Research in Science Education
https://doi.org/10.1007/s11165-019-9846-8

Understanding Science and Language Connections: New Approaches to Assessment with Bilingual Learners

Cory Buxton[1] · Ruth Harman[2] · Lourdes Cardozo-Gaibisso[3] · Lei Jiang[2] · Khanh Bui[2] · Martha Allexsaht-Snider[2]

Published online: 21 May 2019

Abstract
We report on the use of bilingual constructed response science assessments in the context of a research and development partnership with secondary school science teachers. Given the power that assessments have in today's education systems, our project provided a series of workshops for teachers where they explored students' emergent reform-oriented science meaning-making in our project-designed assessments. Within the context of these workshops, we used discourse analysis to explore how three different groups grappled with the new reform-oriented relationship between science and language: (1) the research team's emergent understandings of how to create improved resources for teachers to better integrate science and language; (2) students' emergent understandings as expressed in their assessment responses; and (3) teachers' emergent understandings of how to integrate science and language in their instruction as expressed in interviews in the teacher writing workshops. Implications for curriculum designers, assessment developers, and professional learning facilitators are discussed.

Keywords Bilingual learners · Science assessment · Bilingual assessment · Cultural validity · Legitimation code theory · Systemic functional linguistics

As research on science learning, as well as research on learning more generally, has increasingly highlighted the sociocultural nature of how learning occurs (National Academies of Sciences, Engineering and Medicine [NAS] 2018a), a new wave of frameworks and standards to guide science education has emerged across the globe. From the European Commission's (2015) *Framework for Science Education for Responsible Citizenship* to the Canada 2067 (2017) *STEM Learning Roadmap*, to the *Framework for K-12 Science Education* (NRC 2012)

Edited by: Christine Howitt

✉ Cory Buxton
cory.buxton@oregonstate.edu

1 Oregon State University, 204A Joyce Furman Hall, Corvallis, OR 97331, USA

2 University of Georgia, Athens, GA, USA

3 Universidad ORT Uruguay, Montevideo, Uruguay

and the resulting *Next Generation Science Standards* (NGSS Lead States 2013) in the USA, these documents among others promote new sociocultural understandings of science learning. From this sociocultural perspective, these frameworks highlight the importance of validating and incorporating students' linguistic and cultural repertoires in classroom instruction through multimodal and open-ended learning tasks. From both equity and validity perspectives, the corresponding assessment systems that are being developed to measure student progress toward the goals of these new frameworks must also integrate open-ended and multimodal approaches, in order to both capture and strengthen the dynamic science learning advocated in these documents (Penuel and Shepard 2016; Turkan and Liu 2012).

While ideas about science learning and assessment are shifting rapidly, so too are international migratory patterns, highlighting the need for schools in many countries to serve increasingly multilingual student populations (Banks 2017). Despite growing efforts around developing assessments that align with the new frameworks and the changing student demographics, most classroom science assessments continue in the traditional modes that focus on measuring stored knowledge, as demonstrated through correctly identifying selected response items (Banilower et al. 2013).

Science assessments informed by a sociocultural perspective would document learning using tools that are closely aligned with the vision of how such learning should be constructed (Salleh et al. 2007). To this end, the US National Research Council released guidelines for new assessments, including advocacy for options such as mixed item formats, portfolios, and performances (NRC 2017). These guidelines highlight the importance of considering the sociocultural context of learning from the initial design of assessments, rather than relying on subsequent assessment accommodations and modifications at the classroom level. For example, Solano-Flores and Nelson-Barber (2001) proposed the construct of cultural validity to consider how the sociocultural context of students' lives influence their sense-making, assessment practices, and the nature of their responses. Lyon (2013) further operationalized cultural validity in science assessments as ensuring that assessments include elements of students' language and culture, local ecology, and broader global issues.

In the specific case of assessing bilingual learners, Gottlieb (2016) identified four categories of scaffolds to support bilinguals when completing assessment tasks: linguistic, graphic, sensory, and interactive scaffolds. Linguistic scaffolds include defining key terms, incorporating use of home language, and including sentence starters and language frames to help connect ideas (Unsal et al. 2018). Graphic scaffolds include graphic organizers, charts, and diagrams. Sensory scaffolds include models, manipulatives, and multimedia to support contextualization. Finally, interactive scaffolds include the ability to work in pairs or small groups, or to use interactive technologies that support the assessment task.

