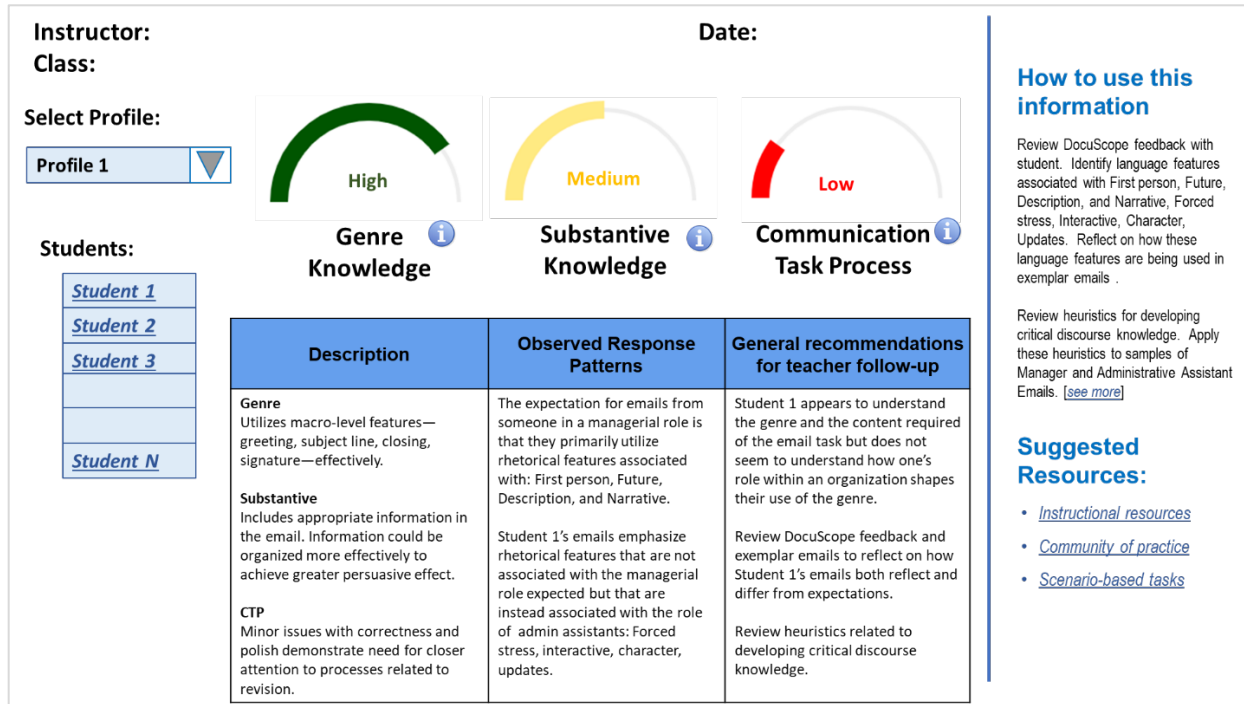
Teacher reports usually include student performance results at the individual, sub-group, and class levels; progress information; information to help understand current performance (e.g., Were my students engaged in the task?); and information that can help inform instruction (e.g., frequent errors and misconceptions; Zapata-Rivera, 2019).

In this case, PSRs for teachers include an overall classroom-level performance report with student performance information on high-level dimensions (e.g., Metacognition, Language - Critical Discourse, Rhetorical Aim). This overall PSR can be used by teachers to explore performance information at different levels.

Figure 2 displays a classroom report for teachers that shows students that belong to a cluster or profile based on response patterns. This figure shows students whose response patterns are consistent with a High level in Genre Knowledge, Medium in Substantive knowledge, and Low in Communication Task Process. The table below the profile provides more information, including sample response patterns associated with the profile, and provides possible recommendations/formative hypotheses for instructors. Data available to produce student profiles may include responses from items (e.g., both automated and human-produced scores) and process data. These profiles can be obtained by applying the scoring approaches described in Section 4.0. In addition, feedback from corpus-based technology-enhanced learning tools such as DocuScope can be used to provide additional recommendations for students and teachers (Helberg et al., 2018; see also Wetzel et al., 2021, this issue).

Figure 2

A Sample Prospective Score Report for Teachers in the Context of the WEC Prototype


Reports for students usually include actionable feedback to guide their learning (e.g., strengths, weaknesses, and recommendations for improvement), progress and performance information (e.g., individual performance and performance of comparison groups such as classroom performance information), and item-level results (Zapata-Rivera, 2019).

Figure 3 shows a prospective individual student report with performance information for Genre Knowledge. In this case, the student has demonstrated a High level on this dimension. Additional information for the student can include sample tasks, observed response patterns, and general recommendations.

In both cases, the PSRs include information to facilitate interpretation and use of the assessment results provided as well as links to additional educational resources.

Figure 3

A Sample Prospective Score Report for Students in the Context of the WEC Prototype
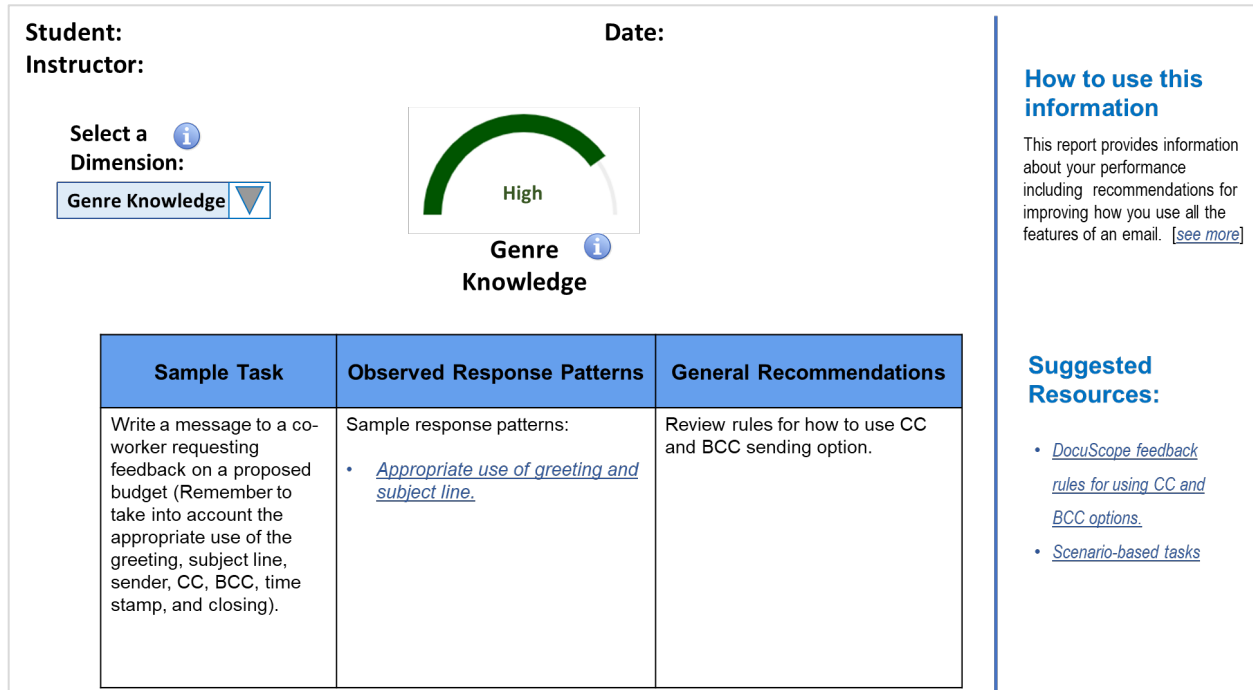
These PSRs have been designed based on feedback from two subject matter experts (SMEs) who have expertise in developing models of writing expertise and teaching the target students. Both SMEs appreciated the use of performance levels (i.e., high, medium, low) rather than numerical values, which is consistent with the formative purposes of the PSRs. The SMEs suggested including PSRs comparing individual and classroom-level performance on particular tasks or groups of tasks. These reports can be used by teachers to identify students who may need additional attention. Finally, SMEs valued the inclusion of information on how to interpret the PSRs and additional resources for teachers and students.