In the current study, we describe our collaborative learning with teachers to develop and interpret culturally valid assessments for bilingual learners. This work was one component of a five-part teacher professional learning framework for supporting bilingual learners in and beyond the science classroom (Buxton et al. 2015). In our first year of the broader project, we designed bilingual (English-Spanish) constructed response assessments to gauge students' understandings of science investigation practices through their written responses. These assessments were administered in participating teachers' classrooms at the start and end of each school year over a 3-year period. Over this time period, we also held workshops where teachers and researchers conducted a collaborative investigation of the student assessments. We view this collaborative role that teachers played in helping to shape both the assessment itself and how we made sense of the student responses as one of the unique features of this work.

We came to conceptualize emergent understandings about the relationship between science and language as existing in three contexts: (1) the research team's emergent understandings of how to create improved resources for teachers to better integrate science and language, (2) students' emergent understandings as expressed in their assessment responses, and (3) teachers' emergent understandings of how to integrate science and language in their instruction as expressed in focus group interviews at the teacher writing workshops. We look across these contexts to consider the following two research questions: (1) What insights on how to better integrate science practices and language practices do science educators gain from analyzing the constructed responses of bilingual students on reform-aligned assessment items? and (2) How can insights about the integration of science practices and language practices support science educators in rethinking how to scaffold student sense-making in constructed response writing prompts?

Conceptual Framework

This study applies a theory of social semiotics informed by systemic functional linguistics ([SFL]; Halliday and Matthiessen 2004). Our exploration of the assessments, therefore, is conceptualized through a view of language as a system that offers a pliable configuration of choices to make meaning in a wide variety of disciplinary and social contexts. In our view then, language is always implicated in knowledge building. With the shift in expectations across the new science learning frameworks as to how scientific reasoning and communication should unfold in classroom instruction, a parallel shift is required in how classroom science discourse is conceptualized. For example, teachers need to support students in making necessary shifts in language use, such as from peer negotiation in a lab group to structured oral explanation of their findings, or written science reports in a lab notebook. The dynamic and complex range of science practices calls for students to develop sophisticated semiotic (i.e., meaning-making) repertoires, including the understanding of multimodal graphics and models.

To support such complex meaning-making in linguistically diverse classrooms involves the integration of all available linguistic repertoires (including languages, modalities, registers) depending on the communicative context (Matthiessen 2015). For example, a pedagogy of translanguaging that values and leverages all the linguistic resources that students bring to the learning environment is essential for validating and supporting bilingual learners in constructing meaning (García 2009). This dialogic view of language has been gaining increased attention in science education (e.g., Karlsson et al. 2018).

A final theoretical perspective that pushed our thinking about integrating science and language practices was Maton's (2013) perspective on the sociology of knowledge, known as Legitimation Code Theory (LCT). LCT allowed us to reconsider the binary relationship that is often used to frame 'academic' discourse and 'everyday' discourse. As Gibbons (2006) has shown, teachers in content areas such as science frequently dilute disciplinary language to make it more easily accessible to bilingual learners, such as by translating technical discourse into concrete examples from everyday life. This simplifying process, which Maton refers to as moving *down the semantic wave*, can limit students' ability to build conceptual understanding if it is not combined with a move back *up the semantic wave* to a more metaphorical explanation of phenomena. This is especially true in the secondary school years, when concrete, everyday language starts to run out of explanatory power for the desired science meaning-making (e.g., Christie 2005).

Methods, Data Sources, and Analysis

The Language-rich Inquiry Science with English Language Learners (LISELL-B) project explored a broad range of questions about teaching and learning science with bilingual learners over a 3-year period in two school districts in the Southeastern USA that have experienced rapidly changing demographics driven largely by Latinx immigration from Mexico, Central, and South America. A total of 44 teachers participated in the project, co-constructing science curriculum and instructional support materials with members of the research team, focused on promoting language-rich, inquiry-rich, and interest-rich science investigations (Buxton et al. 2018). The bilingual constructed response assessments described earlier became the focus of one component of our teacher professional learning: the Teachers Exploring Student Writing Workshops, where teachers and researchers came together four times per year to look at student writing on the project assessments.

Three data sets were analyzed for this study: (1) the assessment forms themselves were explored as data to determine the degree to which they were constructed in ways that supported scaffolding bilingual learners' success in responding to the assessment items; (2) student assessment responses from the start and end of the 2015–2016 school year for all students in the classes of three focal teachers were analyzed, both by the teachers and by members of the research team; and (3) transcripts and field notes were created from the conversations with teachers during the writing workshops and artifacts from those workshops were collected during the 3 years of the project. Names of all people and places mentioned are pseudonyms.