Additional PSRs may include information on factors that may impact student learning, including motivation levels, student interests, and learning approaches (e.g., student information such as number of attempts needed to complete items and reactions to feedback given by avatars on the tasks).

These PSRs are intended to guide assessment development activities. At some point, these PSRs will inform the implementation of the actual report. Evaluations with the intended audience will be carried out. These evaluations will include usability, comprehension, and preference aspects.

6.0 Conclusions and Future Directions

We have described the process of developing initial PSRs in the context of the WEC prototype. The creation of PSRs for different audiences provides the assessment development team with an opportunity to consider the audience requirements, document them, and keep them in mind when making assessment design decisions. Even though the PSRs may not be exactly the same as the final reports, they are always useful to analyze the report components that made it to the final version and the ones that needed to be removed due to operational constraints such as complexity of scoring processes or test length. There are many challenges ahead regarding the analysis and scoring of process data that can be used to augment the current reports.

A clear alignment of task and the intended constructs facilitates the creation of reports for various audiences. PSRs can be used to inform intended users about any changes in the design of the assessment.

Future research in this area includes designing additional PSRs for various audiences, implementing the reports, and evaluating them with the intended audiences following an iterative design and evaluation framework like the ones described here. Other potential audiences for reports resulting from tasks assessing communication and collaboration skills include potential employers and employees. These audiences will have different needs for assessment information. Designing score reports based on the audience's needs, attitudes, and prior knowledge have resulted in innovative and useful types of reports in the past (e.g., Zapata-Rivera & Katz, 2014).

Acknowledgements

Author Biographies

Diego Zapata-Rivera is Distinguished Presidential Appointee in the Learning and Assessment Foundations and Innovations Center at Educational Testing Service.

Jessica Andrews-Todd is Research Scientist in the Learning and Assessment Foundations and Innovations Center at Educational Testing Service.

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

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

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

Andrews-Todd, J., & Forsyth, C. M. (2020). Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. *Computers in Human Behavior*, *104*, 105759. https://doi.org/10.1016/j.chb.2018.10.025

Andrews-Todd, J., Forsyth, C. M., Steinberg, J., & Rupp, A. A. (2018). Identifying profiles of collaborative problem solvers in an online electronics environment. In K. E. Boyer & M. Yudelson (Eds.), *Proceedings of the 11th International Conference on Educational Data Mining* (pp. 239-245). International Educational Data Mining Society.

Andrews-Todd, J., & Kerr, D. (2019). Application of ontologies for assessing collaborative problem solving skills. *International Journal of Testing*, *19*(2), 172-187.

Bauer, M., & Zapata-Rivera, D. (2020). Cognitive foundations of automated scoring. In D. Yan, A. A. Rupp, & P. W. Foltz (Eds.), *Handbook of automated scoring: Theory into practice* (pp. 13-28). Taylor and Francis Group; CRC Press.

Brown, G. T. L., O'Leary, T. M., & Hattie, J. A. C. (2019). Effective reporting for formative assessment: The asTTle case example. In D. Zapata-Rivera (Ed.), *Score reporting research and applications* (pp. 107-125). Routledge.

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

Forsyth, C. M., Andrews-Todd, J., & Steinberg, J. (2020). Are you really a team player? Profiling of collaborative problem solvers in an online environment. In A. N. Rafferty, J. Whitehill, V. Cavalli-Sforza, & C. Romero (Eds.), *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)* (pp. 403-408).

Forsyth, C. M., Peters, S., Zapata-Rivera, D., Lentini, J., Graesser, A. C., & Cai, Z. (2017). Interactive score reporting: An AutoTutor-based system for teachers. In R. Baker, E. Andre, X. Hu, T. Rodrigo, & B. du Bouley (Eds.), *The Proceedings of the International Conference on Artificial Intelligence in Education* (Vol. 10331, pp. 506-509). Springer.