In terms of data analysis, we drew from SFL ideational and textual meaning systems (Martin and Rose 2003) and from the LCT code of semantics (Maton 2013) to analyze both the assessment items themselves and the student responses. Namely, through these lenses, we explored the emergent understandings of students (via their assessment responses), and of researchers (via our analysis of the curriculum materials). Specifically, in the SFL analysis, the ideational analysis helped us to understand how experiences about science were represented, and the textual analysis helped us understand how those experiences were organized in the texts (Martin and Rose 2003). The SFL findings also supported us in measuring the degree of semantic gravity and semantic density in the texts (Maton 2013). Semantic gravity refers to the level of abstraction or the degree to which meaning relates to its context, indicated as high semantic gravity (SG+) for more highly contextualized or concrete examples and low semantic gravity (SG−) for more abstraction with less context. Semantic density refers to the level of symbolic condensation of meaning, or the degree to which meaning is condensed into symbols, indicated as high semantic density (SD+) for more use of symbolic condensation and low semantic density (SD−) for less use of symbolic meaning. By combining SFL and LCT analyses, we hoped to gain emergent understandings about the integration of science and language practices.

To analyze the teacher focus group data, we used conceptual codes developed from our review of the literature to conduct a thematic analysis (Fereday and Muir-Cochrane 2006) of the teachers' understandings about bilingual language use, scientific investigations, and the function of language in the new science classroom. We sought to categorize teachers' emergent understandings about the integration of language and science via their discussions during the assessment workshops.

Findings

In the following three sections, we describe the emergent understandings about the integration of language and science that developed among researchers, students, and teachers as we engaged together in this work. We undertook this multi-perspective analysis, synthesizing insights from student work, teacher reflections, and from the instructional and assessment materials themselves, as we believe that each of these perspectives has different insights to offer in the challenging task of developing educative science assessments that are culturally and linguistically valid. While we see our collaborative explorations with teachers as of utmost importance in this discussion, we begin by describing aspects of the assessments themselves to provide important context, before moving on to describe the student responses and finally the teachers' reflections.

Emergent Researcher Understandings from Analysis of the Assessment Booklets

We begin with our own efforts to rethink how to scaffold student sense-making through the design of constructed response writing prompts. Our analysis of the assessments themselves, designed with the learning of students, teachers, and researchers in mind, was guided by our use of SFL and Maton's semantic codes. We note that we had not yet adopted Maton's framework when we developed the assessments at the start of the project, so our post hoc analysis revealed strengths and weaknesses of our assessment design that we were unaware of prior to project implementation.

Our assessment was developed around seven science investigation scenarios (see Fig. 1 for an example). Semantic analysis showed that each scenario began by providing concrete contexts (high semantic gravity [SG+]) and in most cases, began with less condensed language (low semantic density [SD−]). Nearly all scenarios then shifted toward more abstract, generalizable statements (SG−) with more condensed and symbolic language (SD+).

For example, in a scenario about plant growth, the task begins with the sentence "Carmen knows that plants need light to grow." This point of departure supports the learner in

Weight Lifting Experiment

Luis knows that lifting weights makes people stronger. Luis wonders if different **types** of weight lifting exercises affect how strong a person gets. He decides to design an experiment to find out. Think about the experiment that Luis could design and choose the best answer for questions 15 and 16. Write your answer on the answer sheet.

15) What would be the ***independent variable*** in Luis' experiment about how different **types** of weight lifting exercises affect how strong a person gets? (The independent variable is the variable that Luis changes or manipulates.) The **independent** variable in Luis' experiment would be

A) The type of weight lifting exercise that people do.
B) The length of time the people lift weights.
C) The number of repetitions the people do when lifting the weights.
D) How strong the people get after a length of time.

16) What would be the ***dependent variable*** in Luis' experiment about how different **types** of weight lifting exercises affect how strong a person gets? (The dependent variable is the variable that changes in response to the independent variable.) The **dependent** variable in Luis' experiment would be

A) The type of weight lifting exercise that people do.
B) The length of time the people lift weights.
C) The number of repetitions the people do when lifting the weights.
D) How strong the people get after a length of time.

Next, write your answers for questions 17 and 18 in the boxes on the answer sheet.

17) List **2** other **Variables** that Luis would need to ***control*** (keep the same) in his experiment to be sure that he has a fair test.

18) Use **scientific language** to explain the experiment Luis could do to find out how different **types** of weight lifting exercises affect how strong a person gets.

Experimento de levantar pesas

*Luis sabe que el levantamiento de pesas hace a la gente más fuerte. Luis se pregunta si los diferentes **tipos** de ejercicios de levantamiento de pesas afectan lo fuerte que una persona puede ser. Él decide diseñar un experimento para averiguarlo. Piensa en el experimento que Luis podría diseñar y elije la mejor respuesta para las preguntas 15 y 16. Escribe tu respuesta en la hoja de respuestas.*

*15) ¿Cuál sería la **variable independiente** en el experimento de Luis sobre cómo diferentes **tipos** de ejercicios de levantamiento de pesas afectan lo fuerte que una persona puede ser. (La variable independiente es la variable que Luis cambia o manipula.) La variable **independiente** en el experimento de Luis sería*

A) El tipo de ejercicio de levantamiento de pesas que la gente hace.
B) La cantidad de tiempo que la gente levanta pesas.
C) El número de repeticiones que la gente hace al levantar las pesas.
D) Lo fuerte que las personas son después de un periodo de tiempo.

*16) ¿Cuál sería la **variable dependiente** en el experimento de Luis sobre cómo diferentes tipos de ejercicios de levantamiento de peso afectan lo fuerte que una persona puede ser. (La variable dependiente es la variable que cambia en respuesta a la variable independiente.) La variable **dependiente** en el experimento Luis sería*

A) El tipo de ejercicio de levantamiento de pesas que la gente hace.
B) La cantidad de tiempo que la gente levanta pesas.
C) El número de repeticiones que la gente hace al levantar las pesas.
D) Lo fuerte que las personas son después de un periodo de tiempo.

A continuación, escribe tus respuestas para las preguntas 17 y 18 en las casillas de la hoja de respuestas.

*17) Menciona otras **2 variables** que Luis necesita controlar (mantener igual) en su experimento para asegurarse de que tiene una prueba justa.*

*18) Utiliza **lenguaje científico** para explicar el experimento que Luis podría hacer para averiguar cómo diferentes **tipos** de ejercicios de levantamiento de pesas afectan lo fuerte que una persona puede ser.*

Fig. 1 Weightlifting scenario from bilingual constructed response assessment identifying with the girl, Carmen, and with the concrete experience of plants requiring light to grow. The next two sentences move up the semantic wave by becoming somewhat more linguistically abstract and condensed ("She wonders if different types of light affect how much a plant will grow. She decides to design a science fair experiment to find out by comparing plant growth inside using a 'grow light' and plant growth outside in the sun."). The scenario highlights key scientific relationships (plants' responses to different light sources) and involves students' scientific and analytic thinking (controlling variables to design a fair test), but at the same time, the scientific meaning is grounded in language and experiences that should be familiar to most middle school students (e.g., growing plants, conducting a science fair experiment). As the scenario progresses, the language increases in semantic density as the students are asked to describe "the independent variable in Carmen's experiment about how different types of light affect plant growth." Here, one can see how nominalization (e.g., plant growth) and expanded noun clauses represent a more generalizable principle under investigation. With the movement from more concrete to more abstract questions and patterns of language, the scenario supports students into moving into more abstract scientific thinking and writing.

Our assessment was also designed to support a translanguaging approach to multilingual meaning-making. Our semantic analysis highlighted some of the ways that our materials promoted a translanguaging mindset. For example, if students are given a choice at the start of an assessment between an English version or a Spanish version of the assessment, this implicitly sends the message that students should "pick" the language they feel most comfortable using in a science context and use that language only. The same basic message is conveyed if an assessment is double sided with English on one side and Spanish on the other. To address this issue, we used large format paper so that all students, regardless of their language proficiencies, saw English and Spanish side by side for each scenario, with all images, figures or data located in the middle between the two languages and labeled bilingually (see Fig. 1). Further, for each question, all students were asked to check a box indicating if they read the question in English only, in Spanish only, or in both languages, and students were encouraged to use all available multilingual and multimodal resources to help them understand the questions and to write their answers. When taken together, we found that our assessment was largely supportive of bilingual learners, both through normalizing and centering bilingualism and multilingual resources, rather than normalizing monolingualism and treating bilingual resources as an accommodation (García 2009) and by constructing the text of the items in ways that reflected Maton's ideas of semantic waving.