Goodman, D. P., & Hambleton, R. K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education*, *17*(2), 145-220.

Guo, H., Zhang, M., Deane, P., & Bennett, R. E. (2019). Writing process differences in subgroups reflected in keystroke logs. *Journal of Educational and Behavioral Statistics*, *44*(5), 571-596.

Hambleton, R. K., & Slater, S. (1997). *Are NAEP executive summary reports understandable to policy makers and educators?* (CSE Technical Report 430). National Center for Research on Evaluation, Standards, and Student Testing.

Hambleton, R. K., & Zenisky, A. L. (2013). Reporting test scores in more meaningful ways: A research-based approach to score report design. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology*: *Vol. 3. Testing and assessment in school psychology and education* (pp. 479-494). American Psychological Association.

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.

Hopster-den Otter, D., Wools, S., Eggen, T. J., & Veldkamp, B. P. (2017). Formative use of test results: A user's perspective. *Studies in Educational Evaluation*, 52, 12-23.

Kannan, P., Zapata-Rivera, D., Mikeska, J., Bryant, A., Long, R., & Howell, H. (2018). Providing formative feedback to pre-service teachers as they practice facilitation of high-quality discussions in simulated mathematics and science methods classrooms. In E. Langran & J. Borup (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 1570-1575).

Oliveri, M. E., & McCulla, L. (2019). *Using the Occupational Network Database to assess and improve English language communication for the workplace* (Research Report No. RR-19-28). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12265

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

Oliveri, M. E., Rupp, A. A., & Slomp, D. (February, 2020). *Designing digital modules for teaching and assessing workplace communication: Principles, methods, and challenges* [Paper presentation]. The 8th International Conference on Writing Analytics, St. Petersburg, FL, United States.

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*.

Romesburg, H. C. (2004). *Cluster analysis for researchers*. Lulu Press.

Tannenbaum, R. J. (2019). Validity aspects of score reporting. In D. Zapata-Rivera (Ed.), *Score reporting research and applications* (pp. 9-18). Routledge.

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

Zapata-Rivera, D. (2011, April). *Designing score reports that help teachers make instructional decisions* [Paper presentation]. The Annual Meeting of the American Educational Research Association (AERA), New Orleans, LA, United States.

Zapata-Rivera, D. (2019). Supporting human inspection of adaptive instructional systems. In R. Sottilare & J. Schwarz (Eds.), *Adaptive instructional systems. HCII 2019. Lecture notes in computer science* (Vol. 11597, pp. 482-490). Springer; Cham.

Zapata-Rivera, D., Kannan, P., Forsyth, C., Peters, S., Bryant, A. D., Guo, E., & Long, R. (2018). Designing and evaluating reporting systems in the context of new assessments. In D. Schmorrow & C. Fidopiastis (Eds.), *Augmented cognition: Users and contexts. AC 2018. Lecture notes in computer science* (Vol. 10916, pp. 143-153). Springer; Cham.

Zapata-Rivera, D., & Katz, I. R. (2014): Keeping your audience in mind: Applying audience analysis to the design of interactive score reports. *Assessment in Education: Principles, Policy & Practice*, *21*(3), 442-463.

Zapata-Rivera, D., VanWinkle, W., & Zwick, R. (2012). *Applying score design principles in the design of score reports for CBAL™ teachers* (ETS Research Memorandum RM-12-20). Educational Testing Service.

Zapata-Rivera, D., Zwick, R., & Vezzu, M. (2016). Exploring the effectiveness of a measurement error tutorial in helping teachers understand score report results. *Educational Assessment*, *21*(3), 215-229.

Zwick, R., Zapata-Rivera, D., & Hegarty, M. (2014). Comparing graphical and verbal representations of measurement error in test score reports. *Educational Assessment*, *19*(2), 116-138.