Emergent Student Understandings from our Analysis of Student Responses

Our second area of exploration focused on how analysis of bilingual learners' constructed responses could provide insights into their emergent understanding of new science and language practices. Using SFL and Maton's LCT semantic codes, we examined the interplay between students' scientific knowledge and their language practices, as expressed through their writing.

Overall, we found that students expressed a range of emergent scientific understandings that were at least partially aligned with the expectations of the new science frameworks. First, the analysis revealed that many students were adept at constructing emergent scientific meaning through use of concrete language and examples. In LCT terms, students successfully used high semantic gravity and low semantic density in their responses in ways that were both relevant and mostly accurate. For example, in a scenario that focused on how to design an experiment to test whether different types of weightlifting would make a person stronger, a student responded, “When a person is lifting more heavy weights, that person is struggling the most, and the body muscles are doing the job and that is how a person accumulates more strength.” In this response, the student used concrete and non-symbolic language (SG+/SD−) to express emergent scientific knowledge. Thus, we found that students were generally able to express a basic scientific understanding using language that moved down the semantic wave. However, while such language can lay the foundation for student meaning-making, to fully engage in the generalizable learning advocated in the reform documents requires a shift to more abstract language that moves back up the semantic wave.

Some students did demonstrate generalizable understandings through the use of abstract language. In another example from the weightlifting scenario, one student wrote: “to see how different types of weight lifting exercises affect [how] strong a person gets they would need to test each.” In this example, the student used an expanded noun clause about a generalized situation, thus expressing a more abstract concept about the experiment they would design (SG−/SD+). Similarly, in a scenario describing a sling shot, some students began with a focus on the concrete object “stone”, and then transitioned to discuss the underlying principle of energy transfer (SG−/SD+). Many of these examples of students’ engagement in semantic waving involved their intertextual appropriation of more abstract and dense scientific language from the assessment prompts. In other words, the language of the questions supported students in moving recursively from concrete to abstract explanations.

While we found various examples of students who successfully appropriated semantic resources, we also found that many students struggled in articulating their responses. Their language use sometimes led to off-topic answers, often because of incomplete clauses and lack of semantic waving. For example, when asked to describe possible variables that could be used in an investigation comparing the health of cloned and non-cloned sheep, several students answered in short phrases such as, “same sheep; same place” or “food given; how long there.” Answers like these, while demonstrating an emergent understanding of the role of variables in science investigations, resemble the type of unfinished responses typical in an oral lab group discussion, with minimal semantic density and few abstract generalizations (SG+/SD−). In our view, this tendency of students to abbreviate their responses was most likely due to a lack of explicit scaffolding on how to shift language modes from lab group negotiation to written reporting of phenomenon.

Overall, the patterns we saw in the student writing were twofold. On the one hand, some students succeeded in using semantic waving to describe and explain the phenomena under investigation, especially when they used the language from the questions as an intertextual resource to scaffold their responses. On the other hand, many students seemed stuck in the language of peer negotiation expected in a lab investigation and did not move into the more generalized or specialized discourses called for in new science education frameworks.

Emergent Teacher Understandings

Finally, we turn to how the teachers’ analysis of their students’ constructed responses on our reform-aligned assessment items provided us with insights about the students’ writing and about how we could better integrate science practices and language practices in our curricular and assessment design work. As researchers and teachers working collaboratively in the space of the writing workshops, we helped each other look beyond students’ scores to explore what other information could be learned from their written responses.

Teachers often credited the experience of exploring their students' writing on the project assessments as allowing them to see more clearly the linguistic and science resources students brought to the task. As Shelly explained,

> Looking at my students' papers was really eye-opening. Some of them who wrote some really good answers almost never talk in class, and other ones who I thought would do really well didn't write much at all. So one thing I learned is that we don't ask our kids to write in this way that really requires them to explain their thinking in writing and we need to do that more. (workshop on Feb 16, 2013)

During a focus group interview at a workshop in the second year of the project, multiple teachers discussed and agreed on the importance of making the different language registers that are needed to communicate science thinking explicit for their students:

> Margo: We need to be more intentional about how we scaffold using everyday language and academic language ourselves.
> Stacey: You could tell they knew what was going on but they had trouble with expressing their understandings in either everyday or academic language.
> John: We need to give students everyday language and help them translate it into scientific language and go back and forth from academic language to everyday language and from everyday language to academic language. (workshop on Nov 8, 2014)

While these teachers exhibited somewhat binary views of academic and everyday language registers, they showed clear emergent understanding of the importance for students to learn through the use of multiple registers when communicating their science sense-making. A third emergent understanding for teachers that became increasingly clear as we explored the students' assessments was the need to engage all students in challenging and open-ended tasks:

> Danny: Looking at the assessments helped me see that I'm probably taking it too easy on my English learners. It was really hard for my kids and they left a lot blank. But that's because they haven't been pushed enough and most of them have never seen something with those expectations, so they couldn't handle it. I need to raise my expectations for all my students. (workshop on Dec 6, 2014)

As evidenced in Danny's comments above, the material evidence of the assessment writing and the discussions around them supported more understanding of the importance of grade level scaffolding of bilingual learners within the context of disciplinary instruction.

By exploring their students' writing in this way, teachers also came to see how the students' responses often mirrored features of the language used in the prompts. Thus, when the language of the questions was concrete and simple (e.g., *Maria wants to know what happens when...*), students produced the same style in their response (e.g., *Maria puts the wire...*). However, when the language in the questions became more abstract and dense (e.g., *Describe the result in terms of potential and kinetic energy*), then many students also incorporated this language into their responses (e.g., *Kinetic energy increased because the stone is in motion when...*). Teachers came to see that the insights they gained from careful exploration of students' writing had implications for how they could design instructional and assessment tasks with more intentional language scaffolding.

Finally, the bilingual nature of the assessment helped the predominantly monolingual, English speaking teachers in the project to see the value of supporting their bilingual students with science learning resources in their first language:

> Megan: Because I don't speak Spanish, I haven't really encouraged my Spanish speaking kids to use it in my class. I figured I could help them better if I could always understand what they were saying. But seeing how many of my kids checked the box [on the assessment] that they read the questions in both languages really makes me think that having the materials in Spanish for them could really help them to understand more. (workshop on Dec 5, 2015)

When taken together, we can see that teachers in the project were actively developing new understandings of the connections between science and language as we engaged in the collaborative exploration of their students' writing.

Discussion and Implications

We began this study with the following two research questions: (1) What insights on how to better integrate science practices and language practices can science educators gain from analyzing the constructed responses of bilingual students on reform-aligned assessment items and (2) How can insights about the integration of science practices and language practices help science educators rethink how to scaffold student sense-making when using constructed response writing prompts?

In answering our first question, we found that the three groups involved in this study—the teachers, the students, and the researchers—each developed emergent understandings of integrating science and language but not in identical ways. Teachers' emergent understandings became most clearly visible through the ways they discussed support structures their students needed to further develop science and language practices. For example, teachers increasingly recognized the value of making Spanish language resources available even when they themselves were not proficient in the students' first language. Teachers also highlighted the importance of providing time and opportunities in their classrooms to practice register shifting and use of multiple modalities for all students. At the same time, teachers also expressed some fixed notions about "academic language" and "everyday language," and the role of language instruction in content area learning, making it challenging for them to think about how to alter their science practices and language practices. These findings resonate with the challenges that Penuel and Shepard (2016) raise in supporting teachers to become more critical evaluators of how assessments can and should inform their classroom practices. Overall, however, while the teachers were clear that they still had much to learn about building on the science and language resources of their bilingual students, they also demonstrated growth over the course of their work analyzing student written responses in the assessment workshops. For example, teachers expressed various new ways of thinking about the cultural and linguistic resources that all their students brought to the science classroom.

As for the students themselves, they showed important understandings about the role of disciplinary language on some of the question responses, as they made language choices that began to show awareness of the role that word choice and patterns of language use play in science sense-making. Students were also learning to make use of the bilingual and multi-modal resources embedded in the assessment. Yet, many bilingual learners continued to struggle in using scientific discourse that extended beyond concrete language and examples. While we encountered many examples where students expressed emergent understandings of scientific concepts on the assessment items, their reliance on concrete language served to diminish the accuracy and completeness of their expressions of scientific meaning. That is, many students were able to start their answers with concrete description (SG+/SD−) but failed to advance their thinking and writing through more complex reasoning and discourse (SG −/SD+) to meet the new expectations for grade-level scientific sense-making. This should not be surprising, given that language is always implicated in knowledge building (Halliday and Matthiessen 2004). In other words, without explicit scaffolding of the necessary shifts in genre from students' collective sense-making (such as in a lab group) to individual representation of those ideas in written formal reporting (such as a constructed response assessment), students can remain fossilized in patterns of language use that do not advance their science sense-making.

Developing fluent and flexible use of concrete language is an essential starting point for students who are learning to express their emergent scientific understanding. Teacher acceptance of such language particularly serves bilingual learners as they begin to engage in science reading and writing in a new language. Students' constructed response answers on our assessments show that bilingual learners could express grade appropriate ideas about science content when they were able to start with more familiar language to help them process scientific thinking. However, our findings also reinforce Gibbons' (2006) point that more familiar language (SG+/SD−) should be seen as a starting point for engaging in semantic waving, rather than being accepted as "appropriate" or "good enough" for bilingual learners. Indeed, we found that students who engaged in more robust semantic waving in their responses were better able to articulate their disciplinary knowledge in grade-appropriate ways. We note that student difficulty in articulating understandings of more abstract concepts using denser disciplinary discourse is a pattern also found in studies of bilingual students' attempts to explain their mathematical reasoning (Herbel-Eisenmann et al. 2013).

To answer our second research question, we considered what we have learned from our attempts to scaffold student sense-making through the use of constructed response writing prompts. We began the project aware that the design of written assessments plays a major role in scaffolding students' sense-making (Meyer and Turner 2002). We hoped that our assessment items would help teachers to think about their students' sense-making and communicating in new ways, while also helping us to improve our instructional and assessment materials.

We can conclude that we were partially successful in this goal. Insights we developed through our collaborative explorations with teachers in the project have continued to push our thinking in new directions about how to better support the integration of science and language practices. We believe that the kinds of multi-perspective analysis that we undertook in this study, synthesizing insights from students, teachers, and from the instructional and assessment materials themselves, will be increasingly necessary as a new generation of science learning frameworks are being introduced in diverse parts of the globe. These frameworks converge around sociocultural understandings of learning, highlighting the need to recognize and support a broad range of students' linguistic practices and funds of knowledge through multimodal and open-ended learning and assessment tasks (National Academies of Sciences, Engineering, and Medicine 2018b). To this end, elsewhere (Harman et al. 2018) we have proposed a new framework we refer to as Culturally Sustaining Systemic Functional Linguistics for Science Learning and Assessment (CS-SFL Science) to scaffold the languages, modalities and semantic features of learning and assessment tasks, while attending to culturally relevant contexts and topics.

Thus, our findings have implications for various aspects of science teaching and learning, but perhaps none greater than the implications for the next generation of science assessments that are being developed to both measure and shape students' science learning and teachers' science instructional practices. The developers of these new assessments are working to create culturally valid measures (Solano-Flores and Nelson-Barber 2001) for the multiple facets of the new science learning frameworks. However, based on publicly available documents, such as rubrics currently being used to screen possible assessment items in the US context (Achieve 2018), important questions about how the language of the assessment items is being designed seem not to be considered. We are not surprised by this, as we ourselves did not think deeply about semantic patterning when we began constructing our project assessments. It was only through our collaborative learning with teachers as we studied student responses on our assessments, and then through our subsequent study of Maton's (2013) work and our analyses of the assessment items themselves, that we gained a fuller understanding of how science practices and language practices are mutually constitutive. Curriculum developers and teacher educators can likewise build on these ideas such that teacher preparation and curriculum materials meaningfully integrate semantic features of learning and assessment tasks. (Lemmi et al. 2019). Given the high stakes nature of the future science assessments, however, it is particularly incumbent on assessment developers in all nations to deepen their own emerging understandings of the complex relationships between conceptual learning and language. We hope that our ongoing work in this arena can contribute to the creation of more culturally valid and linguistically supportive assessments for all science learners.

Funding Information This work was supported by the National Science Foundation under Grant Number # DRL-1316398. Opinions and conclusions are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Achieve (2018). Next generation science standards task screener. Downloaded Nov 28, 2018 from: https://www.nextgenscience.org/taskscreener.

Banilower, E., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weiss, A. M. (2013). *Report of the 2012 national survey of science and mathematics education*. Chapel Hill, NC: Horizon Research Inc.

Banks, J. A. (2017). *Citizenship education and global migration: Implications for theory, research, and teaching*. New York, NY: American Educational Research Association.

Buxton, C., Allexsaht-Snider, M., Kayumova, S., Aghasaleh, R., Choi, Y., & Cohen, A. (2015). Teacher agency and professional learning: Rethinking fidelity of implementation as multiplicities of enactment. *Journal of Research in Science Teaching, 52*(4), 489–502. https://doi.org/10.1002/tea.21223.

Buxton, C., Cardozo Gaibisso, L., Xia, Y., & Li, J. (2018). How perspectives from linguistically diverse classrooms can help all students unlock the language of science. In L. Bryan & K. Tobin (Eds.). 1*3 Questions: Reframing Education's Conversation: Science* (pp. 273–291). New York: Peter Lang.

Canada 2067. (2017). Canada 2067 STEM learning roadmap. Accessed on Mar 28, 2019 from http://bit.ly/2FFtFdO.

Christie, F. (2005). *Language education in the primary years*. Sydney: University of South Wales Press.

European Commission. (2015). *Science education for responsible citizenship*. Brussels: Authors.

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods, 5*(1), 80–92.

García, O. (2009). *Bilingual education in the 21st century: A global perspective*. Oxford: Blackwell.

Gibbons, P. (2006). *Bridging discourses in the ESL classroom: Students, teachers and researchers*. London: A&C Black.

Gottlieb, M. (2016). *Assessing English language learners: Bridges for language proficiency to academic achievement*. Thousand Oaks: Corwin.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (3rd ed.). London: Edward Arnold.

Harman, R., Buxton, C., Cardozo-Gaibisso, L., Jiang, L., & Bui, K. (2018). Gotta know the tune to riff? Culturally sustaining SFL praxis in science classrooms. *Presentation at the 2018 meeting of the International Systemic Functional Congress (ISFC)*. Boston, MA.

Herbel-Eisenmann, B., Steele, M., & Cirillo, M. (2013). (developing) teacher discourse moves: A framework for professional development. *Mathematics Teacher Educator, 1*(2), 181–196.

Karlsson, A., Larsson, P. N., & Jakobsson, A. (2018). Multilingual students' use of translanguaging in science classrooms. *International Journal of Science Education*, 1–21.

Lemmi, C., Brown, B. A., Wild, A., Zummo, L., & Sedlacek, Q. (2019). Language ideologies in science education. *Science Education*, 1–21. Early view downloaded May 1, 2019 from https://onlinelibrary.wiley.com/doi/abs/10.1002/sce.21508.

Lyon, E. G. (2013). Conceptualizing and exemplifying science teachers' assessment expertise. *International Journal of Science Education, 35*, 1208–1229.

Martin, J. R., & Rose, D. (2003). *Working with discourse: Meaning beyond the clause*. London: Bloomsbury Publishing.

Maton, K. (2013). Making semantic waves: A key to cumulative knowledge-building. *Linguistics and Education, 24*(1), 8–22.

Matthiessen, C. M. (2015). Register in the round: Registerial cartography. *Functional Linguistics, 2*(9), 1–48.

Meyer, D. K., & Turner, J. C. (2002). Using instructional discourse analysis to study the scaffolding of student self-regulation. *Educational Psychologist, 37*(1), 17–25.

National Academies of Sciences, Engineering, and Medicine. (2018a). *How people learn II: Learners, contexts, and cultures*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24783.

National Academies of Sciences, Engineering, and Medicine. (2018b). *English learners in STEM subjects: Transforming classrooms, schools, and lives*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25182.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

National Research Council. (2017). *Seeing students learn science: Integrating assessment and instruction in the classroom*. Washington, DC: The National Academies Press.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies.

Penuel, W. R., & Shepard, L. A. (2016). Assessment and teaching. In D. Gitomer & C. Bell (Eds.), *Handbook of research on teaching* (pp. 787–850). Washington, D.C.: American Educational Research Association.

Salleh, R., Venville, G., & Treagust, D. (2007). When a bilingual child describes living things: An analysis of conceptual understandings from a language perspective. *Research in Science Education, 37*(3), 291–312.

Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching, 38*(5), 553–573.

Turkan, S., & Liu, O. L. (2012). Differential performance by English language learners on an inquiry-based science assessment. *International Journal of Science Education, 34*(15), 2343–2369.

Unsal, Z., Jakobson, B., Molander, B.-O., & Wickman, P.-O. (2018). Language use in a multilingual class: A study of the relation between bilingual students' languages and their meaning-making in science. *Research in Science Education, 48*(5), 1027–1048.